1290

UNIVERSIDADE Đ COIMBRA

Eva Gouveia Anjo

Radiotherapy Treatment Planning Optimization: an approach using Q-learning

Thesis submitted to the Faculty of Sciences and Technology of the University of Coimbra for the degree of Master in Biomedical Engineering with specialization in Biomedical Instrumentation, supervised by Prof. Dr. Joana Matos Dias and Prof. Dr. Humberto Rocha.

September 2023

1290
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE Đ COIMBRA

Radiotherapy Treatment Planning Optimization: an approach using Q-learning

Eva Gouveia Anjo

Thesis submitted to the Faculty of Sciences and Technology of the University of Coimbra

for the degree of Master in Biomedical Engineering with specialization in Biomedical Instrumentation.

Supervisors:
Professora Doutora Joana Matos Dias
Professor Doutor Humberto Rocha

Coimbra, 2023

Este trabalho foi desenvolvido no âmbito do projeto ORION, financiado pelo Compete 2020 e pelo programa regional Lisboa 2020 com a referência POCI-01-0247-FEDER-047222.

Agradecimentos

Chegando ao fim de 5 anos de muito esforço e trabalho, quero começar por agradecer a todas as pessoas que estiveram presentes nalguma parte deste percurso e que contribuíram para que este se tornasse ainda mais especial.

Aos meus orientadores, Professora Doutora Joana Matos Dias e Professor Doutor Humberto Rocha, o meu mais sincero agradecimento por todo o apoio ao longo deste projeto, no qual foram incansáveis. Demonstraram-se sempre disponíveis para me ajudar no que fosse necessário, mesmo que fosse apenas com uma palavra de incentivo. São pessoas que admiro pelo trabalho exímio que realizam e pela acessibilidade com que transmitem tudo o que sabem. Foi um gosto enorme aprender com ambos.

A todos os amigos que fiz no Departamento de Física, bem como ao departamento em si, que se tornou casa, o meu muito obrigada. Aos Cruzados, Caramelo, Castela, Costini, João, Laura, Luís, Mariana, Martini, Mati, Mendonça, Prata, Tomé, Xanax e Xico, por todas as memórias inesquecíveis, por todas as horas de riso e de festa, mas também por todas as horas de apoio e entreajuda. São o grupo de faculdade que toda a gente teria orgulho em ter. Agradeço em especial à Mariana, por todas as horas de conversa e desabafo, sobre a tese, cadeiras ou sobre a vida. Obrigada por seres tão parecida comigo e por isso nunca me fazeres sentir sozinha. Agradeço ainda em especial à Laura, a minha companheira de faculdade desde o primeiro dia. A pessoa que me conhece melhor profissional e pessoalmente, a que mais me apoia e ajuda e sem a qual verdadeiramente não teria conseguido realizar este percurso. És para mim uma fonte de admiração imensa.

Às Potas, Pipa, Inês, Mariana Marques, Botinas, Mariana Simões, Bia, Sara, Sofia, Laura, Maria, Carolina Antunes, Matilde, Joana e Carolina Novo, por me terem acolhido no primeiro ano e terem tornado todo o meu percurso académico mais memorável. À Jacinta, ao Amado, ao Ivo, ao André e ao Manel, porque merecem ser mencionados pelo apoio e amizade constante, que espero manter sempre ao longo da vida. Gosto muito de vocês e de todas as nossas memórias.

Às minhas amigas de Bologna, Matilde, Sara e Marçal, agradeço-vos muito a amizade, não têm noção o quanto cresci graças a vocês. À Matilde, obrigada por seres a minha amiga de barco e de coração. Não teria chegado aqui sem ti.

Aos que não fazem parte da minha vida académica, mas que mesmo assim não deixam de estar presentes. Ao Francisco, Gui, João e Zé. É muito bom crescer com vocês e sentir-vos sempre comigo em todas as etapas da minha vida. Ao Bonito, obrigada por estares sempre à distância de uma chamada e por seres um apoio constante. Tenho muito orgulho em ser tua amiga e em partilhar tudo da minha vida contigo. À Rachinhas, nunca terei palavras suficientes para te agradecer. Todas as etapas da minha vida são mais especiais graças a ti, és a melhor amiga que alguma vez poderia ter.

Por fim, o meu maior agradecimento é para a minha família. Às minhas irmãs, pai, avós e tios. Não há nada que me deixe mais concretizada do que partilhar o meu percurso com vocês e sentir sempre o carinho e orgulho que sentem por mim. À minha mãe deixo o meu mais sincero e especial agradecimento, por ser a minha maior fonte de apoio e inspiração. És um exemplo de força inigualável e nunca terei palavras suficientes para agradecer o teu amor, carinho e a forma como acreditas sempre em mim.

Resumo

O cancro é um problema de saúde que todos os anos afeta milhões de pessoas, sendo que muitos doentes são submetidos a tratamentos de radioterapia, um dos possíveis tratamentos para a doença. O desenvolvimento de modelos e métodos para um melhor planeamento de tratamentos de radioterapia, com o objetivo de aumentar a precisão da distribuição da radiação no tumor e diminuir a sua distribuição nos tecidos circundantes, é uma importante e ativa área de investigação.

O planeamento radioterapêutico desempenha um papel crucial na radioterapia. Permite criar planos de tratamento individualizados que têm em conta as características do tumor, os tecidos saudáveis circundantes e a saúde global do doente. Trata-se de um processo complexo que envolve escolher os ângulos ou arcos de radiação, bem como a sua intensidade, através de uma análise detalhada do plano de tratamento e correspondente dose depositada nas estruturas de interesse, que é normalmente realizada pelo planeador. O facto do planeamento ser usualmente feito manualmente, com a ajuda de um sistema de planeamento (TPS), através de um procedimento de tentativa e erro, pode influenciar a qualidade final do tratamento. A otimização deste processo, através da automatização da criação do plano de tratamento, pode contribuir para a resolução de algunas dos problemas existentes. A incorporação de Inteligência Artificial potencia a automatização dos planos de tratamento, permitindo a obtenção de planos de tratamento de qualidade consistente, libertando o planeador para outros trabalhos importantes como verificar a qualidade final dos tratamentos.

O foco desta dissertação é a utilização combinada de optimização e *Reinforcement Learning* para a criação automática de planos de tratamento de elevada qualidade num período de tempo reduzido. Optou-se por se trabalhar com *Q-learning*, testando a abordagem desenvolvida considerando casos de cancro de próstata. Numa primeira fase foram consideradas três estruturas de interesse, tendo sido definidos os possíveis estados do sistema através da análise destas três estruturas como um todo. Os resultados obtidos permitiram concluir que, apesar do algoritmo ser capaz de convergir e alcançar um plano admissível, o tempo necessário à construção do plano de tratamento não diminuia quando comparado com a abordagem que apenas recorre à otimização, sem utilização de *Q-learning*.

Na segunda parte deste estudo, foi desenvolvida uma abordagem alternativa, trabalhando de forma específica com cada estrutura, o que contribuiu para atingir planos de tratamento mais rapidamente. Os resultados obtidos foram significativamente mais promissores, uma vez que a redução temporal foi observada consistentemente. Utilizou-se ainda o método de classificação cruzada nesta segunda abordagem, de modo a alcançar uma solução mais robusta. A utilização desta abordagem nos dados já recolhidos por *Q-learning* revelou demonstrar os melhores resultados e representa assim potencial para investigações futuras no campo do planeamento de tratamentos.

Palavras-Chave: Radioterapia, Planeamento de Tratamentos, Inteligência Artificial, *Reinforcement Learning*, *Q-Learning*

Abstract

Cancer represents a worldwide health issue that impacts millions of people annually, with several individuals undergoing Radiotherapy (RT) as a potential therapeutic option within the range of available treatments for this disease. Work is constantly being developed in RT treatment planning to further enhance the treatment quality by accurately increasing the radiation delivered to the tumor and decreasing radiation in the surrounding tissues.

Treatment planning plays a crucial part in radiotherapy. It allows the creation of an individualized treatment that takes into account the characteristics of the tumor, the surrounding healthy tissues and the patient's general health. It is a complex procedure that involves choosing radiation angles or arcs and intensities and therefore demands a thorough analysis, usually made by a treatment planner. The fact that planning is usually done manually through a trial-and-error procedure, resorting to a Treatment Planning System (TPS), can influence the accuracy of the treatment. Optimizing this procedure is a possible solution to tackle this issue. Incorporating Artificial Intelligence (AI) in treatment planning can potentially improve the automation of treatment planning, enabling the consistent creation of high quality treatment plans and releasing the planner for other important tasks such as quality assurance (QA).

The focus of this dissertation is to use in a combined way optimization and Reinforcement Learning (RL), an AI approach, to achieve feasible treatment plans in a reduced time frame. We chose to use Q-learning in this work, a RL method that gathers information from the environment that surrounds it and learns from it. In the first part of this study, considering prostate cancer cases, Q-learning gathered information considering the state of the system as being defined by three structures of interest altogether. The results quickly led to the conclusion that, although the algorithm was able to converge and reach a feasible plan, treatment planning time did not decrease consistently when compared with the use of optimization alone.

In the second approach, Q-learning gathered information looking at each structure in particular, which contributed to a faster calculation of treatment plans. The results were far more promising since time reduction was observed. Cross-validation was also used in this second approach, in order to achieve a more robust solution. Using this approach on the data already collected by Q-learning showed the best results of all and represents real potential for further investigation regarding treatment planning.

