Delft University of Technology

MSc Applied Physics

Feasibility study on the integration of a deep learning dose computation algorithm in an automated proton therapy treatment planning system

Master Thesis

Mitchel Graauw

Supervised by:

Dr. Z. Perkó
Asst. Prof. Dr. Ir. S. Breedveld

March 22, 2024

Abstract

The dose computation algorithm, or dose engine, is one of the fundamental parts of radiotherapy treatment planning. These algorithms predict how the dose will be distributed inside the patient. Current dose engines are mainly based on either Monte Carlo simulations (MC) or pencil beam algorithms (PBA). MC being very precise, but relatively slow. PBA being quicker, but generally lacking accuracy. Since treatment planning requires both high speed and accuracy, one would prefer MC accuracy with even higher speeds than PBA. A recent study showed a possible solution based on deep-learning called the Dose Transformation Algorithm (DoTA). This deep-learning algorithm is capable of doing MC accurate dose calculations and is faster than PBA (Pastor-Serrano and Perkó 2022).

In this project, a feasibility study has been performed on the integration of DoTA as a dose engine in actual treatment planning. The treatment planning system (TPS) in this study is Erasmus-iCycle (Breedveld, Storchi, et al. 2012). This study included the creation of an algorithm to do dose computations with DoTA for any given set of parameters given by the TPS. Subsequently, the dose computations by DoTA have been compared to those computed by Erasmus-iCycle's current dose engine, ASTROID (Kooy et al. 2010). Analyses on these dose computations included comparisons in a homogeneous water box, alternative homogeneous matter and patient geometries.

Two main sources of discrepancy between DoTA and ASTROID where the beam's range and the beam model used by ASTROID, compared to what DoTA was trained on. Both dose engines likely use a different interpretation of the proton stopping power, leading to range discrepancies up to 14.9% for 200 MeV beamlets when projected in a homogeneous matter of 1000 Hounsfield Units (HUs). Comparing the dose distributions in water, the maximum dose discrepancy around the Bragg peak (BP) for a 80 MeV beam was about 60.0%, due to the width of the beam being larger for DoTA. The mean dose discrepancy in water reached a maximum of 18.9%. In a patient geometry, the range differences made the mean discrepancies go up to a maximum of 22.9%, as expected from the range discrepancies found earlier. Implementation of different gantry and beamlet angles increased the discrepancies, likely caused by the interpolation required to perform calculations under these angles. In terms of distributed energy, the models were closer, with the mean discrepancy decreasing to maximum of 7.1%. Computations of two treatment plan dose distributions showed that the discrepancies arising from this beam model and range difference were to large to achieve viable dose value histograms. A two lateral beam plan showed the better results of the two plans with an under dosing of 20.8%, likely due to robustness occasionally compensating for the range discrepancy.

Contents

1 Introduction

Cancer is one of the leading causes of death in the world, still taking millions of lives globally each year (American Cancer Society 2018). With the amount of cancer cases increasing yearly (Siegel et al. 2023), continuing to improve the quality of cancer treatment is a must. Thanks to earlier detection of cancer and improved treatment methods, the survival rate of cancer is slowly but surely increasing over the past decades (Schiffman, Fisher, and Gibbs 2015)(Crosby et al. 2022).

In general there are three main types of treatments for cancer: chemotherapy, surgery and radiotherapy. Radiotherapy can roughly be split into two areas: Internal and external radiotherapy. Internal, meaning that the radiation in coming from an implant or liquid inside the body, while with external radiotherapy, the radiation comes from a source located outside of the body. External radiotherapy has a variety of methods and is performed with ionizing radiation, most commonly in the form of a collection of photon beams. While this type of treatment shows good results in terms of its contribution to cured cancer cases, sometimes in combination with other treatment types (Breedveld, Craft, et al. 2019), there is also a comparable method that uses protons instead of photons , called proton therapy, which is slowly becoming more relevant in the field of radiotherapy (Mohan and Grosshans 2017)(Mohan 2022).

In radiotherapy treatment planning, one of the most important steps is the dose calculation. Within a treatment planning system, a certain dose engine is used to perform this calculation. For the sake of treatment speed and quality, these computations need to happen fast and precise. However, the dose engines currently used in treatment planning are either quick, but somewhat inaccurate in complex patient geometries or precise, but very computation-heavy and therefore very slow (Valk Bouman 2022)(Pastor-Serrano and Perkó 2022). In an ideal case, we would want to be able to adjust treatment plans while the patient is being treated, such that the beams can be adjusted for small deviations in the tumor position, due to for example internal motion. This is called real-time adaptive treatment planning (Albertini et al. 2020)(Jagt et al. 2018).

1.1 Aim of the project

The aim of this project is to do a feasibility study on the integration of a deep learning-based dose calculation model, DoTA (Pastor-Serrano and Perkó 2022), in Erasmus iCycle, an automated treatment planning system for both photon and proton therapy (Breedveld, Storchi, et al. 2012). DoTA is a deep learning based model that has been trained on Monte Carlo dose calculations, and has proven to calculate doses faster than pencil beam algorithms with the accuracy of Monte Carlo dose engines (Pastor-Serrano and Perkó 2022). This would theoretically make this model a very good option for treatment planning, due to its combination of speed and accuracy.

Between dose calculation and realistic treatment planning scenarios lies a gap that needs to be thoroughly analyzed. DoTA is capable of performing single beamlet dose calculations, taking the beamlet energy and a patient volume as inputs. However, it does not take into account the fact that beamlets can exit the source under an angle and that therefore the total beam diverges both inside the patient, as well as on its way to the patient, as the beam's source is located roughly 2 metres away. Therefore this work will focus on two things. Firstly, building an algorithm around DoTA such that it can perform dose calculations of realistic beamlets used in treatment planning, correcting for the different beamlet angles. Secondly, analyze how well DoTA compares to ASTROID (Kooy et al. 2010), the currently used dose computation engine in Erasmus-iCycle, under a variety of different circumstances. Based on the results produced by this analysis, we try to answer the research question of this project:

Could we use a deep learning based dose calculation algorithm (DoTA) as a dose engine in Erasmus iCycle treatment planning environment?

1.2 Thesis Outline

In Section 2, a theoretical background that is deemed necessary to better understand the further contents of the project is provided, including both a background on radiotherapy, treatment planning and its challenges, as well as the deep learning aspect. In Section 3, the algorithmic basis and the different methods of data generation will be described. In Section 4, the results of the project will be presented, containing comparisons between dose distributions and an analysis on certain beamlet behaviour in a variety of circumstances. In Section 5, the results will be analysed and explained. Moreover, it will contain a critical view on both this work and its model's limitations. In Section 6, the main contents of this project will be highlighted, including an answer to the research question.

2 Theoretical Background

In this section, the relevant theory and context on this project is being presented. The first part will give a theoretical basis on radiotherapy treatment modalities and the radiotherapy workflow, as well as a more in depth look into the challenges that research in this field faces and why it is so important to overcome the challenges that are related to radiotherapy, and why proton therapy, while being a solution to some of these problems, still has a way to go. The second part will contain more focus on the computational physics that lie under these dose computation engines and the treatment planning systems.

2.1 Radiotherapy

2.1.1 Treatment Methods

As briefly discussed in the introduction, there are three distinct types of cancer treatment: surgery, chemotherapy and radiotherapy. Often, in order to achieve the best results a combination of these three is required (Breedveld, Craft, et al. 2019). Cancer treatment can be both curative and palliative. Curative treatment aims to cure the patient, while palliative treatment tries to relieve symptoms to improve quality of life, mostly used when curative treatment is not possible.

Surgery generally deals the least amount of damage to healthy tissue in the body. However, in most cases surgery will not be used as a standalone method to cure a patient. Besides playing a role in cancer diagnosis (Wyld, Audisio, and Poston 2015), surgery serves 3 main purposes within cancer treatment: tumor removal, tumor debulking and easing cancer symptoms. Being a local treatment method, surgery is generally not used in cases where the cancer has already spread through the body, or for example Leukemia, where the cancer flows through the blood (National Cancer Institute 2015).

Chemotherapy is a more rigorous approach that kills both cancer cells and healthy cells from the inside out, in the form of medication. Chemotherapy is often used in combination with radiotherapy or surgery, either to shrink the tumor beforehand (neoadjuvant chemotherapy), or target leftover cancer cells that might have been missed by the other modalities (adjuvant chemotherapy) (National Cancer Institute 2015)(McElnay and Lim 2014).

Radiotherapy is most commonly used treatment method, sometimes combined with surgery and chemotherapy. About 50% of patients receive radiotherapy (G. Delaney et al. 2005)(Begg, Stewart, and Vens 2011), with an estimation that radiotherapy contributes to around 40% curative treatment (Barnett et al. 2009). Radiotherapy utilizes ionizing radiation to eliminate both cancer cells and healthy cells, and can be divided into two categories: internal and external beam radiotherapy. Internal beam radiation in commonly used in areas of the body near a natural cavity. The idea is that a radioactive material is implanted inside the body near the tumor, such that it can irradiate the tumor without irradiating too much healthy tissue. The other option is external beam radiotherapy, where an ionizing beam of radiation is aimed at the tumor originating from a source located outside of the patient's body. Generally, photons are being used as the ionized particles. The drawback of this method is that the beam has to propagate through the entire patient in order to reach the tumor. Given the nature of an ionized photon and how it distributes its dose, this also deals a lot of harm to the healthy tissue.

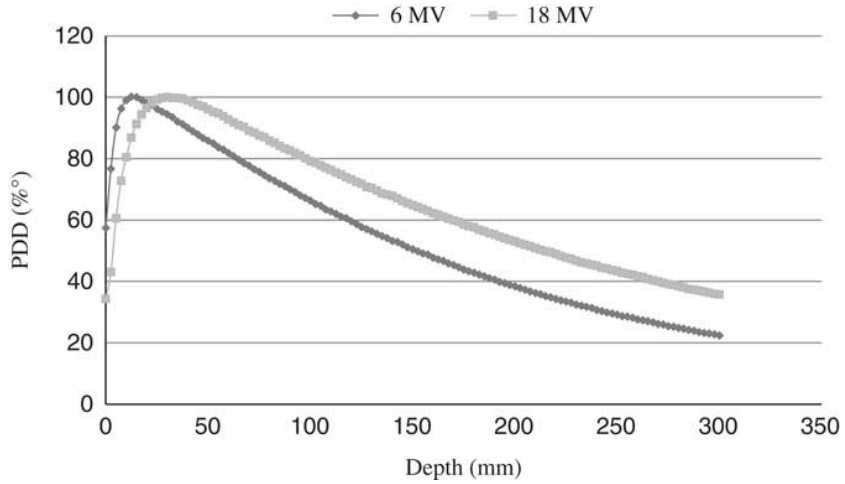


Figure 1: Typical Percentage Depth Dose (PDD) of a photon (Bencheikh, Maghnouj, and Tajmouati 2019).

Figure 1 shows a Percentage Depth Dose (PDD) curve of a photon, which shows that for tumors that lie deeper inside, the majority of the dose is deposited in healthy tissue, as only a percentage of the dose will be distributed at a certain target depth. This is where proton therapy comes in, as using protons over photons can solve this issue.

2.1.2 Proton Therapy

Proton therapy is a different modality for conventional external beam radiotherapy that uses ionized radiation in the form of protons as a radiation source. The major argument in favor of using protons when compared to photons, is their difference in dose distributed along the path in tissue. Figure 2 shows the PDD for a photon beam in comparison to proton beams.

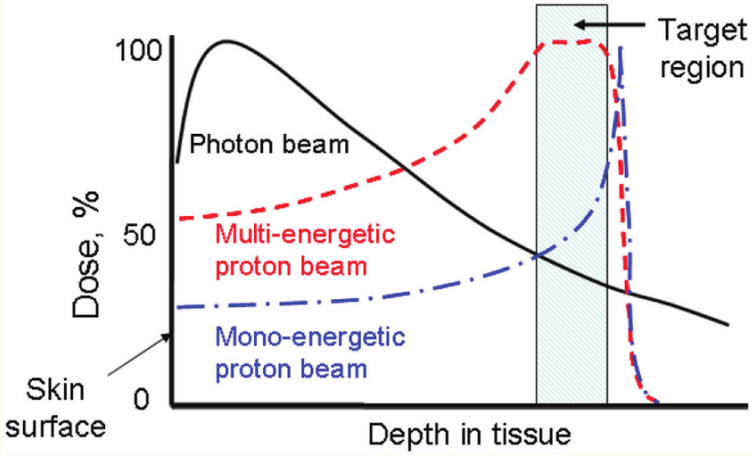


Figure 2: Percentage Depth Dose (PDD) of a photon beam, mono-energetic photon beam and multi-energetic proton beam respectively Azimi and Movafeghi 2016.

The PDD of a proton generally looks like the blue line. The photon distributes a large amount of dose at the entry, with the dose distributed slowly decaying over depth. The proton shows a completely different profile. It reaches its maximum in the form of a single peak, at a certain depth after which it rapidly decreases to zero. This peak is called the Bragg peak (BP), and its depth is dependent on the energy of the proton and the stopping power of the medium it traverses (Yuan, Zhan, and Qian 2019). The BP can be positioned throughout the target region by combining multiple proton beamlets with different energies within a single beam. Additionally, one may assign different intensities to different beamlets. This principle is called intensity modulation, hence why this type of radiotherapy is called Intensity Modulated Proton Therapy (IMPT).