Keywords: Radiotherapy, Treatment Planning, Artificial Intelligence, Reinforcement Learning, Q-learning

Contents

List of Figures xiii

List of Tables xv

List of Abbreviations xvii

1 Introduction 1

1.1 Motivation 1

1.2 Objectives 2

1.3 Thesis Outline 3

2 Radiotherapy treatment planning 5

2.1 External Radiation Therapy 5

2.2 Radiotherapy Treatment Workflow 9

2.3 IMRT treatment planning 11

2.3.1 Beam angle optimization (BAO) 11

2.3.2 Fluence Map Optimization (FMO) 14

2.3.3 Realization problem 15

2.4 Automation of the radiotherapy treatment planning: current state-of-the-art . . 15

2.4.1 Fuzzy Inference Systems for FMO 18

3 Artificial Intelligence in Radiotherapy 23

3.1 Introduction 23

3.2 Reinforcement Learning 24

3.3 Machine learning methods 26

3.3.1 Markov random field (MRF) 26

3.3.2 Artificial Neural Networks (ANN) 26

3.3.3 Deep learning 26

3.3.3.1 Deep Neural Networks (DNN) 27

3.3.3.2 Convolutional Neural Networks (CNN) 27

3.3.4 Q-learning 27

3.4 Areas of implementation 28

3.4.1 Medical imaging 29

3.4.1.1 Medical image registration 29

3.4.1.2 Image segmentation 29

3.4.2 Treatment planning 30

3.4.3 Radiotherapy delivery 32

3.4.4 Radiotherapy verification and patient monitoring 32

3.4.5 Quality Assurance (QA) 33

4 Materials and methods **35**

4.1 Materials 35

4.2 Methods 36

4.2.1 Reinforcement Learning: Q-learning 37

4.2.1.1 First strategy - single Q-table 39

4.2.1.2 Second strategy - a Q-table for each structure 40

5 Results **43**

5.1 Q-learning table with all three structures 43

5.2 Q-learning table for each structure 46

6 Discussion **51**

6.1 First strategy 51

6.2 Second strategy 53

7 Conclusion **57**

Bibliography **59**

List of Figures

2.1 Linear Accelerator (Linac) [2] 6

2.2 (a) Illustration of a beam exiting the head of a gantry rotating around the treatment couch that can also rotate. (b) The head of the gantry is equipped with a multileaf collimator with nine pairs of leaves illustrating the discretization of the beam into small sub-beams called beamlets. 7

2.3 Membership functions 20

2.4 Defuzzification process 21

3.1 Schematic process of reinforcement learning 25

4.1 matRad interface 36

5.1 Two different cases, (a) and (b), with respective action choices 43

5.2 DVH for patient 0 regarding the first strategy 46

5.3 Two different cases, (a) and (b), with respective action choices 46

5.4 Example of Q-tables for the second strategy 47

5.5 DVH for patient 0 regarding the second strategy 48

5.6 Q-tables with average values 49

5.7 Q-tables with maximum values 49

5.8 Q-tables with minimum values 49

5.9 DVH from maximum Q-table values 50

List of Tables

4.1 CT resolution for each five patients 35

4.2 Medical prescriptions for each structure 36

4.3 Example of a Q-table 39

4.4 Matrix regarding the current and previous state 41

5.1 Training sets for each action with various patients 44

5.2 Comparison between training and testing sets with various patients 44

5.3 Q-learning matrix with two actions used for patient 0 45

5.4 Iterations of each training set 47

5.5 Iterations of each test set regarding each trained Q-table 48

5.6 Iterations of each test set 50

List of Abbreviations

3D-CRT Three-Dimensional Conformal Radiotherapy. 6, 7

AI Artificial Intelligence. ix, 1, 2, 14, 23, 26, 28, 30, 32, 33, 34, 37
ANN artificial neural networks. 26, 27

BAO Beam Angle Optimization. 8, 11, 12, 13, 14, 17, 31, 32
BEV beam's-eye-view. 12, 31

CNN convolutional neural networks. 27
CT computed tomography. 9, 17, 29, 35
CTV clinical target volume. 9

DNN deep neural networks. 27, 31, 32
DVH dose-volume histogram. 10, 16, 17, 30, 31, 46, 48, 50

FMO Fluence Map Optimization. 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 22, 31, 36, 37, 38, 42

Gy Gray. 5, 16, 35, 36, 39, 40, 44, 47

IMRT Intensity-Modulated Radiation Therapy. 6, 7, 8, 10, 11, 12, 13, 15, 17, 29, 32, 35, 39

KBP knowledge-based planning. 30, 33, 34

ML Machine Learning. 23, 24, 26, 28, 29, 30, 31, 32, 33, 34
MLC Multileaf Collimator. 7, 8, 15, 17
MRI magnetic ressonance imaging. 9, 29

OAR organs at risk. 8, 9, 11, 12, 14, 15, 16, 17, 18, 19, 31, 35, 36, 37

PSM pattern search methods. 12
PTV planning target volume. 7, 9, 11, 12, 14, 15, 16, 17, 18, 19, 31, 35, 36, 37, 40, 42, 43, 44, 46, 47, 50, 52, 54

QA quality assurance. ix, 2, 11, 33, 34

RL Reinforcement Learning. ix, 2, 23, 24, 25, 27
RT Radiotherapy. ix, 1, 2, 5, 6, 7, 8, 9, 10, 11, 13, 23, 26, 27, 28, 29, 30, 31, 32, 33, 36, 37, 38, 43, 51, 55, 57, 58

SSL semi-supervised learning. 24

TPS Treatment Planning System. ix, 1, 14, 17

VMAT Volumetric Modulated Arc Therapy. 6, 7, 9

1

Introduction

1.1 Motivation

Cancer is the leading cause of death in economically developed countries and the second leading cause of death in developing countries, being the first or second leading cause of death before the age of 70 years in 112 of 183 countries, in 2019 [1]. In 2020, there were approximately 10 million cancer deaths and an estimated 19.3 million new cancer cases worldwide. Multiple factors contribute to the growth of cancer incidence and mortality that is currently happening, such as aging and the growth of the population, as well as changes in the prevalence and distribution of the main risk factors for cancer, several of which are associated with socioeconomic development [1].

One of the possible cancer treatment modalities is Radiotherapy (RT), which is based on the idea that cancer cells are primarily concerned with rapid reproduction and are therefore less capable of repairing themselves through radiation, in contrast to healthy cells [2]. This modality remains an important component of cancer treatment with approximately 50% of all cancer patients receiving radiation therapy during their course of illness and contributing to 40% of curative treatment for cancer [3].

RT treatment planning is a crucial part of the RT treatment workflow. It involves optimizing the treatment delivery in order to maximize the therapeutic effect while sparing healthy tissues and complying with the treatment prescription that was defined by the medical doctor.

Treatment planning plays a huge role in creating patient-specific treatment plans. These plans are based on a detailed imaging analysis that allows the creation of an individualized treatment that takes into account the characteristics of the tumor, the surrounding healthy tissues and the patient's general health. With the constantly evolving environment of cancer care, the importance of treatment planning rests not only in its ability to improve treatment effectiveness but also in its ability to adapt to specific patient demands, resulting in more favorable outcomes and enhanced quality of life.

A treatment plan will mainly define all the treatment parameters that must be considered for treatment delivery, from radiation angles or arcs to radiation intensities. As there are a myriad of possible configurations of all the parameters that define a treatment plan, treatment plan optimization is, in fact, a complex procedure. Nowadays, treatment planning usually consists of a complex trial-and-error procedure done by the planner, based on her/his experience, resorting to a Treatment Planning System (TPS). The use of mathematical optimization approaches and Artificial Intelligence (AI)-based approaches can contribute to the automation of this process, which can have important advantages since it can result in the consistent achievement of more accurate and precise plans, calculated faster and with minimum intervention from the human planner.

1.2 Objectives

Radiotherapy treatment planning is typically a time-consuming process. Despite the consistent planning objectives, defined by the medical prescriptions that usually tend to not differ too much from one patient to the other when the disease is similar, the planning outcomes may vary significantly due to anatomical patient diversity.

AI applications have grown over the last few decades, representing a massive breakthrough in many areas. Regarding RT, AI holds the potential to be applied in numerous aspects of the treatment workflow, spanning from treatment planning and delivery to outcome prediction and quality assurance (QA). AI algorithms are capable of rapidly analyzing complicated medical data, such as patient imaging and clinical history, in order to develop highly accurate and personalized treatment plans. This not only decreases the time required for planning but also improves the precision of targeting cancerous regions while minimizing damage to healthy cells. AI-powered automation can assist medical professionals in making better decisions, leading to higher treatment efficiency and better patient outcomes. Furthermore, the ability of AI to learn from big datasets allows for continuous modification of treatment plans, potentially paving the way for more novel and successful radiation approaches in the future.

In this work, an approach that joins optimization models and algorithms with artificial intelligence for radiotherapy treatment planning will be developed and tested. From the available AI approaches, we have chosen to work with Reinforcement Learning (RL). RL is a machine learning method that intends to foresee the optimum course of action given the current situation. Within RL, Q-learning has been chosen since it is a fully interpretable AI method, which is very important in health-related contexts and it is possible to use it in a way that somehow mimics the learning human planners do when planning treatments during the trial-and-error treatment planning process.

The main goal is to have a learning procedure that is able to guide treatment planning, helping obtain feasible treatment plans in a faster and more robust way, without the need for human intervention and therefore decreasing the need for trial-and-error procedures. Q-learning uses a Q-function that models the action-reward relationship. In this study, it is used to find the actions that quickly lead to convergence and a feasible solution [4].

We aim to verify if Q-learning is a promising new tool for treatment planning, verifying if it helps to reach feasible plans and if it does so in a reduced time window. Ultimately, the algorithm can learn from a given set of cases and then apply this knowledge to other, new cases. Optimizing treatment plans ensures that therapies are adjusted for maximum efficacy by customizing treatment strategies to individual patient characteristics and conditions. This results in better treatment outcomes, fewer side effects, and higher patient satisfaction. Furthermore, optimized plans frequently streamline the use of resources, reducing expenses and treatment length. In summary, incorporating optimization into treatment planning exemplifies the confluence of precision and efficiency, resulting in better overall patient care.

All methods and algorithms will be tested on a set of prostate cancer patients.

1.3 Thesis Outline

This dissertation is organized into 6 Chapters. Chapter 1 presents a brief summary of the problem and the main purposes of the study, as well as its alignment. Chapter 2 provides all the basic principles of radiotherapy, concerning the steps involved in radiotherapy, along with the various types of treatment. A state of the art regarding the methods used in this work is also presented. Chapter 3 summarizes artificial intelligence and machine learning concepts and displays multiple algorithms and their possible applications regarding radiotherapy. Chapter 4 overviews the materials and methods used to acquire and analyze the data, offering a detailed description of the characteristics of the cases considered, along with the two strategies that were implemented for this study. Chapter 5 exhibits the obtained results regarding the two strategies. Chapter 6 provides a discussion based on the results reached. Chapter 7 introduces the main conclusions considering the work developed and presents ideas for future work. Finally, a detailed listing of all the bibliographic references used is presented.

2

Radiotherapy treatment planning

This chapter introduces the main concepts associated with Radiotherapy (RT) treatment planning and delivery. Firstly, Section 2.1 presents some notions and definitions related to external radiation therapy along with its various techniques, followed by an overview of the treatment planning workflow in Section 2.2. Next, Section 2.3 will further explore the various steps involving IMRT treatment planning. Section 2.4 presents a state-of-the-art regarding the current level of automation in the different steps of the RT treatment workflow.