Proton therapy can spare more healthy tissue than photon therapy in both the entry, and the distal side of the tumor. This raises one of the challenges that comes with proton therapy, which is the range error. As briefly mentioned earlier in this section, a proton's range depends on energy and stopping power of the medium it traverses (Yuan, Zhan, and Qian 2019). However, the stopping power depends on many factors, most prominently the type of material (Walters 2017). Since the human body consists of different materials, it can be difficult to precisely predict the range of the proton in a complex patient geometry. Additionally, there are other effects and phenomena, that may influence the trajectory of the proton (Walters 2017).

2.1.3 Treatment Planning Methods

Radiotherapy treatment planning follows a step by step process. This process starts with a consultation, where a radiation oncologist discusses the clinical situation with the patient. This discussion includes consideration of the risks and benefits of the treatment, the patient's goal of care, tumor stage, previous or current treatment received and much more (Mary Feng et al. 2018). After this consult, a treatment strategy is decided upon, which may include radiotherapy, occasionally along with other treatment modalities as mentioned in Section 2.1.1.

If radiotherapy treatment is decided upon, the first step of the treatment planning process is imaging. During this step, the physician wants to get an insight on the locations of the target volume and the organs at risk (OARs). This imaging is generally performed using Computed Tomography (CT). A CT scan makes a two-dimensional image for every few millimeters of a patient's body. In the computer, these images are put together, creating a three-dimensional array. This three-dimensional array consists of a unit we call Hounsfield Units (HU), which are defined as:

$$HU = \frac{\mu - \mu_{water}}{\mu_{water}} * 1000 \tag{1}$$

Once the imaging process is complete, the physician delineates all the important areas: the target volume and the organs at risk (OARs). The target volume can be split in three types: Gross Tumor Volume (GTV), Clinical Target Volume (CTV) and Planned Target Volume (PTV), each one containing the previous, as shown in Figure 3

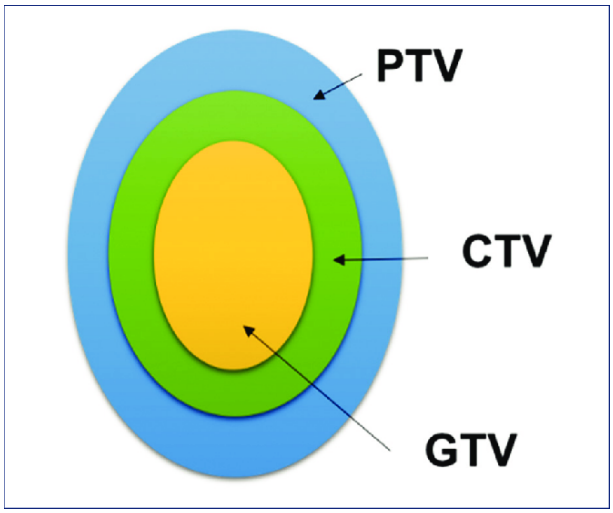


Figure 3: Gross Tumor Volume (GTV), Clinical Target Volume (CTV) and Planned Target Volume (PTV) and their relation to each other (Escobar-Peralta et al. 2022).

The GTV is the visible tumor. The CTV expands upon that, including margins for sub-clinical or invisible tumor tissue. The PTV contains the CTV, plus an extra margin for setup uncertainties. Generally, both the GTV ad CTV are indicated by the physician, though there are studies that focus on extracting the CTV from the GTV (Grégoire et al. 2018).

After the imaging has been performed, the dosimetric goals and constraints are indicated. The goal being sufficient irradiation of the target volume, constraints meaning: setting a maximum dose tolerance for the healthy tissue and OARs near the target volume. This data is combined in list, after which the creation and optimization of the treatment plan commences. This type of treatment planning is called multi-criteria decision making (Breedveld, Craft, et al. 2019).

One part of this process is the dose computation, where the dose engine creates a dose distribution for every spot location that is selected. From these dose distributions we can create a matrix that contains the normalized dose distributions for each beamlet in each area delineated by the goals and constraints. This dose matrix is called a dose influence matrix. An optimization algorithm will then assign values to each beamlet to determine its intensity, resulting in a dose distribution. This distribution is then reviewed manually and can be adjusted by tightening or relaxing some of the constraints (Breedveld, Craft, et al. 2019).

One of methods used in manual reviewing is the dose volume histogram (DVHS). This is a cumulative histogram that shows what percentage of certain body region has received a certain amount of dose. It generally contains the CTV, PTV and OARs and their respective cumulative absorbed dose. This is an easy way of checking if you ticked all the boxes of the list of goals and constraints, and shows an indication of the remaining margins on some OARs.

If the treatment plan is accepted by the physician, it goes through a quality assurance protocol, in which a more precise dose calculation is performed to ensure the plan stability. One of the checks is plan robustness for small deviations , since imaging and treatment do not happen simultaneously (Rojo-Santiago et al. 2021).

After this quality assurance, the patient is ready to be treated. Figure 4 shows the set up of a radiotherapy treatment session.

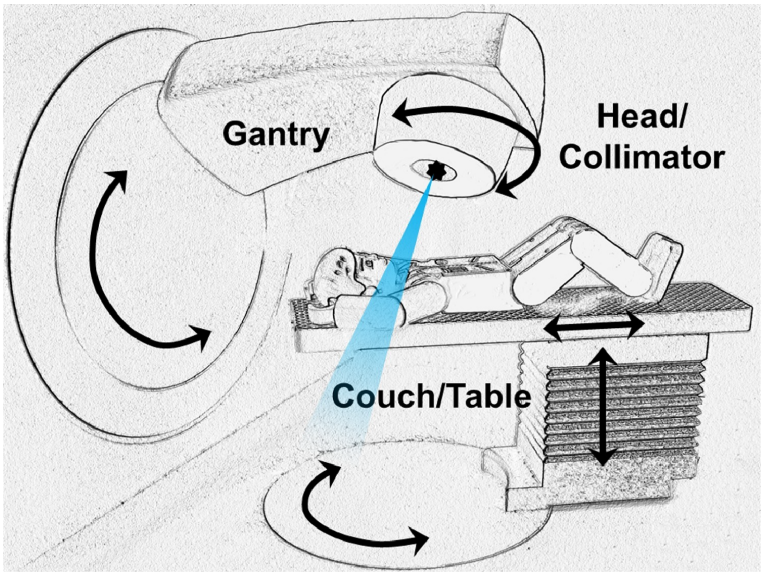


Figure 4: General overview of a radiotherapy machine. The patient lies on a height-adjustable couch and the gantry rotates around the patient. In this work, only beam directions resulting from the gantry rotations are taken into account (Breedveld, Craft, et al. 2019).

The patient is stabilized on the couch to restrict movement. The couch can rotate as well, though this is not taken into account in this work. As discussed in Section 2.1.1 radiotherapy does not only kill tumor cells, but also healthy ones. Therefore, the patient needs time to recover and cannot be irradiated with all the dose at once. Since healthy cells are generally better at recovering from radiation damage then tumor cells Royce, Qureshi, and Truong 2018, granting the patient some time to recover spares healthy cells over the duration of the treatment. This is called fractionation.

2.2 Dose Calculation Algorithms

2.2.1 Current Algorithms

The Dose calculation algorithm is the part of the TPS that computes the dose. It takes a CT image and some beam parameters as input, and produces a dose influence matrix, which contains the dose of every beamlet in every voxel. Besides accuracy, computation speed is an important aspect, since treatment plans generally require over a thousand beamlet calculations.

Proton dose computation algorithms are constantly in development. Most algorithms are based on either Monte Carlo simulation (MC) or pencil beam algorithms (PBA).

MC is a very precise computation method used for many applications in the fields of computational physics. It uses repeated random sampling to obtain numerical results. The underlying concept is the use of randomness to solve problems that are likely deterministic (Y. Lin, Wang, and Liu 2018). MC-based dose computation applies this by doing a simulation of the transport of a single particle inside the patient geometry repeatedly. For example, the MC-based simulations used to train DoTA consisted of 10^9. This transport simulation is based on the probability of each type of interaction the particle might experience in each type of matter inside the patient geometry. MC-based algorithms have been proven to be the most accurate dose computation algorithms, but their major drawback remains the computation load. The large amount of simulations required take

time, though MC-based algorithms have become quicker over the years mainly because of better hardware. However, for potential real-time adaptive treatment planning, where dose computations need to happen fast with high accuracy to adjust for anatomical changes during treatment, faster algorithms with similar precision are required (Kontaxis et al. 2017).

PBA is a faster alternative, but sacrifices accuracy compared to MC-based algorithms (Teoh et al. 2020). It assumes that each beam consists of many, narrow pencil beams, each of which deposits dose along its central axis. The deposited dose of this pencil beam depends on the intensity. Patient inhomogeneities are accounted for by a little shape correction, based on the density of the tissue. The dose distributions of all these narrow pencil beams are added up to compute the entire beamlet dose distribution (Elcim, Dirican, and Yavas 2018). Especially in complex patient geometries - areas with large tissue inhomgeneities, PBA suffers relatively high errors (Teoh et al. 2020).

To improve both quality and calculation time of proton treatment planning, several studies have been performed. Some focus on the improvement of computational methods, like GPU-accelerated MC. To speed up computations times, a graphics processing unit (GPU) is employed in parallel with the computer processing units (CPU) (H. Feng et al. 2022). Another part of research focuses on the implementation of deep learning (Javaid et al. 2021)(Nomura et al. 2020)(Wu et al. 2021)(Pastor-Serrano and Perkó 2022).

2.2.2 Deep Learning Algorithms

Deep learning is a type of Machine Learning (ML) that is starting to play a more prominent role in research over the last few years (Levine et al. 2019). One of the main reasons for this being the increase of available computing power. Machine learning works on the principle that based on training data, an algorithm develops an understanding of the data. This makes it capable of doing predictions or infer conclusions based on that data, without manual programming instructions. ML's automated nature makes it very interesting in research fields that rely on recognition and grouping of patterns. ML can be split into two groups, supervised learning and unsupervised learning. An overview is shown in Figure 5.

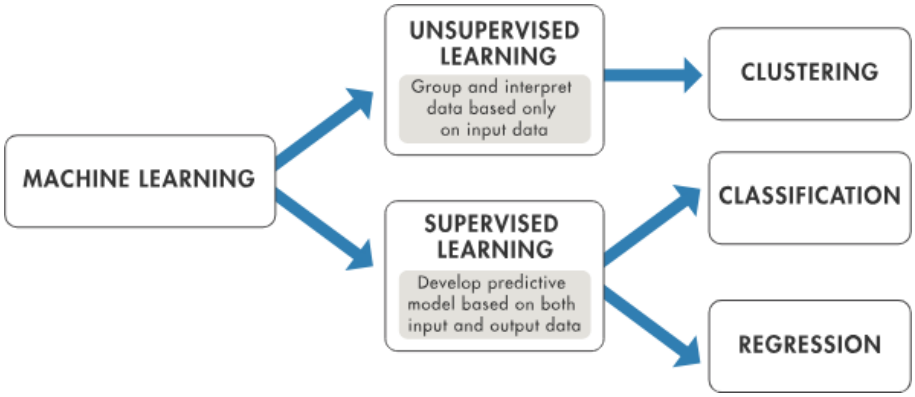


Figure 5: Schematic overview how ML can be split into two groups, supervised learning and unsupervised learning (The MathWorks Inc. 2022).

In unsupervised learning, the model tries to group data points based on patterns or similarities. This is called clustering. In supervised learning, the model knows both the input data and the corresponding output data (label). Therefore it tries to find patterns between input and output, such

that learns to predict an output when presented with an input of which the output is unknown. The desired technique for supervised learning can be split in two categories, classification and regression. Classification being a discrete response, like yes or no, used for example in cyber-security to determine whether an email is spam. Regression is a continuous response, used to determine hard to measure quantities (Rajoub 2020). ML algorithms work generally on relatively simple datasets. If a more sophisticated understanding of the data structure is required, Deep learning is generally required.

A deep learning model is basically a multi-layered ML model, where instead of mapping between directly between one input and output, the information in each layer is extracted from the output of the previous layer. An example of a such an multi-layered ML model is the object recognition feature on a phone, where the first layer identifies the edges, then a second layer identifies the corners, etcetera. The layers between the input and output layers are called hidden layers.

Typical deep learning models are artificial neural networks. These neural networks serve as a simulation of how neurons would signal inside our brains. A neural network generally tries to approximate a function $f^*(x)$ that defines the mapping between input x and label y. By training, the network tries to find a mapping $y = f(x; \theta)$ and learn the values of the parameters θ (Goodfellow, Bengio, and Courville 2016). There are several types of neural network configurations, the most straightforward being a feed-forward neural network, which is displayed in Figure 6.

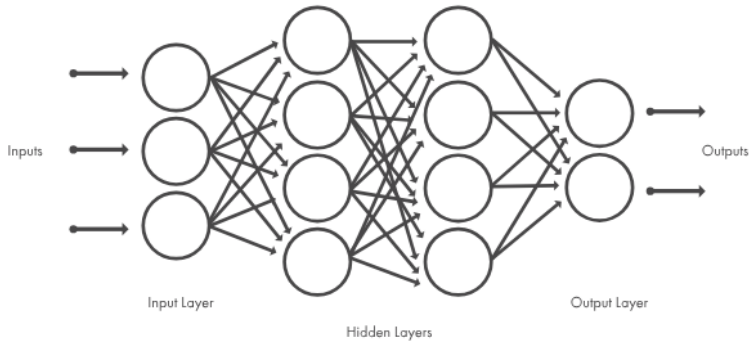


Figure 6: Schematic overview of a neural network (The MathWorks Inc. 2022).

In a feed-forward neural network there is no looping, or going back in layers. The nodes in each layer use the outputs of all nodes in the previous layer as inputs, such that all neurons interconnect with each other through the different layers (The MathWorks Inc. 2022). Some neural networks, besides the conventional layer-to-layer connection, also contain connections that skip between layers. These are called residual connections, which are used to speed up the training process.

Convolutional neural networks (CNN) are an advanced type neural network, capable of processing grid data. As described earlier in this Section, neural networks define the mapping between input and output. For a standard neural network this basically means that for a layer with m inputs and n outputs, the interaction of each layer is described by a matrix multiplication by a parameter matrix of size $m \times n$ (Goodfellow, Bengio, and Courville 2016). In a CNN, the amount of inputs is fixed to a fixed number k instead. The matrix containing the corresponding k parameters is called the kernel. The same kernel is used for each of the n output nodes, hence they all use a shared matrix of parameters, which reduces memory load and computation time. Figure 7 shows such a convolution. Another advantage is that thanks to the parameter sharing, the network learns to detect one relation between all nodes, instead of learning every relation between the input and output.

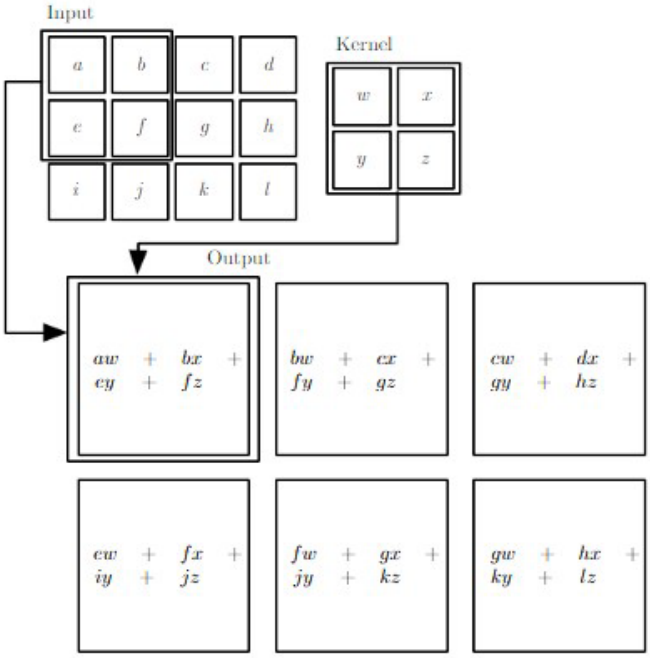


Figure 7: Schematic overview of 2D convolution used in a convolutional neural network (Goodfellow, Bengio, and Courville 2016).

A CNN generally consists blocks each containing the same sequential layers. First, a convolutional layer, performing the aforementioned convolution operation. Secondly, a normalization layer to normalize the processed data. Often this layer is optional. Thirdly, a pooling layer, which groups very similar output data into one single output node. This means that very similar outcomes will be treated as the same outcome to account for small translation errors. Finally, an activation layer may be included. This layer applies an often non-linear activation function to the processed data, to account for non-linear approximation.

A more recent development in deep learning is the transformer model. This is a type model that is able to recognize which data is more important and requires more attention (Vaswani et al. 2017). This is implemented in the form of a Self-Attention (SA) layer. The principle works based on attention units, with dedicated parameters called attention weights.

A Self Attention function can be described as mapping a query and a set of key-value pairs to an output. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a compatibility function of the query with the corresponding key (Vaswani et al. 2017).

Multi-head Self Attention (MSA) is a variant of SA, that works with multiple sets of attention weight, called attention heads. Let N_h be the number of attention heads, that means N_h SA operation are computed in parallel, each capable of detecting different self-attention patterns. A transformer model typically combines this SA layer with a set of feed-forward neural networks with an encoder and a decoder (Vaswani et al. 2017; Valk Bouman 2022). The data is first processed through the encoder layers, with the attention units capturing information the relations between the

several parts of the data. The decoder layers then combine the output and contextual information to generate an output sequence.

Deep learning plays a big role in medical science related research. Some examples: studies show potential deep learning applications in tumor detection (Hamed et al. 2020; Sajid, Hussain, and Sarwar 2019), target volume segmentation (H. Lin et al. 2021), and photon therapy treatment planning (Fan et al. 2019). Focusing on deep learning based research in proton therapy treatment, most focus on the dose computation algorithm. Some examples are MC denoising, where faster, noisy MC dose computations are denoised using deep learning (Javaid et al. 2021) and training PBA to resemble MC based dose computations. (Wu et al. 2021). However, none of these are generally applicable to any non-specific case and therefore do not qualify as a generic dose algorithm (Pastor-Serrano and Perkó 2022). A deep learning algorithm that does qualify as a generic dose algorithm, is the Dose Transformation Algorithm (DoTA) by Pastor-Serrano and Perkó 2022.

2.2.3 Dose Transformation Algorithm (DoTA)

DoTA is a deep learning based algorithm that is capable of computing dose distributions with MC accuracy with computation times faster than current PBA (Pastor-Serrano and Perkó 2022). DoTA is trained to map an input consisting of a CT image and a beamlet energy, to an output in the form of a dose distribution.

In terms of the neural networks described in Section 2.2.2, it tries to learn the function $y = f_{\boldsymbol{\theta}}(\mathbf{x}; \epsilon)$ via a series of artificial neural networks with parameters $\boldsymbol{\theta}$. More specifically, it learn a mapping between a three-dimensional voxelgrid $\mathbf{x} \in \mathbb{R}^{L \times W \times H}$ to a dose distribution $\mathbf{y} \in \mathbb{R}^{L \times W \times H}$ conditioned on the energy $\epsilon \in \mathbb{R}^{+}$ (Pastor-Serrano and Perkó 2022). The voxelgrid used has a fixed resolution of 2 mm $\times$ 2 mm $\times$ 2 mm and the dimensions of both the input voxelgrid and output dose distribution are $L \times W \times H = 150 \times 24 \times 24$ voxels.

The neural network of the DoTA model consists of three parts, an overview of which is shown in Figure 8.

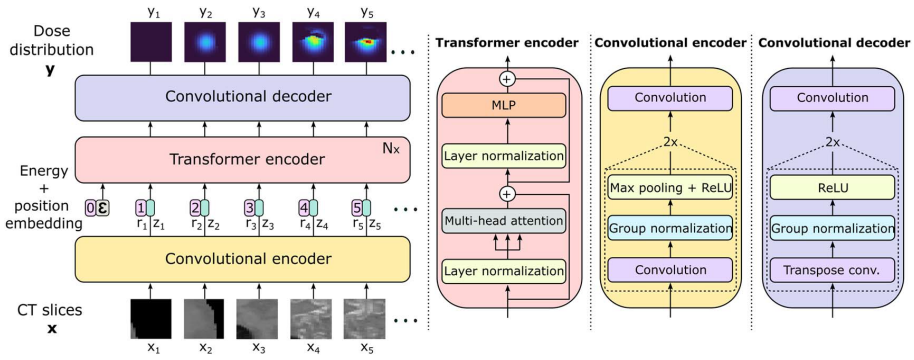


Figure 8: Overview of the DoTA model architecture (Pastor-Serrano and Perkó 2022).

The input is treated as L two-dimensional slices, after which it is passed to the convolutional layer that extracts the important geometrics and turns them into a vector, after which the energy data is added at the beginning of the resulting sequence. A transformer encoder than combines information of the different elements of the sequence. Finally the convolutional decoder transforms the vectors back to two-dimensional slices, which can be combined to a three-dimensional dose distribution.

The range of energies on which DoTA is trained is 70 to 220 MeV, which covers a majority of the proton energy spectrum used in clinical practice (Mohan and Grosshans 2017). The maximum depth in can reach is 150 voxels, which corresponds to 30 cm. These are small limitations that only play a role in very particular cases. More prominent limitations include the assumption of perpendicular beams and the lack of a range shifter. In clinical practice, beams more often than not enter the patient under a slight angle, hence a correction for that is required. A range shifter is a small plate of material that reduces the range of the proton in order to make high energy protons usable in low depth patient geometries, like the neck. The lack of a range shifter limits the types of treatment DoTA can potentially be used.

3 Methodology

In this section, the methodology of the project will be introduced, giving an overview and explanation of the algorithm and the experiments and analyses that have been done during this work.

3.1 Algorithm

The dose computation algorithm is based on the deep learning dose engine DoTA Pastor-Serrano and Perkó 2022. Besides DoTA, the algorithm has a pipeline that processes the inputs from Erasmus iCycle (Breedveld, Storchi, et al. 2012) and perform some corrections and changes to the CT data to enable realistic dose computations.

One of the corrections in particular that need to be performed is a rotation to account for the beamlet angles. As mentioned in Section 2.2.3, DoTA is can only compute doses that enter the geometry perpendicular. Figure 9 shows why the assumption of perpendicular beamlets is not feasible and why beamlet angles need to be accounted for.

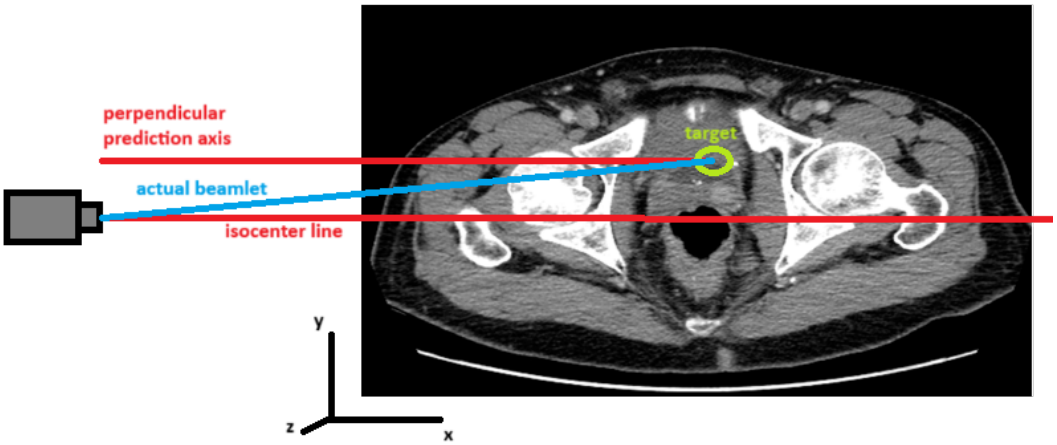


Figure 9: Example of a spot located relatively far from the isocenter line and the potential error emerging when assuming a perpendicular entry

Note that in reality, this beam source is located roughly 2 meters away. Figure 9 only shows the beamlet angle in the same plane as the gantry angle (xy-plane rotation), but the same correction has to be applied in the xz-plane, because the CT represents a 3D volume. The target region might be located in a different slice than the isocenter, in which case the beamlet has to travel 'into the paper' in the case of Figure 9. Since the locations of both the isocenter and the target spots are known, the beamlet angles can be computed by:

$$\alpha = arctan\left(\frac{dy}{SAD + dx}\right) \tag{2}$$

$$\beta = arctan\left(\frac{dz}{SAD + dx}\right) \tag{3}$$

Where α and β are the beamlet angles in the xy and xz-plane respectively, dx, dy, dz the distances to the isocenter in mm and SAD the Source Axis Distance, which is the distance from the source to the isocenter.

A simplified full overview is given of the pipeline, which starts by collecting the necessary data from Erasmus iCycle, and outputs a dose distribution or normalized dose matrix:

- Convert Erasmus iCycle spots (B, E, dx, dy) to (B, E, x, y, z) using proton ray-tracing, where B is the beam index, E the beamlet energy, dx, dy the x and y distances to the isocenter respectively, and x, y, z the actual coordinates of the spot.
- Load the CT to the python environment and apply the correct translations occurring from convention differences between MATLAB and python (axis inversions, different bases etc.).
- Apply a calibration scaling to the input energy (more on this in Section 3.3)
- Load in the transformer with appropriate weights, prepare DoTA.
- Shift the CT towards the isocenter, scale the voxels to the 2 mm $\times$ 2 mm $\times$ 2 mm grid and rotate the CT towards the desired gantry angle for each beam and apply the same shifts, scaling and rotation to the original spot coordinates.
- For each spot compute the beamlet angle based on Formula 2 and 3, apply an additional rotation and interpolation for this beamlet angle and slice a (150,24,24) sized array around it (corresponding to DoTA's desired input).
- Compute the dose distribution with DoTA.
- Assign the dose array in the original full CT grid and apply the inverse of all the previous translations to regain the dose of each beamlet in the original CT coordinate system.
- Assign the dose to the original spot, creating a (B, N, Dx, Dy, Dz) matrix where B is the beam index, N the spot index and Dx, Dy, Dz the dose in each dimension.
- Either sum them all up to a full dose distribution of the form (Dx, Dy, Dz), or normalize for each beamlet to get a starting point for a dose influence matrix

Since we need to integrate the beamlet angle, a separate interpolation is required for each beamlet. In the case of realistic treatment plans with over a thousand spots, this means an equal amount of additional matrix computations is required.

3.2 Erasmus iCycle & ASTROID

Erasmus iCycle is the TPS used in this project, which is an algorithm for integrated, multi-criteria optimization of beam-angles and intensity modulated radiotherapy (IMRT). Erasmus iCycle is an automated TPS. This means that it performs the spot selection, dose computation and initial optimization without manual interference (Breedveld, Storchi, et al. 2012).