2.1 External Radiation Therapy

A range of therapeutic approaches are employed in cancer treatment, including surgical intervention, chemotherapy, radiotherapy, and other methods. RT is recommended for the majority of patients, and it can be used as a standalone treatment or combined with other treatment approaches, like chemotherapy or surgical procedures.

RT involves the targeted delivery of radiation to cancerous cells, aiming to destroy or inhibit their growth while minimizing damage to surrounding healthy tissues [5]. The primary intent can be therapeutic/curative or palliative, to control the symptoms in patients with incurable cancer [6].

Radiation interacts with matter through energy transfer, which can result in not only the excitation but also in the ionization of atoms or molecules, which leads to biological damage that can be mediated by direct or indirect action of radiation. Irradiation will cause cell sterilization as a direct effect of radiation, and one that is mediated by breaking DNA strands, most effectively if both strands are broken close to each other (double strand) since single-strand breaks are repaired more effectively through the structural support by the leads unbroken strand. For a dose of 1 Gray (Gy), which corresponds to one Joule of energy deposited in one kilogram of matter, approximately 10^5 ionizations per cell occur, but the yield of double-strand is only about 40 per cell [7]. The indirect effect is related to water radiolysis since the interaction between radiation and the water molecules induces the formation of free radicals. These radicals create reactive oxygen species (ROS), which then cause oxidative stress and eventually damage the cancer cells [8].

One of the factors that determine how radiation will interact with both tumorous and healthy cells is radiosensitivity, which translates into the susceptibility of cells, organisms or tissues to the damaging effects of ionizing radiation [8]. RT is based on the principle that tumorous cells have fast and uncontrolled growth, making them more radiosensitive than healthy cells, whose repair capacity is higher.

In the field of RT, several techniques have been developed to effectively deliver radiation to the tumor site. This is difficult to do due to the inherent conflict between precisely delivering radiation to the tumor and sparing the surrounding organs and tissues. As a result, extensive research is being conducted to optimize and achieve the maximum potential degree of this delicate compromise. Within radiotherapy treatments there is teletherapy (also referred to as external beam therapy) and brachytherapy. In brachytherapy, radioactive seeds are inserted within the tumor region. In teletherapy, radiation is delivered from outside the body and directed at the location of the patient's tumor, which can be accomplished by employing a range of equipment that allows for different delivery modes.

External radiation therapy, the one that is the focus of this work, is performed with the patient laying on a couch that can rotate. Radiation is generated by a linear accelerator (linac) mounted on a gantry that can rotate along a central axis (Figure 2.1). The gantry rotation and couch rotation combined allow radiation to be delivered from almost any angle around the tumor. The point where the linac gantry rotation axis intersects with the central axis of the linac is called the isocenter, a geometric reference point, typically placed inside the tumor, that is strategically intersected by the radiation beams [2]. A treatment plan is considered coplanar if the couch is fixed at a 0° angle during the whole treatment, being noncoplanar otherwise [9].

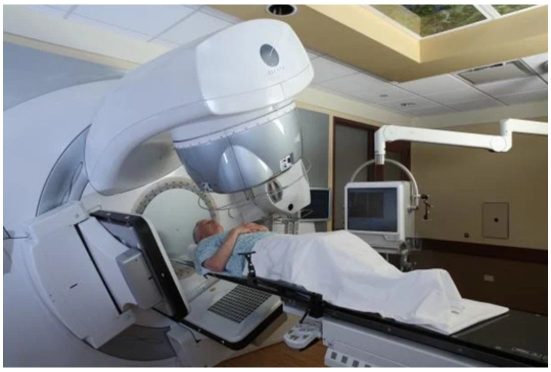


Figure 2.1: Linear Accelerator (Linac) [2]

In terms of external radiation treatment, there are also two types of RT: conventional and conformal. In conventional RT the radiation dose is delivered to the target through high-energy radiation beams, these beams being large enough to irradiate the whole volumes that need to be treated. In conformal RT the objective is to be able to achieve a high conformity between the volume to be treated and the doses absorbed by the tissues [2].

RT incorporates a range of techniques, and among them, Three-Dimensional Conformal Radiotherapy (3D-CRT), Volumetric Modulated Arc Therapy (VMAT) and Intensity-Modulated Radiation Therapy (IMRT) are the most commonly used external radiation therapy techniques. All three fall into the domain of conformal radiation therapy [2].

3D-CRT is an advanced technique that incorporates the use of imaging technologies to generate 3D images of a patient's tumor and nearby organs and tissues. The clinical introduction of 3D-CRT radically changed the RT workflow from a simulator-oriented "beam-adjusted" philosophy to an information-driven, computer-based process where the machine settings for treatment delivery were determined with the help of a treatment planning system, used by the treatment planner, where it is possible to include realistic dose calculations [7]. The beam directions, beam weights, field shapes, etc., constituting the free variables in 3D-CRT could be manually optimized. However, despite the advances achieved, the uniform radiation fields produced did not yet allow conformal treatments to be obtained. The modulation of the radiation intensities would only become possible when new modalities were developed, namely IMRT and VMAT.

Both VMAT and IMRT emerged as modality treatments with notable advancements in RT over the past decades, providing cancer patients with a wider variety of treatment options. The most cutting-edge RT systems use a variety of technological innovations to approach the physical limits of photon beam dose delivery. The radiation beams are modulated by a Multileaf Collimator (MLC), which is a system of adjustable leaves located in front of the linear accelerator treatment head, allowing the patient to be exposed to nonuniform radiation fields from specific angles [10,11]. With the assistance of multileaf collimators, the beam can be divided into a grid of smaller beamlets with independent intensities (Figure 2.2). The beamlet intensities can be optimized (Fluence Map Optimization (FMO) problem) leading to nonuniform radiation fields that can be sequenced and delivered while the gantry is halted at the given beam irradiation directions- static intensity-modulated radiation therapy (IMRT)- or can be delivered while the gantry rotates around the patient with the treatment beam always on, rotational/arc IMRT-VMAT.

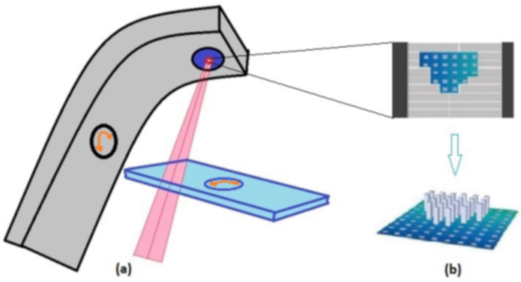


Figure 2.2: (a) Illustration of a beam exiting the head of a gantry rotating around the treatment couch that can also rotate. (b) The head of the gantry is equipped with a multileaf collimator with nine pairs of leaves illustrating the discretization of the beam into small sub-beams called beamlets.

VMAT has gained growing attention because of its decreased dose delivery time compared to fixed-field IMRT, representing a development of IMRT that offers potential advantages to the delivery of the radiation beam for various tumor sites [12]. VMAT comprises an arc trajectory and distributes doses dynamically during gantry rotation, in contrast to IMRT, which typically involves less than 10 fixed-field beam angles. VMAT delivers dose dynamically while the gantry rotates along an arc trajectory. The level of modulation from each beam direction in VMAT is still significantly smaller than that from each beam in fixed-field IMRT, even if more than one arc is sometimes used in this technique [13].

IMRT represents one of the most used RT techniques and, with the achieved scientific advances in treatment planning and delivery, it is now possible to shape photon beam irradiation dose distributions to the planning target volume (PTV) uniformly, whilst sparing surrounding tissues and organs at risk (OAR) [14]. These objectives are conflicting because radiation must pass through healthy tissues to reach the tumor. Furthermore, while practically any direction can be used to efficiently irradiate the tumor, properly sparing the normal tissues can only be accomplished if the beam irradiation directions are carefully chosen [9]. This suggests that, as a first stage in the treatment planning decision-making process, Beam Angle Optimization (BAO) is frequently required. This is done to obtain an arrangement of beam angles that leads to treatment plans with improved sparing of OAR, aiming to minimize their radiation exposure. With the advent of highly non-coplanar plans, finding a fast method for beam orientation selection is very useful, making IMRT more appealing [15]. For a given beam ensemble, the radiation to be delivered from each beam is then optimized (FMO), aiming to fulfill the prescribed and the tolerance doses. The last step involves a leaf sequencing problem, where it needs to be determined how the leaves of the MLC should move so that the optimal beamlet intensities calculated in the previous step are, in fact, delivered [16].

Leaf sequencing can be performed for multistatic or dynamic delivery. MLCs have movable leaves on both sides that can be positioned at any beamlet grid boundary. There are two possible ways of using collimation: dynamic and multiple static. In the multiple static collimation, also known as the "step-and-shoot mode", the leaves are set to open a desired aperture during each segment of the delivery, and radiation is on for a specific fluence time or intensity. When changing segments, the beam is off and there is no radiation being delivered. In contrast, in dynamic collimation (dMLC) the leaves move continuously during irradiation. Although generally considered faster than the step-and-shoot technique, the calculations for dMLC are more complicated because the discretization of the beams into well-defined beamlets is not as straightforward as for the step-and-shoot technique. Additionally, MLC leaves often have thin parts that overlap the sides of adjacent leaves in order to minimize interleaf leakage. Since the penumbras formed over the overlapping sections do not add up to produce a homogeneous fluence, this may result in an unintended underdosage. MLC movements are of the same order of magnitude as breathing; therefore, movement interference can cause significant deviations between planned and delivered doses [2, 7].

Besides the aforementioned modalities, there are other treatment techniques in RT, such as helical tomotherapy. In helical tomotherapy, the accelerator rotates in the gantry around the patient while the couch moves the patient slowly through it thus creating a spiral (helical) pattern of beam delivery. During this rotation, a computer-controlled MLC with two sets of interlaced leaves continuously modulates the radiation beam [7].

Regardless of the treatment modality considered, there is ongoing research into adaptive radiation therapy. Adaptive radiation therapy is a closed-loop radiation treatment process where the treatment plan can be modified using systematic feedback of measurements. Adaptive RT intends to improve radiation treatment by systematically monitoring treatment variations and incorporating them to re-optimize the treatment plan early on during the course of treatment. In this process, field margin and treatment dose can be routinely customized to each individual patient to achieve a safe dose escalation [17]. Adaptive RT can lead to the idea of "treatment of the day" where the treatment plan is adapted daily to better account for all the changes that may occur from one day to the other.