This algorithm works using advanced multi-criteria decision making. In advanced multi-criteria decision making, the TPS sets a list of conditions and constraints, combined with a certain ascribed priority. The higher this priority, the higher the chance that this condition will be met (Breedveld, Storchi, et al. 2012).

The dose engine currently used by Erasmus iCycle is ASTROID. This dose engine is based on PBA. It is based on the implementation of pencil-beam scanning (PBS), a technology whereby a

focused beam of protons, of variable intensity and energy, is scanned over a plane perpendicular to the beam axis and in depth (Kooy et al. 2010). During the experiments, DoTA will be compared to this dose engine.

3.3 Calibration

Since DoTA and ASTROID are both based on different beam models, some sort of calibration was required to make a good comparison between the two. Figure 12 shows the difference in center line dose distributions between DoTA and ASTROID before calibration for a single 175 MeV beamlet.

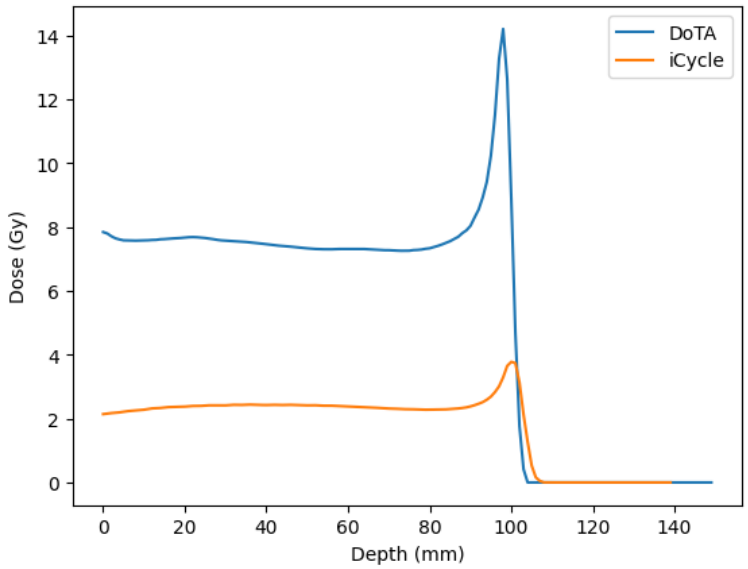


Figure 10: Difference between DoTA and Erasmus-iCycle ASTROID center line dose for a 175 MeV beam in a perfect water box.

There is a big difference between the two models in both input and output. ASTROID works based on Monitory Units (MU). This is a parameter in radiation therapy that determines the amount of radiation to be delivered to the patient during treatment (Ma et al. 2004). Depending on the MU used to do the calculation, it outputs a dose distribution in Gy. DoTA works based on a dose per particle scale. Since it is trained on a set beam model with a sample size of 10^9 particles (Pastor-Serrano and Perkó 2022), it computes the dose per 10^9 particles. However, the amount of particles in a beamlet is not known, as Erasmus iCycle works with a different type of intensity scaler.

The goal of this calibration is to relate the MU's used by Erasmus-iCycle to DoTA. To this end, 23 single spot dose distributions are predicted with both DoTA and ASTROID, for different energies between 80 MeV and 205 MeV in a 'perfect' water tank (homogeneous field of 0 HU), at a perpendicular entry. The depth of the BP is then determined for each beamlet. Subsequently, for all 23 energies, a manual, iterative process of modifying the energy input of the DoTA model and scaling the dose distributions towards each other based on the dose in the BPs is performed. This process obtains the difference in energy required by DoTA to reach equal BP depth (ΔE) and a scaling factor to link the MU's from Erasmus-iCycle's ASTROID and DoTA's particle stream.

By accumulating the appropriate and dose scaling for each energy, an interpolation function can be constructed to determine these scaling factors for all energies and beam intensities. Note that, unless stated otherwise, all future experiments are performed with this calibration to be able to make sensible, direct comparisons.

3.4 Hounsfield Unit Analysis

After obtaining a calibration in water, an analysis is performed in different HUs. As the beams did not fully overlap correctly in water, a full investigation in a variety of Hounsfield Units is warranted.

For this experiment, 3 beams under perpendicular conditions were tested in a range of HUs. The range (-250,1000 HU) was chosen to correspond to the HUs of most prevalent tissues in patients. The investigated beams were chosen to be 80 MeV, 140 MeV and 200 MeV, each representing a part of the spectrum of proton therapy beams. Clinical proton energies typically vary between 70 and 250 MeV (Mohan and Grosshans 2017). However, DoTA is trained for proton energies in the range of 70 to 220 MeV (Pastor-Serrano and Perkó 2022) 2.2.3. For the different HUs, a homogeneous box of a set amount of HUs was created as CT input for both models and the corresponding dose distributions were computed. The locations of the BPs from each dose distribution were determined. The relative discrepancy between the paired BP's is defined as:

$$\Delta x = \frac{x_{DoTA} - x_{iCycle}}{x_{DoTA}} \tag{4}$$

Where x_{DoTA} and x_{iCycle} are the depths of the DoTA and Erasmus-iCycle's ASTROID predicted BPs respectively.

3.5 Water Phantom Experiment

After calibrating DoTA in water, its overall quality and comparability in water was investigated for different energies, gantry angles and beamlet angles.

First, an analysis is performed on the mean and maximum relative absolute discrepancies of the beamlets for 7 different energies in a range of 80 MeV to 200 MeV, under gantry rotations in a range of 0° to 180° in 30° increments. The metrics used to determine the mean and maximum relative absolute discrepancies are described in Formula 5 and 6.

$$\Delta D_{mean} = \sum_{(i)} \left| \frac{D_{DoTA}[i] - D_{ASTROID}[i]}{D_{DoTA}[i] * N} \right| * 100\% \tag{5}$$

ΔD_{mean} being the mean discrepancy and D_{DoTA} and $D_{ASTROID}$ refer to the dose distributions produced by DoTA and ASTROID respectively, in each voxel (i). N represents the total number of summed values, over which is divided to obtain the mean.

$$\Delta D_{max} = max \left| \frac{D_{DoTA} - D_{ASTROID}}{D_{DoTA}} \right| * 100\% \tag{6}$$

In Formula 6 the maximum value of the relative absolute difference is computed, obtaining the maximum discrepancy ΔD_{max}.

In addition to the 30° increment analysis on the influence of the gantry angle, an analysis is performed on a 0° to 90° interval with 10° increments to quantify the influences of gantry rotation on the models.

Moreover, an analysis on the mean discrepancy has been performed on the dose distributions under a variety and combination of beamlet angles. In clinical practice, the protons and therefore the resulting proton beamlets are steered towards the head of the beam source by steering magnets. Rather than perpendicular, these beamlets may exit the nozzle and enter the patient under a variety of angles, roughly in the interval of [-5.16°,5.16°] (Valk Bouman 2022). In this case, an interval of [-3.0°,3.0°] was more realistic, as larger angles would miss the sample. Positive and negative angles in either plane were investigated for a perpendicular incident beam.

3.6 Patient Geometry Experiment

To further test the feasibility of DoTA as a dose engine in Erasmus-iCycle, an analysis is needed on the comparability to ASTROID inside a patient geometry.

The patient geometry used in this experiment is a prostate cancer patient. This patient has a sufficient variety of different HUs. As soft tissue is present as well as some bone tissue due to the presence of the and the pelvic bones. Furthermore, the total CT volume is relatively large, meaning that available energies have BPs inside the patient and no range shifter is required. Figure 11 shows a prostate CT and a collection of beamlets arriving from different gantry angles.

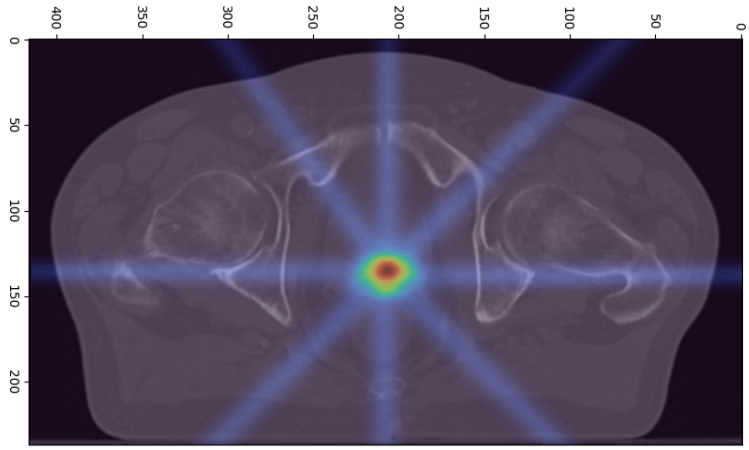


Figure 11: A prostate patient geometry with some example beamlets under 45° gantry angle increments.

First, a similar experiment on the mean absolute discrepancy as in the water sample has been performed. Again, for 7 energies in the range of 80 MeV to 200 MeV, under gantry rotations of 0° to 180° in 30° increments, the mean discrepancy between the models has been computed using Formula 5. Secondly, again similar to the water phantom experiment, an analysis on the mean absolute discrepancy has been performed under a set of beamlet angles in the interval of [-3.0°,3.0°] in each direction, for a perpendicular beam.

Thirdly, because the dose computation models in this work, DoTA and ASTROID, are inherently different, we want to analyze another metric on which the two can be compared. An additional test has been performed on the energy deposition, rather than dose. Based on the principle of conservation of energy, the total amount of energy deposited in the patient's body should be close.

Since absorbed dose in radiotherapy is measured in Gy, a link can be made between dose and energy. The unit Gy is actually defined as energy per mass, $[J/kg]$. Therefore, converting the CT to a mass density field and then eliminating the volume - which is just a multiplication with the size of the corresponding voxel - gives the distribution of energy inside the patient. Hounsfield Units are generally a function of linear attenuation coefficients, as shown in Formula 1 back in section 2.1.3. However, using the curve in Figure 12, a link between Hounsfield Units and mass density can be made.

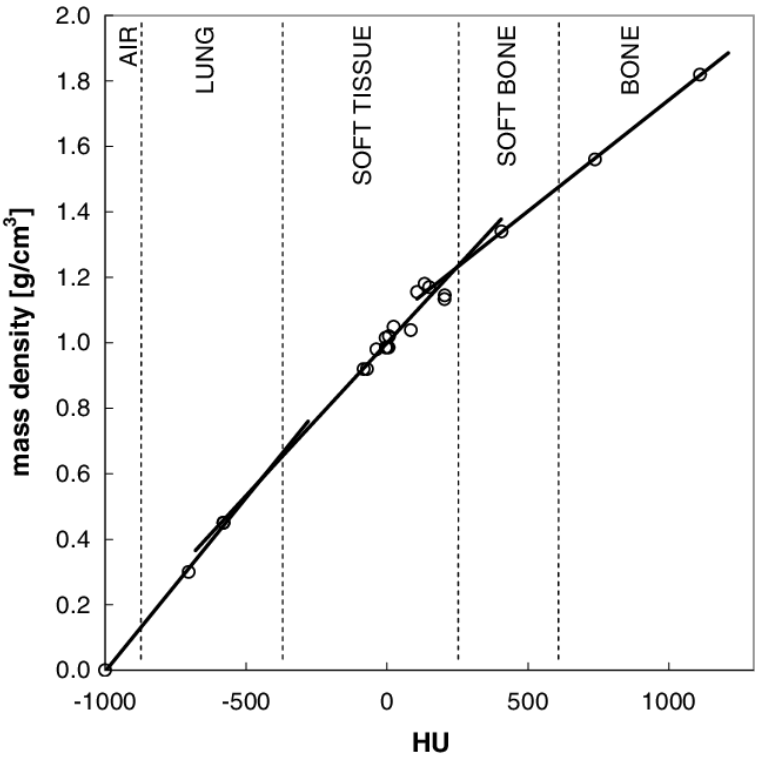


Figure 12: A (HU,ρ) calibration curve used for conversion of HU into mass densities (Bazalova et al. 2008).

Using the data from Figure 12, three linear fits are used to convert the CT image to a mass density field, from which a mass density field $M[kg/cm^3]$ can be created. The energy distribution E is now obtained via Formula 7

$$E = M * D * V_{voxel} \tag{7}$$

Where E is the obtained Energy distribution, M the mass density field, V_{voxel} the voxel volume.

Lastly, a comparison on a full dose distribution between the two models has been made. The doses have been compared by their respective DVHS. Generally, all the organs are visible in the dose

value histogram, but in this case only the CTV has been plotted, as applying the correct dose of the tumor volume is the first thing to get right.

A brief analysis has been performed on computation time. DoTA has proven itself to be quicker than PBA, which is the type of algorithm ASTROID is based on (Pastor-Serrano and Perkó 2022) (Valk Bouman 2022) (Kooy et al. 2010). However, as explained in Section 3.1, there are a lot of additional computations required in order to predict the doses. For ASTROID, only the total computation time was known. For DoTA, several timings could be recorded, therefore the computation time can be more thoroughly analyzed. This should give an insight on how the two different models compare in terms of overall computation time.

Note that these computations where performed on a node of a high computing cluster, consisting of 18 CPU cores and 128GB RAM. In actual large treatment plan computations, there are sophisticated methods to sometimes speed up computations. This was not taken into account, as a direct comparison per dose calculation would already give an insight on which method is faster in general.

4 Results

This section contains the results of each of the experiments described in Section 3. In addition to that, a short interpretation and comparison of some of the data displayed in the figures is made.

4.1 Beam Calibration In Water

The initial discrepancy between the depth of the BPs from DoTA and Erasmus-iCycle beam calculations in water is displayed in Figure 13.

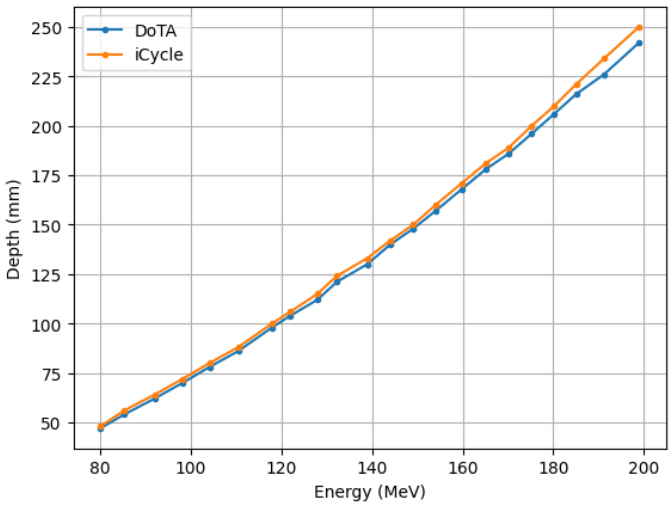


Figure 13: Initial difference in the depth of the Bragg peaks predicted by DoTA and Erasmus iCycle's ASTROID in water.

Notice that the BP predicted by Erasmus iCycle's ASTROID lies systematically deeper than the one predicted by DoTA. Moreover, figure 13 shows that there are two areas where the discrepancies between the BP depths are largest. For most energies the depth discrepancy is about 1 to 2 mm, but for energies between 120 and 140 MeV the difference becomes up to 5 mm. Moreover, for energies above 170 MeV the discrepancy rapidly increases in size, eventually up to 8 mm at 200 MeV.

4.1.1 Energy Scaling

An energy calibration was performed on a set of 23 energies as described in Section 3.3. The range of a proton can be extended by increasing its energy to overcome the discrepancy in range for the water samples displayed in Figure 13. The results of the calibration in energy are displayed in Figure 14.

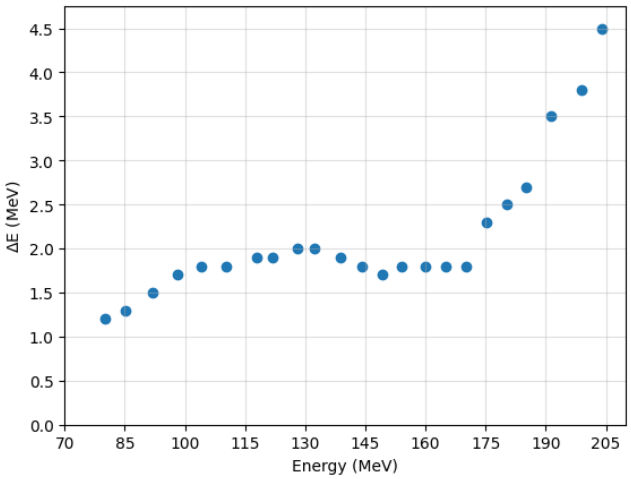


Figure 14: Calibration results of the depth of the Bragg peak of the center line dose.

In Figure 14, the x-axis represents the initial energy - the energy used by ASTROID and the y-axis contains the difference in energy ΔE required to be added to DoTA in order to reach the same depth for the center line Bragg Peak in water. In Figure 14, ΔE seems to stabilise around 130 MeV and it increases after 170 MeV. These findings correspond with the depth discrepancy displayed in Figure 13, since these are the areas where the depth differences were the largest.

4.1.2 Dose Scaling

The particle stream is unknown in Erasmus-iCycle (Section 3.3). Therefore the spot intensity needs to be linked to the particle stream required by DoTA via a separate calibration. The results of this calibration are shown in Figure 15.

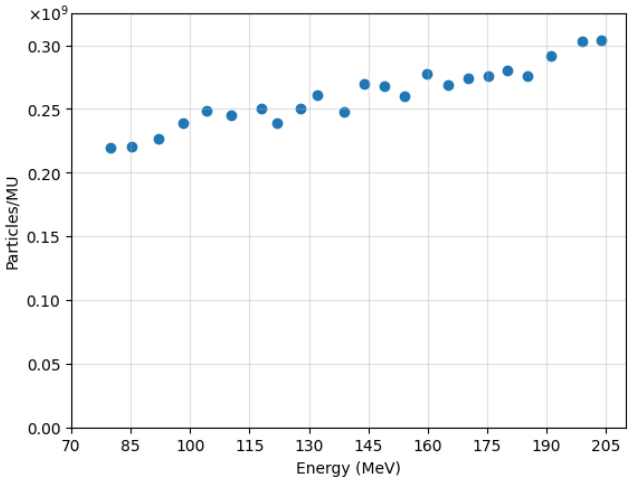


Figure 15: Calibration of dose scaling factors between DoTA and Erasmus-iCycle's ASTROID (i.e. it connects DoTA's output ($Gy/10^9$) to ASTROID output (Gy) for 1 MU).

The original energy (i.e. the energy from Erasmus-iCycle) of the beams is represented on the x-axis, and the scale is represented on the y-axis, with a unit of number of particles per MU.

4.2 Beam Comparison

With DoTA calibrated in water, qualitative comparison between DoTA and Erasmus-iCycle ASTROID can be made.

4.2.1 Hounsfield Unit Analysis

An analysis is performed on a range of HUs that are common in a patient for three different energies: a low energy (80 MeV), a moderate energy (140 MeV) and a high energy (200 MeV). The findings are displayed in Table 1. The relative discrepancy has been plotted in Figure 16.

Table 1: The depth of the Bragg peak of a 80, 140 and 200 MeV beam respectively in a variety of homogeneous Hounsfield Units fields for both DoTA and Erasmus iCycle's ASTROID generated beams and their discrepancy.

	DoTA (mm)	iCycle (mm)	Discrepancy
-250 HU	60	64	-6.7%
-100 HU	54	56	-3.7%
0 HU	50	50	0.0%
250 HU	45	44	2.2%
500 HU	41	40	2.4%
750 HU	38	37	2.6%
1000 HU	37	33	10.8%

(a) 80 MeV beam

	DoTA (mm)	iCycle (mm)	Discrepancy
-250 HU	158	170	-7.6%
-100 HU	146	152	-4.1%
0 HU	136	136	0.0%
250 HU	120	119	0.8%
500 HU	112	108	3.5%
750 HU	106	100	5.7%
1000 HU	100	88	12.0%

(b) 140 MeV beam

	DoTA (mm)	iCycle (mm)	Discrepancy
-100 HU	272	284	-4.4%
0 HU	252	252	0.0%
250 HU	226	225	0.4%
500 HU	212	201	5.2%
750 HU	202	182	9.9%
1000 HU	194	165	14.9%

(c) 200 MeV beam

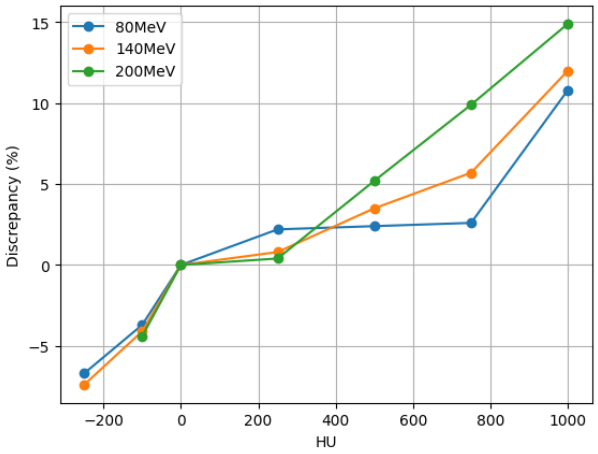


Figure 16: Discrepancy of the BP depth between the DoTA and ASTROID dose distribution, displayed as a percentage of the depth of the DoTA BP.

Figure 16 plots the difference between the BP depths predicted by the two models. We see that for 1000 HU, range discrepancies occur between 10.8 and 14.9%. Note that a positive percentage refers to the DoTA beamlet getting deeper than the ASTROID beamlet and a negative percentage means the exact opposite.

4.2.2 Water Phantom

Analyzing where the majority of the difference in dose distributions between DoTA and Erasmus-iCycle's ASTROID come from, the models have been examined and compared in a water box. Figures 17 and 18 show the mean and maximum absolute relative discrepancy between the two dose distributions in water and their dependence on energy.

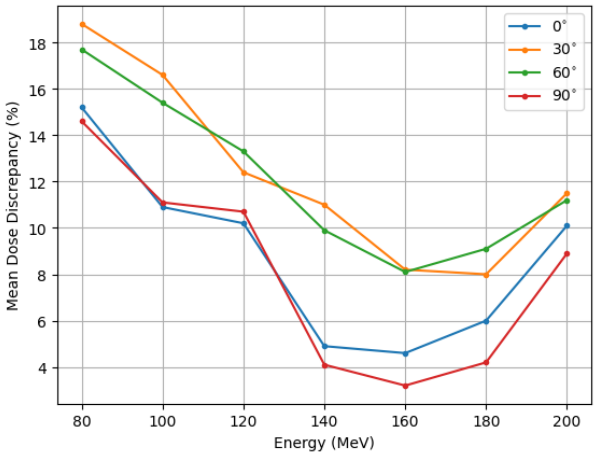


Figure 17: Mean relative absolute discrepancy between the DoTA and ASTROID dose distributions in water as a function of energy for different gantry angles.

Figure 17 shows relative absolute mean dose discrepancies of up to 15.2% for perpendicular beamlets, and up to 18.9% for non-perpendicular ones. The maximum discrepancy shows similar results for the energy dependence as the mean discrepancy. The results are displayed in Figure 18. Unlike the mean relative absolute discrepancy, the maximum absolute discrepancy does not vary much with the gantry angle.

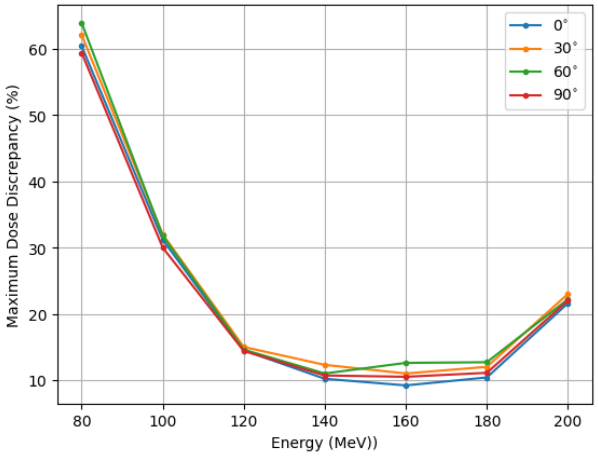
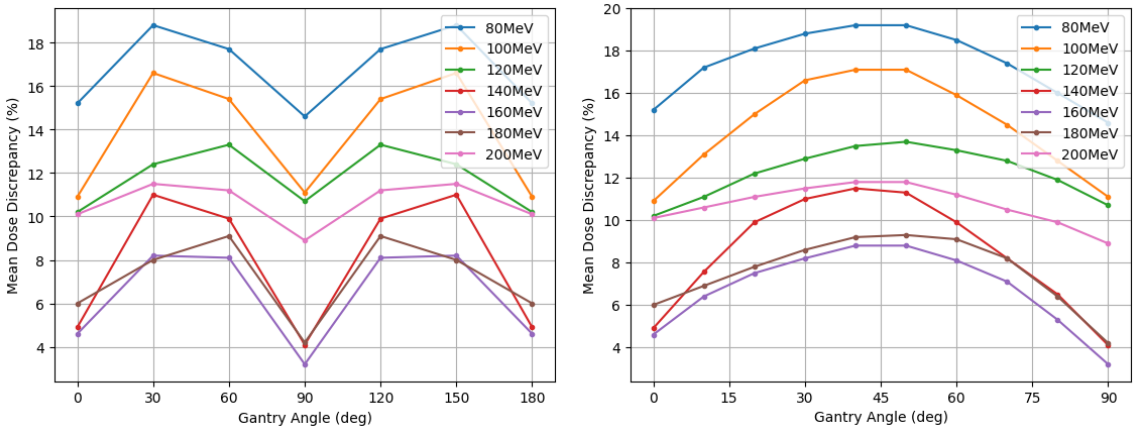


Figure 18: Maximum relative absolute discrepancy between the DoTA and ASTROID dose distributions in water around the BP as a function of energy for different gantry angles.

Figure 18 does show maximum relative absolute dose discrepancies up to 64.8% for 80 MeV beamlets. Figure 19 shows the dependence on the gantry angle of the mean relative absolute dose discrepancy.



(a) [0°,180°] gantry angle, 30° increments.

(b) [0°,90°] gantry angle, 10° increments.

Figure 19: Mean relative absolute dose discrepancy between the DoTA and ASTROID dose distributions in water as a function of the gantry angle for different energies.

Figure 19a shows symmetry around the 90° gantry angle, which is expected in a homogeneous water box. Moreover, the mean discrepancy increases for non-perpendicular gantry angles. Figure 19b shows that all curves reach their maxima between 40° and 50°.

Figure 20 shows the gantry angle dependence of the maximum relative absolute dose discrepancy.