2.2 Radiotherapy Treatment Workflow

The process of RT is complex and involves an understanding of medical physics, radiobiology, radiation safety, dosimetry, radiation treatment planning, simulation and interaction of radiation with other treatment modalities [6]. It consists of three distinct steps:

1. immobilization, imaging and target volume definition
2. treatment planning
3. treatment delivery and set-up verification.

The first step is crucial to ensure the accurate delivery of RT. For that purpose, over the several weeks of treatment, patients must be properly immobilized and positioned on the treatment couch in such a way that the patient's position is comfortable, reproducible and optimal for the way the treatment will be administered [6]. The volume to be treated is determined by performing a planning computed tomography (CT) scan. In this planning CT, all the volumes of interest (both volumes to treat and organs to spare) must be delineated. The OAR are defined as the organs in the neighborhood of the tumor that could be damaged by radiation. The tumor should be defined taking into account all the diagnostic data, including clinical examination findings and results from diagnostic imaging or techniques like CT, magnetic ressonance imaging (MRI), positron emission tomography (PET), and endoscopic ultrasound [6]. The planning CT study is regarded as a snapshot of the patient at a specific point in time, and any deviation between the snapshot and the time-averaged tumor position (which is not available) results in a systematic error throughout the delivery sequence. Other deviations can be caused by the patient's movement since some movements cannot be suppressed completely, e.g. breathing and heartbeat [7]. In order to ensure consistent delivery of radiation to the designated area on a regular schedule, the patient must be positioned in the exact same orientation as during the planning scan prior to each treatment session [6].

Considering the volumes to be treated, three volumes are usually delineated: GTV, clinical target volume (CTV) and PTV. The most common target volume used during treatment planning is the PTV, which is created by adding a margin to the CTV to account for organ movement and set-up errors. CTV is defined by the gross tumor volume (GTV) with margin regions to cover the assumed spread of invading tumor growth. The PTV is usually the structure used for designing treatment plans, and adding a marginal volume to the CTV is a safety measure to prevent possible inaccuracies or variations.

Treatment planning, the second step of the RT process, involves creating a combination of RT beams (or arcs, for VMAT) and corresponding fluences that will deliver the required dose to the PTV. The optimization of treatment planning can be interpreted as the optimal selection of a number of parameters corresponding to a given treatment, among a range of acceptable solutions. Because of the complexity of the treatment planning process, treatment planning often involves a trial-and-error process. For IMRT, firstly, the medical physicist, based on her/his own experience, manually selects a beam ensemble. Then, the FMO takes place, considering these beam angles fixed. The resulting solution is then analyzed by the medical physicist who may change some of the beam angles, or other input parameters considered in the FMO, aiming at increasing the quality of the treatment plan. The FMO problem is solved again, and this laborious and time-consuming process is repeated until a satisfactory treatment plan is obtained. The procedure ends when the dose distributions are considered acceptable according to the medical prescription. If not, the treatment planning continues to be changed manually, which is a time-consuming process with no assurance of providing high-quality treatment plans. It is not possible to ensure that the treatment plans obtained by this process are the best possible for each patient.

Inverse planning emerged as an alternative process that involves using optimization models and algorithms that have been tuned by the treatment planner to determine an optimal treatment plan. Inverse planning is usually considered for FMO and it typically involves formulating an optimization problem, which comprises a set of constraints that must be satisfied and an objective function to be minimized or maximized with respect to a set of independent variables. The constraints serve as non-negotiable requirements for the problem, such as ensuring that the dose does not exceed a certain critical threshold.

The objective function, which can be composed of different functions combined, is optimized to the extent permitted by the constraints, with the aim of achieving the desired clinical outcomes. Although mathematical optimization is being used to determine treatment planning parameters, the iterative process of setting up the optimization problem can be very time-consuming. There is also the possibility of prioritizing treatment goals instead of using an objective function consisting of a weighted sum of conflicting goals. This involves including only the most critical category in the optimization and transforming the result into constraints before introducing the next category as objectives. This guided problem modification aims to reduce the number of iterations by providing tools to analyze and modify the problem [7].

Inverse treatment planning helps the human planner during the trial-and-error procedure, but it does not eliminate the need for human intervention: in every trial-and-error iteration what the human planner does is to define new parameters for the underlying optimization mathematical model considering the results of the previous treatment plans calculated.

The treatment quality is then usually assessed by its cumulative dose-volume histogram (DVH) and by analyzing isodose curves, among other quality assessment metrics and tools [2,9]. As some of the treatment objectives are conflicting with each other, each treatment plan can be seen as a compromise solution and treatment planning can be understood as a multi-objective optimization problem. Taking this into account, several approaches were developed to navigate the solution space in the large-scale problem that the IMRT optimization represents, using, for example, deterministic methods to determine search patterns, often by using information about the gradient of the objective function [7].

Treatment delivery, the last step of the RT process, depends on the type of treatment being used. In this step the treatment that was planned is delivered to the patient, meaning that the treatment parameters that were defined are now translated into the machine configuration parameters. For treatment delivery, the optimal fluences calculated have now to be transformed into MLC leaves movements and corresponding apertures. This is done by solving an aperture optimization problem, that optimizes leaves movements so that the desired fluences are obtained.

There is also an alternative approach, called direct aperture optimization, where fluence and apertures are simultaneously optimized.

Along every stage of the RT process, quality assurance (QA) is used, playing a vital role in the procedure. To ensure safety and quality, a feedback mechanism is also crucial since clinical practice may change over time. QA is especially important for the minimization of systematic errors. According to medical physicists, "quality assurance" refers to more than just lowering the incidence of obvious errors in radiation planning or delivery. QA is understood as including components of a patient's treatment requiring expertise regarding the patient and her/his condition as well as the best way to administer therapeutic radiation [18].

Simulation and delivery processes in RT have QA procedures that are considered "patient-independent". These procedures, such as dose calculation verification and IMRT QA, apply universal criteria to ensure the accuracy of treatment plans for all patients. However, contouring and plan optimization are highly specific to each patient, making standardization challenging [19].

2.3 IMRT treatment planning

In this work, we will mainly focus on IMRT treatment planning. The IMRT treatment planning is a sequential process that starts with the image acquisition, followed by the contouring of the PTV and OARs. Afterwards, the treatment planning usually starts by selecting a given number of beams and their angles, followed by calculating the optimal intensities (FMO) for each selected beam. Finally, delivery is done, in the various ways that were previously discussed. The major challenges concerning treatment planning and delivery will be presented in greater detail now.

2.3.1 Beam angle optimization (BAO)

A very important aspect to keep in mind is the fact that the integral dose to a patient's body tends to be fairly constant, regardless of the treatment planning approach considered, meaning that the best a planner can do is to decide where the excess radiation should be positioned, which healthy tissue to irradiate and which to spare. The configuration of beam directions has a major effect on the quality of the treatment plan [15].

BAO, which involves choosing the proper radiation incidence directions, may have an impact on the quality of IMRT plans particularly by enhancing organ sparing and increasing tumor coverage [15]. This involves calculating the ideal number of beams and determining the ideal beam angles. It is an important step in IMRT optimization since it has a direct impact on both the effectiveness of the therapy and the total treatment time, which increases as the number of beams increases [16].

An efficient BAO engine is especially critical when noncoplanar treatments are involved since this additional degree of freedom increases the complexity of beam selection and makes the number of possible iterations in a trial-and-error process even higher. This is why, in most cases, coplanar treatments with equally spaced beams are chosen, with the number of beams being predetermined by the planner based on prior experience and taking into account the location of the PTV and the OAR [20].

There are, however, other ways of calculating the optimal number of beams, such as classification methods. Applying methods like Support Vector Machines, Random Forests, Neural Networks, etc., to a database of previously treated patients is one way to calculate the number of beams for a new patient [11]. Then, an optimal IMRT plan is obtained by solving the FMO problem for a given beam angle set. The problem revolves around determining the optimal beamlet weights for the fixed beam angles since each beamlet has its own intensity. This approach allows the delivery of radiation in a more precise and controlled way, improving the quality of treatments delivered [10].

Regarding the beam directions, the current clinical treatment planning workflow involves manual selection by the planner. To overcome the problem of time-intensive calculations for dose-based metrics, several researchers have used purely anatomical metrics for BAO. Some of these approaches include sorting potential beam orientations based on their distance to the PTV and OAR, selecting beams in a specific order subject to a minimum distance threshold, using computer vision and beam's-eye-view (BEV) techniques to define a treatment plan based exclusively on geometric information or training a random forest regression algorithm using approved treatment plans to learn the relationship between patient anatomy and beam orientations [15].

BEV concept, which uses topographic criteria to rank the candidate beam directions, is one way of addressing the BAO problem. BEV dose metrics assign a score to each radiation beam direction based on topographic criteria. Beam's-eye-view dose (BEVD) evaluates each possible beam direction using a score function that accounts for beam modulation. In IMRT, beam directions are non-intuitive and may have to go through sensitive organs to achieve an optimal compromise between target coverage and organ sparing, which makes the geometrical criteria used by BEV limited. The optimal beam configuration for an IMRT treatment should balance the BEVD score and the beam interplay as a result of the overlap of radiation fields [10].

BAO can also consider the use of pattern search methods (PSM), which are directional direct search methods that belong to a broader class of derivative-free optimization methods, such that iterate progression is solely based on a finite number of function evaluations in each iteration, without explicit or implicit use of derivatives, which can avoid local entrapment. PSM are organized in two steps at every iteration [10]. The first step, called search step, provides the flexibility for a global search since it allows searches away from the neighborhood of the current iterate and influences the quality of the local minimizer or stationary point found by the method. The second step, called poll step, performs a local search in a mesh neighborhood and ensures the convergence to a local minimizer or stationary point. If the search step fails to produce a decrease in the objective function, the poll step is performed around the current iterate. As for the stopping criteria, usually, they are based either on the maximum number of function value evaluations allowed or on convergence criteria related to the mesh size. Pattern search methods have the ability to converge globally from arbitrary points to good local minimizer candidates [10].

Several other methodologies have been used to tackle the BAO problem, like the use of scoring methods, which assign scores to beam angles by considering geometric and dosimetric factors [21]; response surface approaches, which focus on generating beam data for promising directions to explore all potential beam orientations [22]; mixed integer programming approaches, that utilize treatment planning models incorporating two classes of decision variables to simultaneously capture the beam configuration and intensities [23]; amongst others. In addition to the aforementioned methodologies, metaheuristics have also been applied to address the BAO problem. For instance, simulated annealing [24,25], particle swarm optimization [26], and evolutionary algorithms have been utilized [16]. In the context of conformal RT treatment planning, Wu et al. [27] explored the use of a genetic algorithm to determine beam directions and intensities. It is important to note that while these global heuristics have the potential to avoid local optima, obtaining globally optimal or clinically superior solutions typically requires a large number of objective function evaluations [10].

Despite the BAO problem being the first one to be solved in treatment planning, its optimal solution will depend on the optimal solutions of the two other sequential problems (FMO and aperture optimization), being BAO specially dependent on the optimal solution of the FMO. So, since optimal beam angles for IMRT are frequently found to be counter-intuitive, the resulting beam angle set has no guarantee of optimality and has questionable reliability unless it takes into account FMO. It requires considerable time to calculate and produce the appropriate optimal FMO solution for a beam angle set, and even if only one beam angle is changed in that set, a full new dose computation is necessary [15,16], which explains the computational burden associated with BAO.

FMO is usually solved resorting to an optimization mathematical model, that has some parameters that are usually fixed *a priori* (like weights and lower/upper bounds). Actually, it is possible that these parameters should also be a function of the beam angles considered. For instance, if it is necessary to spare femur heads and the beam angles are going through these structures, then they should have adequate weights in the optimization model to be properly spared. However, if the beams chosen are such that radiation beams will not go through these structures, then lower weights could be considered. Although the choice of beams should clearly influence the choice of the FMO parameters, the majority of BAO techniques do not take into account any adjustment of these parameters for different beam angle sets, and FMO is only solved once for each beam angle configuration taking into account a predetermined set of weights and lower/upper bounds. The drawback of this approach is that the parameters used in FMO are not guaranteed to generate feasible plans for every BAO solution [20].

Dias et al. [20] developed an alternative approach. To ensure that the resulting treatment plan is clinically feasible, an automatic Fuzzy Inference System (FIS) FMO approach can be utilized, where the FMO model parameters are automatically tuned to optimize the solution. By adopting this approach, the BAO procedure aims to optimize the treatment plan while ensuring that the final solution is clinically feasible and can be implemented in practice.

Dias et al. [20] compared two approaches that optimize treatment plans. One algorithm used 7 noncoplanar beams and obtained solutions referred to as BAOFIS, while the other utilized the most commonly used 7-beam equispaced beam ensemble configuration and was referred to as FIS. In both algorithms, FMO was performed using the same FIS approach, indicating that the two alternatives are solely differentiated by the integration of BAO, and not by different

FMO approaches. The results demonstrate that target coverage is very similar considering both FIS and BAOFIS treatment plans, with all the calculated solutions presenting dosimetric values above the admissibility threshold, thus fulfilling the desired PTV coverage. Nevertheless, it is possible to observe that organ sparing is clearly enhanced in the case of BAOFIS treatment plans. Compared to FIS treatment plans, BAOFIS obtained an improved average sparing for the spinal cord and brainstem of 4.4 and 7.3 Gy, respectively.

BAO approaches considering Artificial Intelligence (AI) will be discussed in Section 3.4.2.

2.3.2 Fluence Map Optimization (FMO)

FMO can be defined as the problem of finding the optimal intensity of beam profiles to generate a high-quality plan [15]. It is frequently based on nonlinear continuous programming problems, requiring the planner to specify *a priori* weights and lower/upper bounds that are iteratively changed inside a trial-and-error approach until an acceptable plan is found [28].

A mathematical model is implicitly constructed in clinical practice by the explicit choices made by the planner. The planner uses a software, known as Treatment Planning System (TPS), where she/he can set weights and lower/upper bounds for each structure. The planner cannot, however, be certain of the underlying mathematical model that is being optimized because TPS optimization models and algorithms are black boxes from the user's point of view. This adds to the procedure complexity because it is not trivial to understand how a change in a specific weight or bound will affect the treatment plan that has been calculated. The planner is aware of the requirements a treatment plan has to comply with to be admissible, but it is not straightforward what information should be provided to the TPS in order for the optimization model to actually produce an acceptable plan. Furthermore, the planner knows that the medical prescription establishes thresholds, but it is often possible to go beyond those boundaries. If the planner believes that better outcomes (better tumor coverage and/or better OAR sparing) can be obtained, she/he will continue the process, being increasingly more demanding (as if the planner was changing the admissibility threshold established by the medical prescription). This trial-and-error procedure continues until the planner is satisfied or runs out of time, whichever comes first. Manually choosing the parameters that make up the underlying optimization model is far from intuitive and, depending on the planner's experience and time constraints, it can result in treatment plans of inconsistent quality [20].

The volume of each structure is discretized into small cubes called voxels, and the dosage is calculated for each voxel by considering the contribution of each individual voxel combined. Typically, a dose matrix D is built from the set of all beamlet weights, by indexing the rows of D to each voxel and the columns to each beamlet, so that the number of rows of matrix D equals the number of voxels (N_v) and the number of columns equals the number of beamlets (N_b) from all beam directions considered. Thus, using a matrix format, the total dose received by the voxel i can be given by $\sum_{j=1}^{N_b} D_{ij}w_j$, with D_{ij} the unitary dose delivered to voxel i by beamlet j and w_j the weight (intensity) of beamlet j. Defining S as the set of structures to be considered, and U_s the upper bound associated with structure $s \in S$, L_s the lower bound associated with the structure $s \in S$, $\underline{\lambda}_s$ and $\bar{\lambda}_s$, the penalty weights of underdose and overdose of structure s, respectively, the FMO model can be defined as follows:

$$f(w) = \min_{w} \sum_{s \in S} \sum_{i \in S} \left[\underline{\lambda}_s \left(L_s - \sum_{j=1}^{N} D_{ij} w_j \right)^2 + \bar{\lambda}_s \left(\sum_{j=1}^{N} D_{ij} w_j - U_s \right)^2 \right] \tag{2.1}$$

Generally, the total number of voxels considered reaches the tens of thousands; therefore, the row dimension of the dose matrix is of that magnitude. The size of D originates large-scale problems, being one of the main reasons for the difficulty of solving the FMO problem [10, 28].

The quality of the results can be perceived by considering a variety of metrics. One that is usually clinically used for plan evaluation is the volume of PTV that receives 95% of the prescribed dose (D_{95}). Typically, 95% of the prescribed dose is required ($D_{95\%} \geq 95\% D_P$) [15]. Regarding the OAR, the metrics vary depending on whether a serial or parallel OAR is being considered. Serial OAR are the ones that can have their functionality compromised even if a small portion of the OAR volume is damaged, so maximum-dose constraints are considered. Parallel OAR can keep their functionality even if a small volume target is damaged, so mean-dose constraints are used [20].

2.3.3 Realization problem

Once an acceptable set of intensity maps has been calculated, the delivery problem (also known as the realization problem) must be solved by selecting one of the existing appropriate methods ([29], [30], [31], [32], [33]) for creating apertures and intensities that approximate the previously determined intensity maps. Several papers propose algorithms for the realization of arbitrary fluence distributions by means of multileaf field segmentation, which involves superimposing differently shaped beams in order to freely shape the delivered fluence pattern [7].

It is a difficult optimization challenge to efficiently reproduce the optimized intensity maps. The intensity maps that are actually delivered may differ from the ones that were optimized due to leaf collision problems, leaf perturbation of adjacent beamlet intensities, and tongue and groove limitations, amongst others. Although most of those issues have been solved, the realization problem continues to be a thriving area of study [34].

There are other techniques, such as direct aperture optimization (DAO), that eliminate the requirement for a separate leaf-sequencing stage by including all of the MLCs constraints into the optimization process. DAO is designed to offer the dosimetric advantages of IMRT while maintaining conventional radiation therapy's simplicity and effectiveness [35].

2.4 Automation of the radiotherapy treatment planning: current state-of-the-art

IMRT treatment planning may be a lengthy and frustrating process and different treatment planners and institutions will provide different levels of quality treatment plans for patients with similar target dose prescriptions and normal tissue constraints [36]. Several attempts have already been made in order to contribute to the elimination of the treatment planning trial-and-error process, many of which involve the automation of the process [19]. Some of these attempts will now be presented, as they hold relevance to our current study.

Automating treatment planning has the potential for extreme time reductions in the treatment design procedure. The difficulty in planning eventually comes down to an issue of quality discrimination, that is how to assess good plans as “good” and bad plans as “bad”. There are countless ways for the treatment planning process to go wrong and produce a suboptimal planning result, such as errors in simulation (incorrect representation of the patient or of the radiation distribution in patients), errors in contouring (incorrect target or normal tissue delineation), errors in plan optimization (planner fails to meet achievable plan quality for the patient) and/or errors in delivery (treatment delivery differs from representation in the treatment planning) [19].

Zhang et al. [36] tested a methodology to automate the IMRT planning process for lung cancer. The methodology claimed to automatically set beam angles based on a beam angle algorithm, effectively designing the planning structures and automatically changing the objectives of the objective function based on a parameter automation algorithm. Using a database of treatment plan experts, the beam angles are chosen, resulting in the initial selection of 19 beam angles (14 coplanar and 5 non-coplanar). Then, based on physician/dosimetrist-contoured PTV, an automatic set of structures and optimization parameters is generated. For the optimization, a dose or dose-volume-based objective function was combined with an equivalent uniform dose (EUD)-based objective function. The main benefits of an EUD-based objective function over a dose-volume-based objective function include the fact that only one parameter (the target EUD) is changed in the objective function parameter automation loop, making it incredibly quick and simple; the EUD-based objective function is a convex objective, which makes the optimization algorithm well behaved, and optimizing it will optimize the entire DVH at the same time. The objective will be determined based on the prediction of whether or not the mean lung dose can be constrained to 22 Gy. The target plan was PTV-based, which results in a plan with high PTV coverage. Optimization parameters related to OAR can be changed throughout the process depending on the objective function values of the current solution. The 19 beams were ranked, and no meaningful difference was shown between using 19 and 11 beams, so the 11 best beams were chosen. This algorithm underlying principle implies that if fewer beams are better, the optimization algorithm should turn off the extra beams automatically.