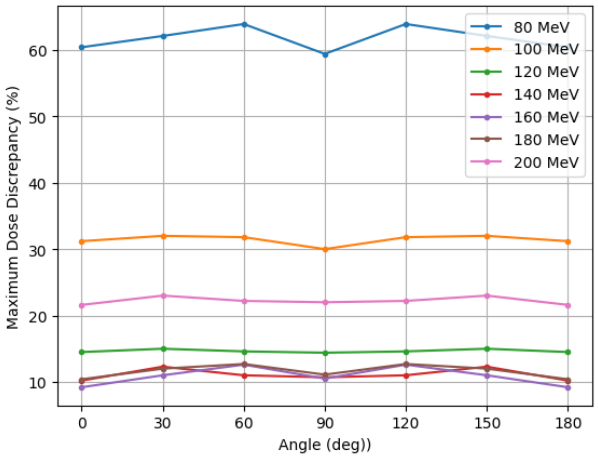
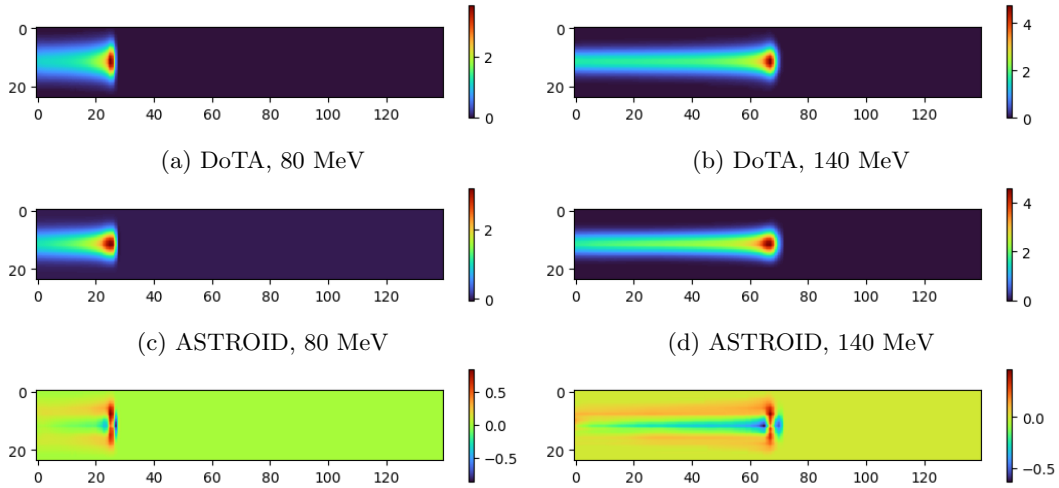


Figure 20: Maximum relative absolute discrepancy between the DoTA and ASTROID dose distributions in water around the BP as a function of gantry angle for different energies.

Though the mean relative absolute dose discrepancy showed a clear dependence on the gantry angle, the maximum relative absolute dose discrepancy seems to have very little to no gantry angle dependence at all.



(a) DoTA, 80 MeV

(b) DoTA, 140 MeV

(c) ASTROID, 80 MeV

(d) ASTROID, 140 MeV

(e) Difference, 80 MeV

(f) Difference, 140 MeV

Figure 21: Dose distributions of a 80 MeV and 140 MeV beamlet dose in water by DoTA and ASTROID and their differences.

Figure 21 shows the single spot dose distributions of a 80 MeV and 140 MeV beamlet by DoTA, ASTROID, and their difference. This gives a more visual representation of the dose discrepancy. Notice how for both the 80 MeV and the 140 MeV beam, the widths of the beams and their BPs are different. In both cases, DoTA seems to distribute the dose over a wider area. ASTROID has a more narrowly distributed dose, resulting in more dose in the center line. Especially the BP seems to be 'rounder' for ASTROID and flatter for DoTA.

The mean discrepancies for three beamlets under a variety of different beamlet angles and combinations of beamlet angles are shown in Figure 22.

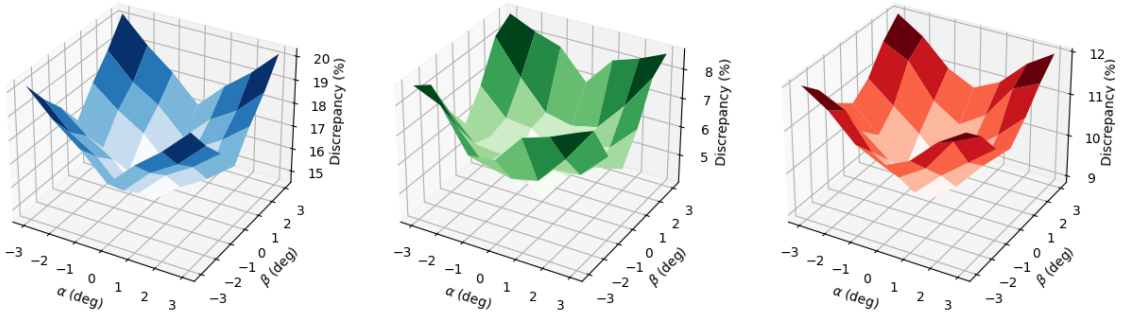


Figure 22: Mean relative absolute discrepancy between the DoTA and ASTROID dose distributions in water as a function of the beamlet angle for a 80 MeV (Blue), 140 MeV (Green) and 200 MeV (Red) beamlet respectively.

Like in Figure 19, the symmetry is present for rotations in both the xy-plane or the xz-plane. Moreover, Figure 22 shows higher mean relative absolute discrepancies for combined gantry angles, than single ones.

4.2.3 Patient Geometry

Similar analyses as in water have been performed in a patient. Figure 23 shows the mean relative absolute dose discrepancy between DoTA and iCycle.

First, notice that for 0° and 180° there is no entry for 200 MeV. This is due to the fact that the 200 MeV overshot the patient from these angles, as at these angles the distance traveled through the patient is shortest, which can be seen in Figure 11 in Section 3.6.

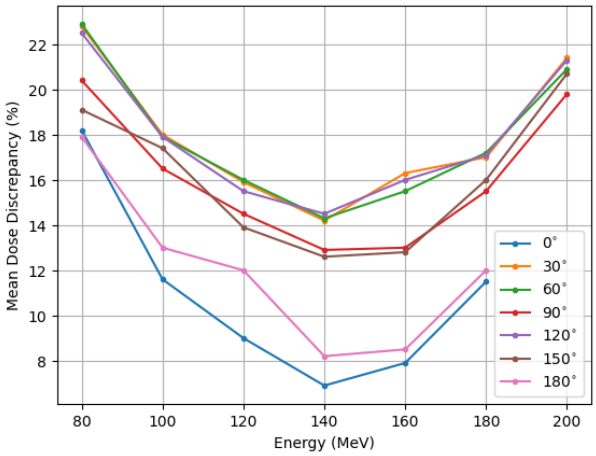


Figure 23: Mean relative absolute discrepancy between the DoTA and ASTROID dose distributions in a patient geometry as a function of energy for different gantry angles.

Figure 23 show mean relative absolute dose discrepancies up to 23.2 and 21.8% for the 80 and 200 MeV beamlets respectively. An observation is that the curves take a different shape compared to the mean discrepancy in water. The difference in mean relative dose discrepancy between the water sample and the patient geometry for the 80 MeV beamlets ($4 \pm 1\%$.) is smaller than that same difference for the 200 MeV beamlets ($10 \pm 1\%$.).

An analysis on the influence of the gantry angle is shown in Figure 24.

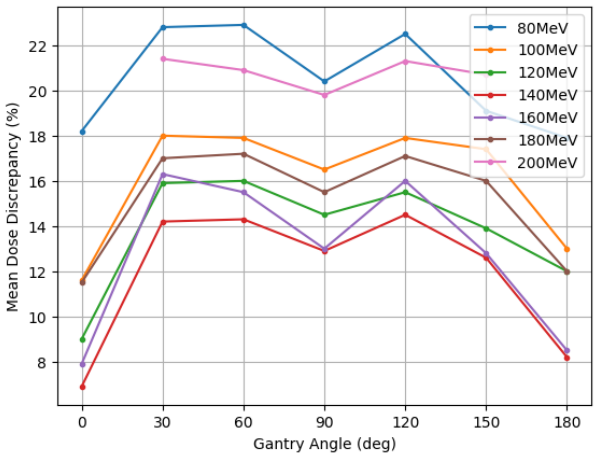


Figure 24: Mean relative absolute discrepancy between the DoTA ans ASTROID dose distributions in a patient geometry as a function of angle for different energies.

Figure 24 shows that the mean discrepancy increases as the gantry angles moves further away from perpendicular angles. Due to the asymmetry of a patient geometry, the discrepancies are no longer symmetric in the 90 ° gantry angle.

An analysis on the mean relative absolute discrepancy between DoTA and ASTROID for different beamlet angles inside this patient geometry, is displayed in Figure 25.

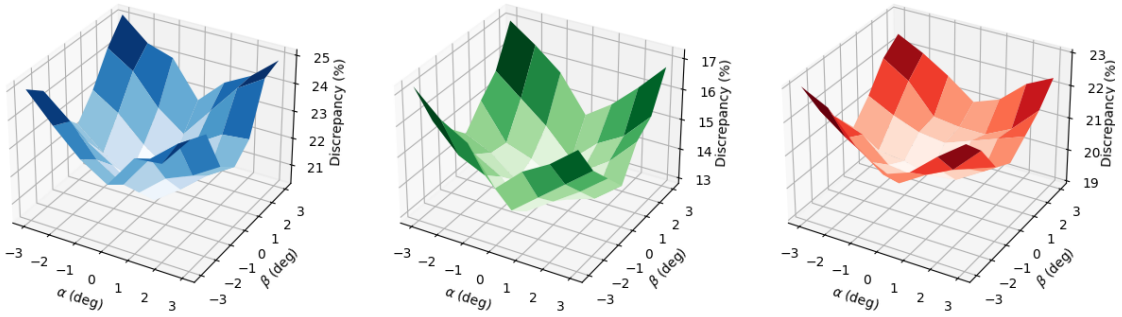


Figure 25: Mean relative absolute discrepancy between the DoTA and ASTROID dose distributions in a patient geometry as a function of the beamlet angle for a 80 MeV (Blue), 140 MeV (Green) and 200 MeV (Red) beamlet respectively.

Figure 25 shows that the mean relative absolute dose discrepancies tend to grow as we move away from the perpendicular axis, and grow even faster when combining beamlet angles in both directions. Moreover, it shows higher mean relative absolute discrepancies for combined gantry angles, than single ones.

The results of the comparison on the total energy by the two different beamlet models, is displayed in Figure 26

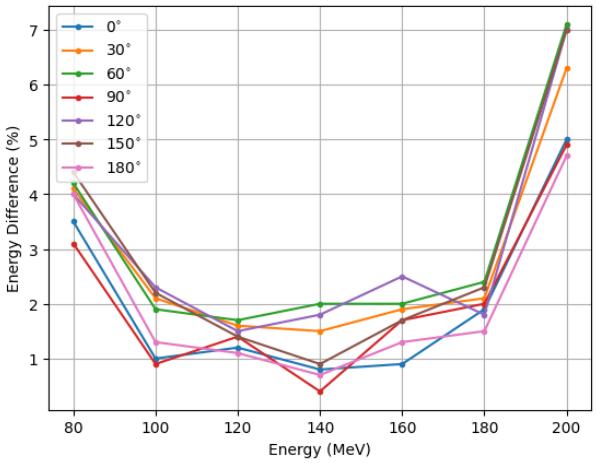


Figure 26: Discrepancy in the total energy distributed as a function of the beamlet energy for different gantry angles.

The discrepancies in the total distributed energy are $< 3\%$ for the majority of the beamlet in the range 100 MeV to 180 MeV. For 80 MeV and 200 MeV beamlets, these energy discrepancies increase up to of 4.5 and 7.1% respectively.

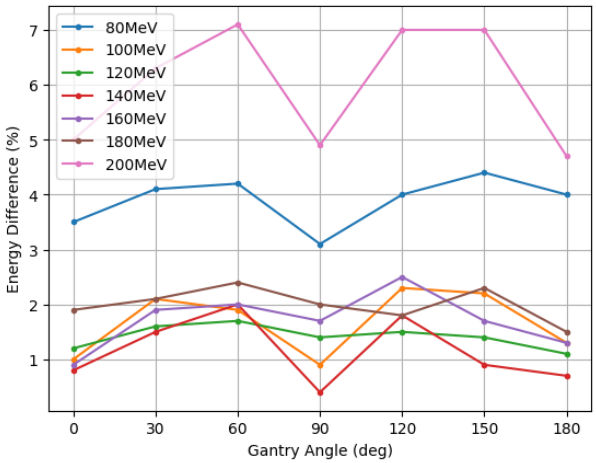
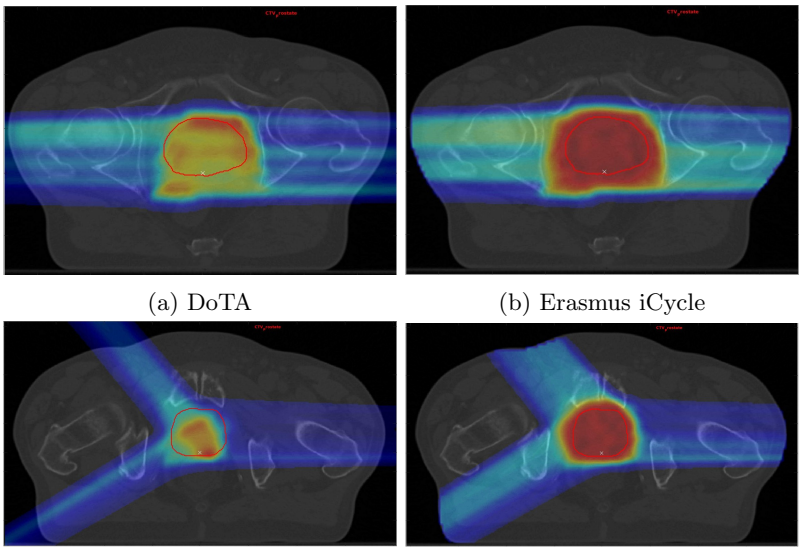


Figure 27: Mean discrepancy in the total energy distributed as a function of the beamlet energy.