Zarepisheh et al. [37] considered a treatment planning optimization based on an algorithm guided by the DVH curves of a reference plan. The percentage of delivered dose-containing volumes is related to those volumes by the DVH. The reference plan is selected from a library of clinically approved and delivered plans of previously treated patients with similar medical conditions and geometry. The algorithm navigates the vast voxel-based Pareto surfaces using a voxel-based optimization model. Voxel weights are iteratively adjusted to approach a plan that is similar to the reference plan in terms of the DVHs. Seven equispaced beam angles were considered.

An algorithm was proposed in [38], that optimizes both the intensity-modulated beam (IMB) and the normal tissue prescription. The IMB optimization employed a fast-monotonic-descent method, which has the advantage of quick and monotonic convergence to the minimum for a constrained quadratic objective function. To convey the vague understanding of the importance of matching the calculated dose to the prescribed dose in the normal tissue, a fuzzy weight function was used. This function partly expresses the multiplicity of the objective and the complexity of the planning problem in radiation therapy.

Jia et al. [39] based the suggested treatment planning process on an OAR-3D dose distribution prediction. The OAR-related constraints that support FMO are defined by the dosimetric values predicted, which take all the voxels within an OAR as research subjects, their doses as output, and the individualized geometrical features, including its location and volumetric information, as inputs. Using hard constraints to ensure PTV dose coverage, an artificial neural network (ANN) was used to first predict the dose distributions for OAR and use them as an objective goal to quickly guide the current dose distribution to the prediction as closely as possible. BAO is not considered. The dosimetric achievements of the current solution are not taken into account while updating the FMO objective function dynamically.

Dose prediction was also considered in [40], training a deep learning-based 3D dose distribution prediction and automatic plan generation based on the predicted dose distribution. The model input consists of CT images and contours delineating the OAR and PTVs. The algorithm output is trained to predict the dose distribution on the CT image slices. The obtained prediction model is used to predict dose distributions for new patients. Then, an optimization objective function based on these predicted dose distributions is created for automatic plan generation. Besides the DVH curves, this method also gives voxel-level feedback to planners about where the dose distribution could be improved. Then a voxel-by-voxel dose optimization using the predicted voxel is performed. This optimization does not require any specific dose–volume objectives.

A deep learning-based approach was used in [41] to build an IMRT plan by generating predictions of fluence maps using just patient anatomy. No inverse planning is required. The predicted fluence maps are converted into a deliverable treatment plan by delivery parameter generation and dose calculation in a commercial TPS, which means there is no need for an optimization phase.

Pencil beam fluence maps were proposed to be submitted to a fully automated prioritized multicriteria optimization (AUTO MCO) with integrated noncoplanar BAO in [42], followed by the generation of MLC segments. Instead of replicating the fluence maps, the segmentation algorithm then faithfully reproduces the AUTO MCO 3D dose distribution while taking into account every potential beam at once. A stand-alone version of the clinical dose calculation engine is used to calculate pencil beams, segment dose depositions and final doses.

Based on a set of cost functions that are either established as hard constraints or planning objectives with assigned priority and goal values, Bijman et al. [43] generated plans using a wish-list that defines the protocol for automated plan generation. Planning objectives are sequentially optimized following their priorities while never violating all established constraints. To guarantee that the previously acquired function value is maintained while minimizing lower-priority objectives, a new constraint is applied to the optimization problem after each objective function optimization. Wish-lists are treatment site-specific and are created in an iterative tuning process with the treating physician. This final strategy is based on Erasmus-iCycle, a prior strategy created by the authors. To employ this procedure, the wish-list must first be defined, so that it can be used on all patients with the same tumor site. In Erasmus-iCycle, the selection of beam angles is based on a greedy iterative selection procedure, where one beam is fixed at each iteration. As a result, the search space for each new beam is greatly influenced by prior decisions. Breedveld et al. [44] proposed a novel *a priori* multicriteria approach to integrate beam angle and intensity optimization, which may be a drawback if a patient that differs from the most common cases is being treated.

Protocol-based automatic iterative optimization was used to develop automatic plans in [45]. Pinnacle3 16.2 Auto-Planning was used in the automatic planning approach. The authors state that more complex protocols containing additional optimization targets and support structures were necessary because protocols that only contained the prescribed target dose and OAR limitations frequently did not produce optimal plans. Depending on their geometry and restrictions, different tumor sites required different techniques.

Two separate automated engines were employed in the Pinnacle treatment system in [46]. These two automated engines, the currently used Autoplanning and the new Personalized, are both template-based algorithms that use a wish-list to build the planning goals and an iterative technique capable of simulating the planning process typically used by experienced planners. Personalized algorithms present an advanced technology called Feasibility which allows an estimation of the best possible sparing of the OARs to inform the planner *a priori* about the achievability of treatment planning goals.

In order to evaluate how well-automated planning will perform in comparison to current manual processes, the primary validation procedure is the deployment of the automated planning routine to a representative sample of previously treated patients. A well-set-up automated planning system should consistently balance the same clinical trade-offs across all patients, which is one of its selling benefits. However, if new trade-offs or information need to be considered, this could be a drawback. A crucial issue to keep in mind while utilizing an automated planning system is that an incorrect dose prediction could occasionally lead to the optimization being misled [19].

In this work, the FMO approach that is going to be considered is a totally automated FMO based on Fuzzy Inference Systems described in [28]. Since FMO is highly important for the Q-learning approach developed, the FIS FMO is now described in the next subsection.

2.4.1 Fuzzy Inference Systems for FMO

Fuzzy logic is used in this work for the automation of FMO. A fuzzy inference system can be used to iteratively change parameters associated with the PTV and OARs in the FMO mathematical optimization model in order to achieve desirable doses. The weights and lower/upper bounds are automatically changed by this approach depending on how distant the present solution is from a desirable one or, in other words, how far the current treatment plan is from the medical prescription. The objective function typically considers a weighted sum of dose deviations [28]. The FMO objective function evaluates the sum of the weighted squared difference between the dose delivered to each voxel within the structures of interest (such as PTVs and OARs) and the dose desired. The optimization parameters include the weights and lower/upper bounds associated with each structure of interest, which are the ones that need to be adjusted, as illustrated in FMO model (Figure 2.1) [20]. It is not straightforward to understand or estimate the impact that changing one or more of the existing parameters will have on the quality of the generated treatment plan. As we already stated, in practice planners spend a lot of time testing the FMO model through a process of trial-and-error until they reach a treatment plan with the best possible quality [20].

Fuzzy logic can represent a great tool for the FMO problem. Fuzzy logic allows the creation of sets with unclear boundaries so that a given element can belong to a set with only a partial degree of membership. The concepts and their relationships will be represented by a membership function, which can vary between 0 and 1, that must represent the concepts as well as the relationships between the concepts [28]. This fuzzy inference system implements a set of rules that decide which parameters should change, as well as the magnitude and direction of this change, taking into account the distance between the current plan (corresponding to the optimal solution of the current FMO model) and the constraints defined by the medical prescription [20]. All rules are assessed at the same time [13]. The purpose of this system is to mimic, up to a certain level, the decision-making process of a planner. If the current FMO optimal solution for a specific structure of interest is still far from the desired result, then the parameters linked to that structure must be modified to increase its importance in the optimization process [20]. It is possible to change the importance of a structure in FMO in two different ways: increasing the weights associated with that structure or changing the corresponding bounds. According to [20, 28], changing bounds first produces a smoother iterative process, that converges faster.

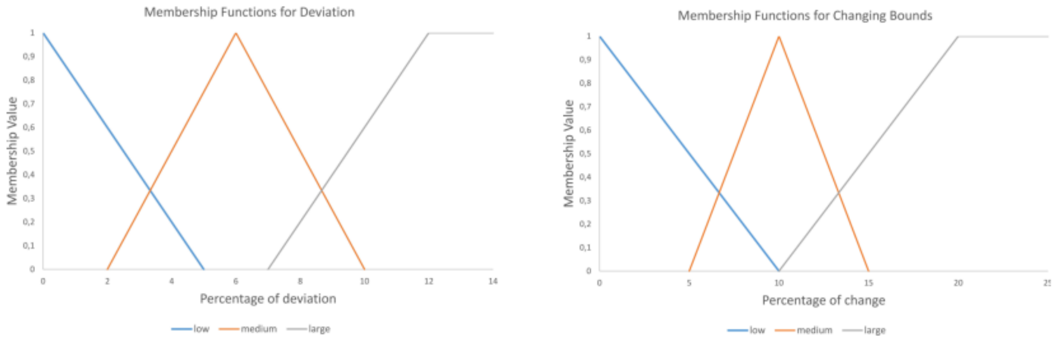
In its initial state, two different phases constitute the Fuzzy Inference System (FIS) procedure. The first phase entails the computation of a treatment plan that complies with all constraints defined by the medical prescription. The second phase recognizes the possibility of improving the quality of the treatment plan, by enhancing PTV coverage and/or better sparing OARs. Therefore, this phase takes a more demanding approach than the initial medical prescription by trying to establish lower values for the tolerance doses for OARs and/or higher values for the dosimetric values associated with the PTV s [20]. This process is repeated until either the predetermined number of iterations is reached, or it is no longer possible to enhance the current solution [28].

Inferior and superior limits of dose for each structure are automatically adjusted by the algorithm according to fuzzy logic, until a plan that complies with all the restrictions is found, for the PTV and the OARs. This fuzzy inference system represents a set of simple rules that in a way follow the manual and iterative process normally executed by the planners. Thus, for each structure of interest:

- If, for a given structure, the deviation between the prescribed and the delivered dose is low, then the parameters should only be slightly changed.

- If, for a given structure, the deviation between the prescribed and the delivered dose is medium, then the parameters should suffer medium changes.

- If, for a given structure, the deviation between the prescribed and the delivered dose is large, then the parameters should be largely changed.



(a) **(b)**

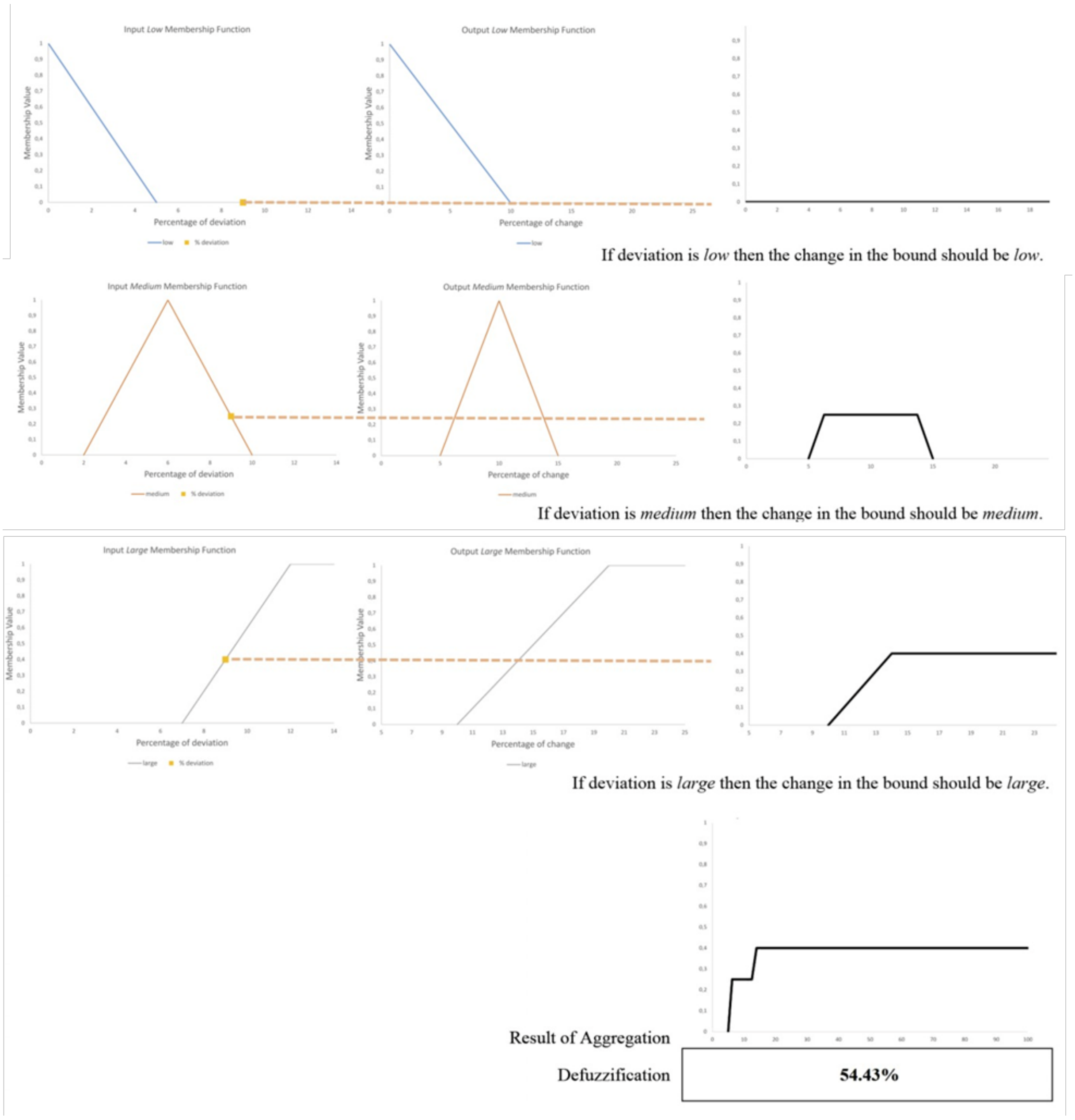
Figure 2.3: Membership functions

In fuzzy logic, the concepts "small", "medium" and "large" can be represented by a membership function. According to this logic, any given element does not belong to just one determined group, meaning that it does not abide by a binary logic of "belongs" or "does not belong". A given deviation between the prescribed dose considered and the delivered dose can simultaneously belong, with only a partial degree of membership, to different membership functions that represent each set. This means that a given deviation can be considered, simultaneously, as being small and medium with different degrees of membership. After evaluating the degree of membership for each concept, every rule mentioned above is activated with a different level of membership, which is then necessary to calculate a single crisp value that will be the output of the FIS. This process is known as *defuzzification*, and it was implemented by the centroid calculation defuzzification procedure. After the degree of membership in each function and the degree to which each rule is activated are evaluated, the area under each curve is aggregated and the centroid value of that region is calculated. In this work, triangular and trapezoidal memberships functions were used and a fuzzy inference system was created based on each structure, taking into account the structures objectives.

Figures 2.3a and 2.3b illustrate the membership functions considered for determining the changes in lower and upper bounds. As shown, there are some values in which the percentage of deviation only belongs to a specific set: for the input function, deviations belonging to [12%,100%] are considered large with membership equal to 1. However, deviations of 8% belong simultaneously to the medium and large sets, which means the values of the respective membership functions are higher than zero.

Figure 2.4 exemplifies the rules that constitute a fuzzy inference system, for a deviation of 9% that belongs simultaneously to the medium and large input membership functions.

In [20], the authors propose using three fuzzy rules to ascertain the extent to which deviations between desired dosimetric values and those obtained by a given solution determine the changes in the bounds. Specifically, small, medium and large deviations are found to produce small, medium and large changes, respectively. These deviations are represented as fuzzy numbers and each concept is represented by triangular or trapezoidal membership functions, which allows one to handle imprecise or uncertain information in decision-making and control systems. Output fuzzy sets are then obtained and, by a defuzzification step that is applied, all output



If deviation is *low* then the change in the bound should be *low*.

If deviation is *medium* then the change in the bound should be *medium*.

If deviation is *large* then the change in the bound should be *large*.

Figure 2.4: Defuzzification process

fuzzy sets are aggregated producing a crisp value, which will represent the percentage of change of the corresponding bound.

Fuzzy logic, and more specifically fuzzy inference systems (FIS), can potentially be used to optimize other parameters in inverse planning such as the beam orientation and the dose prescription. As the configuration of FIS is flexible, it provides us with a wide space to customize the FIS for different applications. Fuzzy logic has also been used in various stages of treatment planning. Dias et al. [28] created a completely automated FMO technique where all the objective function parameters are adjusted based on fuzzy inference systems that resemble the rationales planners use on a daily basis. Only beam-equidistant solutions are taken into account in this situation. The system analyzes how far the current solution is from a desirable one, changing in a completely automated way both weights and lower/upper bounds.

Yan et al. [47] also used FIS to improve normal tissue sparing. Three main modules that made up the FIS principle were defined: the Fuzzifier, which represents the membership function specified for the inputs, the Inference Engine, which implements the operations of inference using fuzzy rules, and finally the Defuzzifier, which represents the membership functions specified for the outputs.

One major advantage of using fuzzy inference systems for treatment plan optimization is that it eliminates the need for human interaction during the optimization process. The planner is only required to define constraints based on the medical prescription and, if desired, the priorities for each structure, before the start of the optimization. This allows for a more efficient and streamlined optimization process.

3

Artificial Intelligence in Radiotherapy

This chapter presents an overview of the current state-of-the-art of Artificial Intelligence (AI) in Radiotherapy (RT). Section 3.1 will provide information about the basis of AI and its concepts. Next, Section 3.2 explores Reinforcement Learning (RL), the learning process used in our study. Section 3.3 will describe the commonly used techniques and features of Machine Learning (ML). Furthermore, Section 3.4 presents all the areas of implementation related to RT regarding AI, as well as a deeper look at the various studies that implemented the discussed methods in the respective areas.

3.1 Introduction

Artificial intelligence is a viable option for RT since it can be helpful during the various treatment planning and delivery stages. All tasks the planner performs manually in the current workflow, such as selecting the number of beams and their directions, can be replaced or supported by AI algorithms. This integration of AI in the workflow has the potential to not only improving the accuracy of treatment planning but also to significantly accelerate the RT workflow process.

AI can be defined as the capability of a machine to mimic human intelligence. It can be classified into two branches based on its application: virtual and physical. The physical component can be represented in medical devices or sophisticated robots, whilst the virtual component can be represented in ML [48]. AI has the potential to optimize the various stages of the complex process of RT; however, it is often perceived as a “black box”. As a result, it can be challenging to integrate AI into clinical practice as human operators may only comprehend input and output predictions [49].

ML refers to the concept of a machine gaining the ability to do tasks based on pratical learning. ML models are usually divided into four main groups: supervised learning, unsupervised learning, semi-supervised learning and RL.

While unsupervised learning seeks to identify the unknown pattern underlying the observation or identify the relationship between samples, supervised learning seeks to develop a functional relationship between inputs and outputs from training data that generalizes to testing data [48].

The semi-supervised learning (SSL) approaches were developed to address the issue of having limited availability of labeled data, which can significantly impact the effectiveness of supervised learning approaches. The main challenge lies in developing ML algorithms that can extract knowledge from existing data. Traditionally, supervised learning algorithms train classifiers using labeled samples and explanatory attributes. The objective is to construct a model, such as a predictor or a classifier, that can accurately predict class labels for samples where only attribute information is available. However, supervised learning approaches require large labeled datasets to generate more accurate classification rules. In contrast, semi-supervised learning leverages both labeled and unlabeled data to construct a decision rule, enhancing the performance of classifiers trained solely on the labeled data. In situations where datasets are small and imbalanced, SSL classifiers exhibit slight improvements compared to the use of supervised learning methodologies. The inclusion of unlabeled data enriches the information extracted from the labeled samples, leading to enhanced performance of the classifiers. Overall, semi-supervised learning provides a promising approach for addressing the limitations posed by small and unbalanced datasets, allowing for more effective and accurate classification in various applications [50].

RL is considered a learning process that evolves based on the feedback from previous actions. As it is the main focus of this work, it will be detailed in Section 3.2.

As for assessing the effectiveness of a model, different resampling methods have been employed. The most common types include k-fold cross-validation, leave-one-out cross-validation, Monte Carlo cross-validation and bootstrapping.

In k-fold cross-validation, the existing dataset is divided into k subsets. The learning step takes into account all these sets but one, that is used for testing. The assessment is done considering the testing results for all k subsets. Leave-one-out cross-validation, a variant of k-fold cross-validation, chooses one element from the training set is in each iteration, and the remaining training set is used to train the models. The trained models are then employed to predict the class of the selected element. This process is repeated for each component of the training set and the model accuracy is calculated as the percentage of correct predictions [11].

The bootstrap method involves using the available sample data as a “surrogate population” to approximate the sampling distribution of a statistic. It creates numerous “phantom samples”, known as bootstrap samples, by resampling (with replacement) from the original data. The sample summary is then computed on each bootstrap sample, typically generating thousands of samples [51].