Figure 27 shows (again) the overall increase in discrepancy as the gantry angle becomes non-perpendicular.

4.3 Full Dose Comparison

Figure 28 shows the dose distributions made by Erasmus-iCycle using ASTROID as well as an attempt to reproduce this by DoTA for two separate patients. The first plan consists of two lateral beams at 90 and 270 degrees. The second plan has a three beam set up to avoid the pelvic bone.

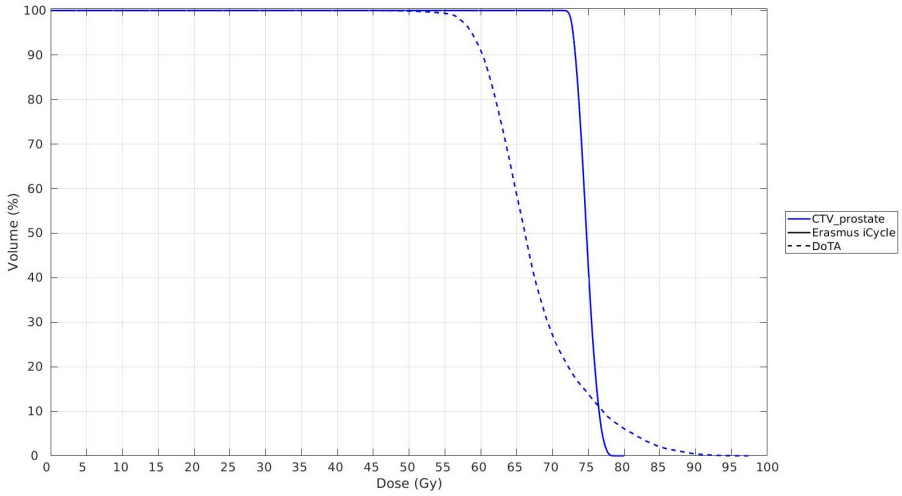


(a) DoTA (b) Erasmus iCycle

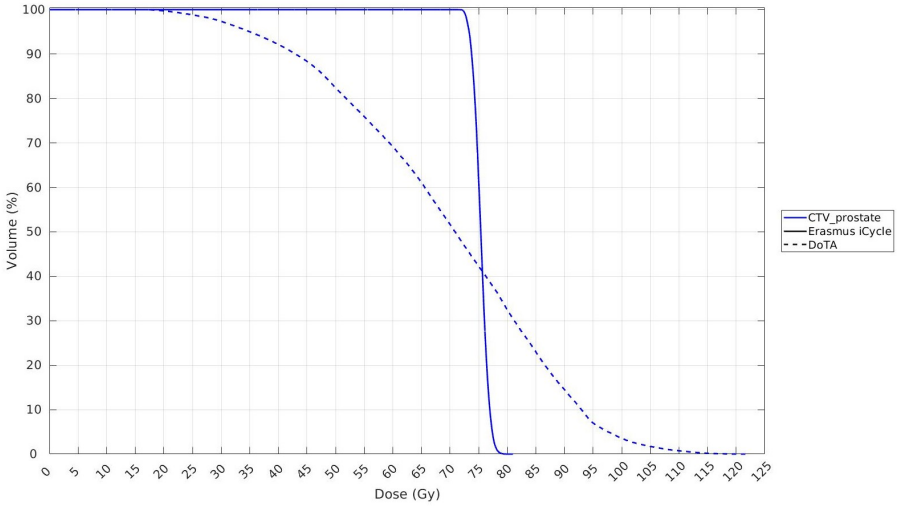
(c) DoTA (d) Erasmus iCycle

Figure 28: Dose distributions of two multi-beam plans computed by DoTA (a)(c) and Erasmus-iCycle ASTROID (b)(d).

We observe that the distributions show similarity in shape, but are different in terms of distributed dose, which can be confirmed by looking at the dose value histograms of the CTV for these two dose distributions.



(a) Two beam plan



(b) Three beam plan

Figure 29: Dose volume histograms of the CTV of the two-beam plan (a) and three beam plan (b).

The continuous line, representing the dose computed by Erasmus-iCycle's ASTROID, is what one would expect from a dose value histogram: 100% of the CTV receives the prescribed dose (70 Gy) after which the line radically drops to 0%. The DoTA doses however, show a major drop around 55 Gy, under-dosing the tumor by roughly 15 Gy. Figure 29b shows a drop already at 25 Gy, where 75 Gy was the target. Moreover, In both plans an overdosing on some parts of the CTV is observed.

The average computation time of a single spot has been calculated, the results of which are shown in Table 2. Note that the time to load the transformer model (about 10 seconds) are not taken into

account for DoTA in the total computation time, as this needs to happen once at the start of the computation. For a treatment plan generally consisting of a thousands of spots this is negligible.

Table 2: Average computation time of a single beamlet by both DoTA and ASTROID.

	DoTA time (s)	ASTROID time (s)
Preparation	7.18 s	
Dose computation	0.50 s	
Dose re-interpolation	7.86 s	
Total computation*	15.54 s	18.90 s

The actual dose computation by DoTA is only a fraction of the total computation time required for a single beamlet. The additional steps required to do a dose calculation slow it down by approximately 15 seconds, making it only 3.36 seconds faster per spot on average.

5 Discussion

This section will discuss the results collected over the course of this work. In addition to that, as in every field of research, there is always room for improvement and some limitations that bounds that improvement, which also will be looked into. Finally, some recommendations for further research will be made.

5.1 Results Interpretation

In this work, discrepancies were considered instead of errors. Despite the differences between DoTA and ASTROID, one is not necessarily more true than the other. Both are models used to describe dose disposition inside a Patient, and both depend on a different method to do perform this computation. One could even argue that DoTA, since it is trained on MC accurate dose computations, is likely more accurate than ASTROID, as it is a PBA.

First of all, a beam calibration was deemed necessary to be able to make a comparison between DoTA and ASTROID, the dose computation model used by the current of Erasmus-iCycle. The calibration was performed in water, and linked the MU's used by ASTROID with the particle stream used by DoTA.

From analysis in different media than water, it became clear that the calibration was not going to be uniformly valid in a patient geometry, as the range errors were still present in other media than water. Since the beamlets where calibrated in water, there is no discrepancy between the paired BPs for 0 HU. For the 80 MeV beamlet in HUs in the [-100,750] range, the BP depth discrepancies were within 2 mm (3.7%). For the 200 MeV however, the depth discrepancy would exceed 5.2% as soon as it passes through 500 HU. In 1000 HU, all of the BP pairs lined up poorly, with depth discrepancies ranging from 10.8 to 14.9%.

Secondly, performing tests in water, it became evident that the depth of the beamlets was not the only source of discrepancy between the two dose engines. Since both the depth as the center line BP dose of the beamlets in water has been calibrated, the discrepancies are likely a result of a difference in beam model. Both the mean and the maximum relative absolute discrepancy in water is largest for 80 MeV beams. Figure 21 shows that the dose distributed by DoTA is spread over a larger width. Due to the Gaussian nature of a beam, this means that the spread or full width at half maximum (FWHM) of the beam predicted by DoTA is larger than the one predicted by ASTROID, which is clearly a beam model difference. When analyzing how different angles influence the dose discrepancy, we observed that the mean relative dose discrepancy increases when the gantry angle moves away from the perpendicular axis. This is likely due to the interpolation that needs to be done in order to align the beams predicted by DoTA back on the original CT grid. This interpolation consists of a rotation part (gantry angle, beamlet angle) and a scaling part (voxelsize scaling). For perpendicular angles, less interpolations are required, since only the voxelsize scaling needs to be accounted for. The maximum relative absolute discrepancy showed very little dependence on gantry angle, compared to the mean discrepancy. This makes some sense if we'd assume this gantry angle dependence arises from the rotation part of the interpolation. Since the mean discrepancy takes into account the change in every voxel, while the maximum discrepancy only looks at the change in the voxel containing the maximum discrepancy, it is likely that the mean discrepancy experiences a bigger interpolation error. For the beamlet angle analysis, the mean relative absolute dose discrepancies tended to increase for increasing beamlet angle. Combination of beamlet angles in both plains resulted in higher discrepancies, likely due to the additional rotation required to compute these beams, after which they have to be re-interpolated an additional time to arrive back at the original CT grid.

For the Patient geometry, similar test were performed, but most discrepancies turned out larger. This was expected due to the examined range discrepancies in other HUs than water. Even though the calibration in water was performed on a 90° gantry angle, the discrepancies under this angle are overall larger than the beamlet under 0° and 180°. This likely has to do with the fact that the 90° beamlet has to pass through the pelvic bone, which has HUs going up to around 1000. Table 1 containing the BP range discrepancies in different HUs already showed range errors up to 12.0 and 14.9% for 140 and 200 MeV energy beams respectively passing through HUs in that range. For the beamlet angles, similar results as in water were found. In terms of total distributed energy, the two models did become closer, especially in the 100 MeV to 180 MeV area. Overall is the [120,160 MeV] energy range the area where the beam models showed the most similarity.

A comparison for a full treatment plan showed poor similarity, which was expected considering the dose and range discrepancies obtained comparing the dose distributions in patient geometries. In the case of the lateral beam plan, due to robustness effects, the discrepancies in range can sometimes compensate for each other. This is likely why the two beam plan still showed a DVHS of the CTV that somewhat contained the expected shape. However, the 100% drop is at 55 Gy, which means an under dosing of roughly 15 Gy (20.8%) for this plan. The three beam plan however, aimed to avoid the pelvic bone, showed worse results. This is likely due to the range discrepancies arising from the patient geometry, beam model differences and the couch on the bottom side of this image through which the beamlet passes.

5.2 Limitations

Firstly, the algorithm needs to interpolate between different voxelsizes in order to bring the CT to the grid on which DoTA was trained. DoTA's voxelsize is 2 mm × 2 mm × 2 mm, which is relatively large. In the case of this work, the voxelsizes of the CT image were always smaller than DoTA's. This means that the CT image is first brought to a larger voxelgrid, meaning less voxels, meaning a sacrifice of resolution in the CT. Afterwards, the dose distribution suffers the reverse translation, meaning it is translated from the 2 mm × 2 mm × 2 mm grid to the smaller grid, meaning an increase in voxels, leading to an extrapolation of the dose distribution. This interpolation of the CT, and extrapolation of the dose distribution are both sources of error. Secondly, the DoTA model used in this work has some practical limitations. First of all, this version of the model cannot make angles, it can only perform computations for a perpendicular beamlet. This means that in order to achieve realistic beamlets used in clinical practice, which enter the patient under a range of small angles, an additional interpolation of the CT is required to align the beamlet on the correct beamlet angle. Note that CT's are large three dimensional arrays, that require quite some computation time to interpolate for a given angle. A second limitation is the maximum depth that DoTA can predict, which is 30 cm. Generally, 30 cm is deep enough for most cases. However, during the research presented in this work there a CT of a larger patient geometry has been encountered, where the range of 30 cm was not large enough for the beamlet to reach the spot position. Moreover, DoTA does not have a rangeshifter, meaning it can also not compute high energy doses in smaller geometries, like head and neck patients. The combination of these range limitations allow for a smaller selection of patient geometries in which the model can be applied.

5.3 Further Research

There is still potential for DoTA as a treatment planning dose engine, as it has proven its accuracy in the past (Pastor-Serrano and Perkó 2022; Valk Bouman 2022). Now that there is also an algorithm that makes it able to perform either full or nominal dose computations for any given set of treatment

planning parameters, a lot of pieces of the puzzle are present. However, the beam models used by DoTA and ASTROID are currently too different to replace ASTROID by DoTA as Erasmus-iCycle's dose engine. To overcome this difference, DoTA needs to be retrained. We want to keep the MC accuracy, so we cannot simply train DoTA on ASTROID dose computations. Best would be to find out the exact beam parameters used by Erasmus iCycle, and do MC simulations with these parameters, then train DoTA on that.

DoTA in this state is quicker than ASTROID, however it would be preferable if it'd be faster, as the eventual goal is real time adaptive treatment. For this, the dose engine needs to quick enough to do full dose (re-)computations during treatment. A potential solution to reduce computation times is the DoTA-S, or DoTA Shape model (Appendix A). DoTA-S is a version of DoTA that has been retrained on MC accurate doses that include beamlet angles, which was unfortunately not available in a usable state during this work. As it is capable of handling beamlet angles, it can potentially save computation time by sparing the separate CT interpolation per spot that is now required to account for the beamlet angle.

Since DoTA needs a retraining anyway, it would be interesting to create an alternate version that is trained on beams containing a range shifter. Including the option to enable or disable a range shifter will increase the potential amount of patient geometries on which it can be used by a lot.

Lastly, Erasmus iCycle is mainly written in a MATLAB environment, DoTA is a deep learning model in Python. MATLAB and Python can be very difficult to work with together, due to certain conventions and conflicting libraries. There are sophisticated ways to still make a direct integration possible, but those required too much time and was therefore not possible to implement within the time span of this project. However, making that work would be a great step forward as you would no longer have to manually bring your data back and forth between the two.

6 Conclusion

Currently, the results show that dose distributions computed by DoTA are quite different than those computed by Erasmus-iCycle's current dose engine ASTROID. Since ASTROID is a PBA based on the actual beam model used in Erasmus iCycle, we may assume it to be very accurate in water, since it is the simplest of geometries. Since DoTA - after calibration to account for the initial range discrepancy and dose conventions - is still experiencing dose mean and maximum relative absolute dose discrepancies up to 18.9 and 60% respectively for 80 MeV beamlets, it may be concluded that the beam model on which DoTA was trained is different than that used by Erasmus-iCycle for treatment planning. Additionally, the interpolations and translations required to compute a dose distribution with DoTA seemed to induce an additional error, as both beamlets under non-perpendicular gantry angles and beamlets under small beamlet angles had higher mean dose discrepancies than perpendicular ones.

The analysis in different HUs showed that the range discrepancy between DoTA and ASTROID encountered in water, was also present in different HUs than water, from which can be concluded that both dose engines handle the proton stopping power differently. With depth discrepancies ranging from -7.6% up to 14.9%, the potential BP range discrepancy in a patient is too large.

The combination of range discrepancies and beam model differences led to mean relative absolute dose discrepancies up to 23.2% in a patient geometry, which is not even including errors arising from interpolation of the beamlet angle. In terms of total energy distributed, the discrepancy reduced to $< 3\%$ for beamlet energies on the interval [100,180 MeV]. From this it can be concluded that the dose discrepancies are, at least in this interval, mainly the result of beam model differences and less the scaling and dose re-interpolation errors.

The combination of differences in dose (due to the difference in beam models) and range (due to a difference stopping power interpretation) led to poor similarity of complete treatment plan doses. The lateral beam plan by DoTA showed some characteristics of a typical DVHS, but under-dosed the CTV by roughly 20.8%. Therefore it can be concluded that in this current state, DoTA is currently not feasible as Erasmus-iCycle's dose engine.

A DoTA-S

DoTA-S is a variant of the DoTA model that is retrained to handle beamlet angles between. It works similar as the original DoTA, but with an extra shape parameter, defined as: $s \in \mathbb{R}^{L \times H \times W}$ (Valk Bouman 2022). This shape is a 3D computed array of size 150×24×24, the same size of the CT-image input. Figure 30 shows the construction of this shape array.

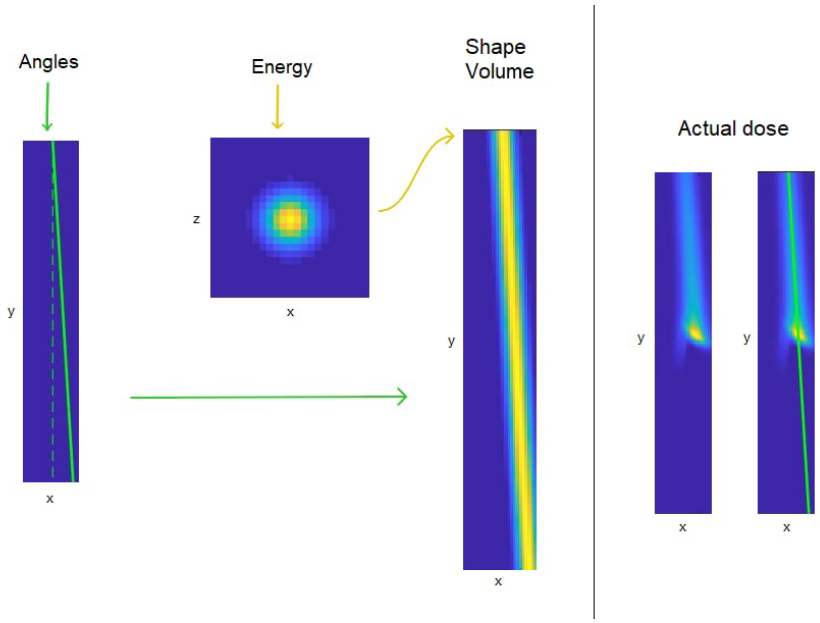


Figure 30: Working principle of the DoTA-S model. Left: A line representing the beamlet angle is projected along the beam axis. When combined with a predicted energy distribution of the beam in the first layer, it can project this energy distribution along the line to create a 3D shape volume. Right: The expected direction due to the angle aligns with the actual dose distribution (Valk Bouman 2022).

The energy distribution is based on a two dimensional normal distribution based on the variance of the beam model (Valk Bouman 2022). A drawback compared to the original DoTA is that it is trained on a smaller energy range, namely $E = [70, 140]$ instead of $E = [70, 220]$.

References

Albertini, Francesca et al. (2020). "Online daily adaptive proton therapy". In: *The British journal of radiology* 93.1107, p. 20190594.

American Cancer Society, The (2018). "Global cancer facts and figures 4th edition". In: pp. 1–73.

Azimi, Parisa and Amir Movafeghi (Aug. 2016). "Proton Therapy in Neurosurgery: A Historical Review and Future Perspective Based on Currently Available New Generation Systems". In: *International Clinical Neuroscience Journal* 3, pp. 59–80. DOI: 10.22037/icnj.v3i2.13324.

Barnett, Gillian C et al. (2009). "Normal tissue reactions to radiotherapy: towards tailoring treatment dose by genotype". In: *Nature Reviews Cancer* 9.2, pp. 134–142.

Bazalova, Magdalena et al. (2008). "Dual-energy CT-based material extraction for tissue segmentation in Monte Carlo dose calculations". In: *Physics in Medicine & Biology* 53.9, p. 2439.

Begg, Adrian C, Fiona A Stewart, and Conchita Vens (2011). "Strategies to improve radiotherapy with targeted drugs". In: *Nature Reviews Cancer* 11.4, pp. 239–253.

Bencheikh, Mohamed, Abdelmajid Maghnouj, and Jaouad Tajmouati (2019). "Percentage depth dose fragmentation for investigating and assessing the photon beam dosimetry quality". In: *Journal of Radiotherapy in Practice* 18.3, pp. 280–284.

Breedveld, Sebastiaan, David Craft, et al. (2019). "Multi-criteria optimization and decision-making in radiotherapy". In: *European Journal of Operational Research* 277.1, pp. 1–19.

Breedveld, Sebastiaan, Pascal RM Storchi, et al. (2012). "iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans". In: *Medical physics* 39.2, pp. 951–963.

Crosby, David et al. (2022). "Early detection of cancer". In: *Science* 375.6586, eaay9040.

Delaney, Geoff et al. (2005). "The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines". In: *Cancer: Interdisciplinary International Journal of the American Cancer Society* 104.6, pp. 1129–1137.

Elcim, Yelda, Bahar Dirican, and Omer Yavas (2018). "Dosimetric comparison of pencil beam and Monte Carlo algorithms in conformal lung radiotherapy". In: *Journal of Applied Clinical Medical Physics* 19.5, pp. 616–624.

Escobar-Peralta, Elizabeth et al. (2022). "Helical tomotherapy: advanced radiotherapy technology". In: *Revista médica del Hospital General de México* 85.4, pp. 179–184.

Fan, Jiawei et al. (2019). "Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique". In: *Medical physics* 46.1, pp. 370–381.

Feng, Hongying et al. (2022). "GPU-accelerated Monte Carlo-based online adaptive proton therapy: a feasibility study". In: *Medical physics* 49.6, pp. 3550–3563.

Feng, Mary et al. (2018). "Machine learning in radiation oncology: opportunities, requirements, and needs". In: *Frontiers in oncology* 8, p. 110.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. http://www.deeplearningbook.org. MIT Press.

Grégoire, Vincent et al. (2018). "Delineation of the primary tumour clinical target volumes (ctv-p) in laryngeal, hypopharyngeal, oropharyngeal and oral cavity squamous cell carcinoma: Airo, caca, dahanca, eortc, georcc, gortec, hknpcsg, hncig, iag-kht, lprhht, ncic ctg, ncri, nrg oncology, phns, sbrt, somera, sro, sshno, trog consensus guidelines". In: *Radiotherapy and Oncology* 126.1, pp. 3–24.

Hamed, Ghada et al. (2020). "Deep learning in breast cancer detection and classification". In: *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*. Springer, pp. 322–333.

Jagt, Thyrza et al. (2018). "An automated planning strategy for near real-time adaptive proton therapy in prostate cancer". In: *Physics in Medicine & Biology* 63.13, p. 135017.

Javaid, Umair et al. (2021). "Denoising proton therapy Monte Carlo dose distributions in multiple tumor sites: A comparative neural networks architecture study". In: *Physica Medica* 89, pp. 93–103.

Kontaxis, C et al. (2017). "Towards fast online intrafraction replanning for free-breathing stereotactic body radiation therapy with the MR-linac". In: *Physics in Medicine & Biology* 62.18, p. 7233.

Kooy, Hanne M et al. (2010). "A case study in proton pencil-beam scanning delivery". In: *International Journal of Radiation Oncology* Biology* Physics* 76.2, pp. 624–630.

Levine, Adrian B et al. (2019). "Rise of the machines: advances in deep learning for cancer diagnosis". In: *Trends in cancer* 5.3, pp. 157–169.

Lin, Hui et al. (2021). "Deep learning for automatic target volume segmentation in radiation therapy: a review". In: *Quantitative Imaging in Medicine and Surgery* 11.12, p. 4847.

Lin, Y, Fuming Wang, and B Liu (2018). "Random number generators for large-scale parallel Monte Carlo simulations on FPGA". In: *Journal of Computational Physics* 360, pp. 93–103.

Ma, CM et al. (2004). "Monitor unit calculation for Monte Carlo treatment planning". In: *Physics in Medicine & Biology* 49.9, p. 1671.

McElnay, Philip and Eric Lim (2014). "Adjuvant or neoadjuvant chemotherapy for NSCLC". In: *Journal of Thoracic Disease* 6.Suppl 2, S224.

Mohan, Radhe (2022). "A review of proton therapy–Current status and future directions". In: *Precision Radiation Oncology* 6.2, pp. 164–176.

Mohan, Radhe and David Grosshans (2017). "Proton therapy–present and future". In: *Advanced drug delivery reviews* 109, pp. 26–44.

National Cancer Institute, The (2015). *Surgery to Treat Cancer*. URL: `https://www.cancer.gov/about-cancer/treatment/types/surgery#types-of-cancer-treated-with-surgery`.

Nomura, Yusuke et al. (2020). "Fast spot-scanning proton dose calculation method with uncertainty quantification using a three-dimensional convolutional neural network". In: *Physics in Medicine & Biology* 65.21, p. 215007.

Pastor-Serrano, Oscar and Zoltán Perkó (2022). "Millisecond speed deep learning based proton dose calculation with Monte Carlo accuracy". In: *Physics in Medicine & Biology* 67.10, p. 105006.

Rajoub, Bashar (2020). "Supervised and unsupervised learning". In: *Biomedical signal processing and artificial intelligence in healthcare*. Elsevier, pp. 51–89.

Rojo-Santiago, Jesús et al. (2021). "Accurate assessment of a Dutch practical robustness evaluation protocol in clinical PT with pencil beam scanning for neurological tumors". In: *Radiotherapy and Oncology* 163, pp. 121–127.

Royce, Trevor J, Muhammad M Qureshi, and Minh Tam Truong (2018). "Radiotherapy utilization and fractionation patterns during the first course of cancer treatment in the United States from 2004 to 2014". In: *Journal of the American College of Radiology* 15.11, pp. 1558–1564.

Sajid, Sidra, Saddam Hussain, and Amna Sarwar (2019). "Brain tumor detection and segmentation in MR images using deep learning". In: *Arabian Journal for Science and Engineering* 44, pp. 9249–9261.

Schiffman, Joshua D, Paul G Fisher, and Peter Gibbs (2015). "Early detection of cancer: past, present, and future". In: *American Society of Clinical Oncology Educational Book* 35.1, pp. 57–65.

Siegel, Rebecca L et al. (2023). "Cancer statistics, 2023". In: *CA: a cancer journal for clinicians* 73.1, pp. 17–48.

Teoh, Suliana et al. (2020). "Is an analytical dose engine sufficient for intensity modulated proton therapy in lung cancer?" In: *The British Journal of Radiology* 93.1107, p. 20190583.

The MathWorks Inc. (2022). *What is Machine Learning?* URL: https://nl.mathworks.com/discovery/machine-learning.html.

Valk Bouman, Jim van der (2022). "Modeling ray angles in deep learning based dose calculation algorithms". In.

Vaswani, Ashish et al. (2017). "Attention is all you need". In: *Advances in neural information processing systems* 30.

Walters, R (2017). "Stopping-Power & Range Tables for Electrons, Protons, and Helium Ions". In: *NIST Standard Reference Database 124, NIST*.

Wu, Chao et al. (2021). "Improving proton dose calculation accuracy by using deep learning". In: *Machine learning: science and technology* 2.1, p. 015017.

Wyld, Lynda, Riccardo A Audisio, and Graeme J Poston (2015). "The evolution of cancer surgery and future perspectives". In: *Nature reviews Clinical oncology* 12.2, pp. 115–124.

Yuan, Tai-Ze, Ze-Jiang Zhan, and Chao-Nan Qian (2019). "New frontiers in proton therapy: applications in cancers". In: *Cancer Communications* 39, pp. 1–7.