Each sampling scheme possesses its own characteristics in terms of variance and bias. For instance, the bootstrap method tends to have low variance but significant bias, while k-fold cross-validation exhibits small bias but higher uncertainty, depending on the value of k [52]. The measured performance of classifiers varies depending on the validation method employed.

3.2 Reinforcement Learning

Reinforcement learning requires learning what to do (how to connect situations to actions) in order to maximize a numerical reward signal [53]. The absence of input/output pairs is the primary distinction between supervised learning and RL. Instead, after making a decision, the

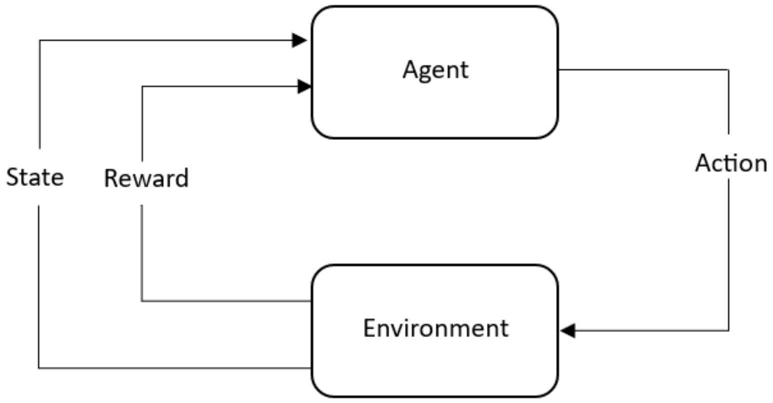


Figure 3.1: Schematic process of reinforcement learning

agent is informed of the immediate reward and subsequent state, but not of the action that would have been better for it in the long term. For the agent to behave optimally, it must actively gather useful experience regarding the potential system states, actions, transitions, and rewards [54].

The main goal of RL is to foresee the optimum course of action given each situation. RL offers resources to optimize a series of decisions for long-term outcomes. The input for RL algorithms is typically a history of interactions between the decision-maker and their environment via perception and action, as shown in Figure 3.1. A Markov decision process is usually used to model it, with a set of environment states and actions employed to train an artificial agent to maximize its cumulative expected rewards. The algorithm selects an action at each point by its policy and receives new information as well as immediate outcomes (reward). Formally, RL models consist of:

- a discrete set of agent actions, a;
- a discrete set of environment states, S;
- a set of scalar reinforcement/reward values, r.

The training process often involves an exploration-exploitation trade-off. Exploration refers to exploring the whole space to learn additional information and potentially discover better strategies. It is vital for long-term performance development since it allows the agent to gain a better awareness of the environment and discover actions that might lead to higher rewards. Exploitation refers to exploring the prospective areas based on current data. It means implementing the learned policy to exploit previously known actions in order to maximize short-term gains [4, 48].

One of the most common RL algorithms, and the one explored in this work, is Q-learning. Further information about Q-learning will be presented in Subsection 3.3.4.

3.3 Machine learning methods

Numerous ML methods could be used in RT, including ensemble learning, cluster analysis, K-means, linear feature extraction, deep learning (deep neural networks and convolutional neural networks) and linear models for classification and regression, amongst others [15]. These algorithms are very appealing for RT because they could significantly reduce time-consuming operations in segmentation and planning, reduce deviations from expected dose distribution caused by treatment delivery issues and improve the predictability of adverse effects of RT [49]; however, they still have a lot of drawbacks: they are very prone to make some mistakes that a human would not make, they require large data sets and they always need to be extensively trained and tested for accuracy before being implemented clinically [48].

Every model is adjusted during the training/validation process to meet clinical demand. Availability of detailed information concerning the model intended use and limits, description of training and validation set, used standards, metric and overall validation protocol is highly recommended [55]. In the test phase, the model ultimate performance is evaluated independently, its robustness is examined and the patient types to which the model can be applied to are determined. This phase should be applied to all AI models used clinically. The model is assessed both qualitatively and quantitatively, using an independent dataset that should accurately reflect the data for which the model will be used clinically and exhibit similar variation as in the training data.

An in-depth look at the multiple ML algorithms and their applications to RT and medical imaging will now be provided, presenting some of the existing ML algorithms.

3.3.1 Markov random field (MRF)

MRF is a conditional probability model, where the probability of a pixel is affected by its neighboring pixels. MRF is a stochastic process that uses the local features of the image. It is a powerful method to connect spatial continuity due to prior contextual information [56].

3.3.2 Artificial Neural Networks (ANN)

The artificial neural networks (ANN) are a very streamlined representation of the human brain. They are made up of nodes and interconnections, where nodes, despite having little computational capability, act as a triggering mechanism that builds up to activate a neuron in a mechanism comparable to how neurotransmitters work. The output of a node is determined by the weighted sum of its inputs and the network is trained by incrementally altering the weights in an effort to reduce an error function. The major advantage of ANN is their ability to tackle a wide variety of problems, including those that are not linearly separable ([48], [11]).

3.3.3 Deep learning

Multiple processing layers are used in the deep learning approach to find patterns in a vast amount of data. Deep learning is a subfield of AI and a group of computational models made up of several data-processing layers. The system calculates the error between the observed output and the desired output during the training phase and modifies its internal parameters, known as weights, to minimize this error. Additionally, the system calculates a gradient vector for every weight that shows the error deviation as a result of weight adjustment. The weight vector is altered in the gradient vector's opposite direction. Recently, a variety of deep learning-based networks, including deep neural networks and convolutional neural networks, have been developed and used in RT [48].

3.3.3.1 Deep Neural Networks (DNN)

There are supervised and unsupervised deep neural networks (DNN). DNN individually learn the order representation of the input data, requiring a significant amount of data for effective learning. Multiple neural networks or multilayered ANN are combined to form DNN. The output from the first layer becomes the input of the next layer, and so on, with the final layer's output being the system's derived output. DNN compute the data using nodes initially, using the same ANN principles. Due to the requirement for less labeled data, unsupervised deep learning techniques are chosen over supervised DNN. In unsupervised DNN, it is more difficult to ensure that the learned representation will be meaningful, which comes as a disadvantage [48].

3.3.3.2 Convolutional Neural Networks (CNN)

The convolutional neural networks (CNN) are suitable for processing data that is presented in arrays, such as medical images. The three main types of layers found in CNN are the convolutional layer, the pooling layer, and the fully connected layer. The functions of convolutional layers include learning feature representations of the input and spotting similarities between features from prior layers. The quantity of convolution kernels utilized in the computation of feature maps is represented by the number of convolution layers. The role of the pooling layer is to achieve shift variance by lowering the resolution of the feature map. Each feature map of the pooling layer is coupled to a corresponding feature map of the preceding convolutional layer, which is often positioned between two convolutional layers. The purpose of the fully linked layer is to do high-level reasoning by connecting each and every neuron in the current layer to every single neuron in the preceding layer [48].

3.3.4 Q-learning

Q-learning is a value-based RL algorithm that aims to learn a Q-function that models the action-reward relationship [4]. Q-learning consists of an agent learning an optimal policy in a Markov decision process without having any prior knowledge of the environment. It is based on the Bellman equation (3.1), which is defined below:

$$Q(S_t, A_t) \longleftarrow Q(S_t, A_t) + \alpha \left[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t) \right] \tag{3.1}$$

$Q(S_t, A_t)$ represents the current action, α the learning rate, R_{t+1} the reward, γ the discount rate and $\max_a Q(S_{t+1}, a)$ the maximum expected future reward given the new state and possible actions at that new state.

In Q-learning, a Q-table that helps determine the best action for each state is created, which is initially empty. The agent first chooses its action at random and the corresponding Q-value is calculated according to the equation, which iteratively updates the Q-table. The expected reward is maximized by selecting the best of all possible actions.

In practice, the Q-learning method entails the agent exploring to make better future action selections and exploiting what it already knows to obtain a reward. It begins by exploring the environment and updating the Q-table. When the Q-table is complete, the agent will begin to exploit the environment and take better actions, following the policy of selecting the action corresponding to the highest value [53].

The Q-learning algorithm is shown in procedural form as follows:

```
1. Initialize $Q(s,a)$, $\forall s \in S, a \in A(s)$, arbitrarily, and $Q(terminal\text{-}state,.)= 0$
2. Repeat (for each episode):
    (a) Initialize $S$
    (b) Repeat (for each step of the episode):
        i. Choose an action for the current state using policy derived from $Q$ (e.g., E-greedy)
        ii. Take action $A$, observe $R$, $S'$
        iii. $Q(S,A) \leftarrow Q(S,A) + \alpha\Big[R + \gamma \max_a Q(S',a) - Q(S,A)\Big]$
        iv. $S \leftarrow S'$
    (c) until $S$ is terminal.
```

This means the algorithm operates as follows, in the exploitation phase:

1. Set the current state as the initial state.
2. From the current state, find the action that produces the maximum Q value.
3. Set the current state as the next state.
4. Go to step 2 until the current state is the goal state.

The best way to conceptualize Q-learning is as a stochastic approximation technique for computing the Q-values. No explicit expected values are computed by the technique, even though the definition of the optimal Q values for each state depends recursively on the expected values of the Q values for the following states (and on the expected values of rewards). Instead, utilizing the real stochastic mechanism that generates succeeding states, iterative sampling is used to approximate these values [57].

3.4 Areas of implementation

This Section provides an overview of the opportunities for AI in each step of the RT treatment workflow, as well as various studies that implemented the previously discussed ML methods in each of those steps.

3.4.1 Medical imaging

Modern medical imaging systems, such as computed tomography (CT) and magnetic ressonance imaging (MRI), use a variety of machine-learning techniques. CT is a type of radiological imaging that uses a medical image to gather volumetric and morphological data about the anatomy of the patient. MRI can differentiate soft tissue extremely well, making it ideal to examine a joint or ligament; however, MRI may be used to obtain images practically anywhere on the body, if it involves soft tissue density difference [48].

Each of these imaging technologies has drawbacks, such as radiation exposure, sensitivity to its surroundings and expense. The two areas of medical imaging where AI-based algorithms are most commonly used are illustrated below [48].

3.4.1.1 Medical image registration

In RT, images from several patients, times, or modalities usually need to be registered in order to combine their relevant data in a common coordinate. It is frequently necessary to do multimodal image registration to improve the visibility of organs or tissues compared with what is obtained with a single image modality [58]. To aggregate the results and produce a more precise diagnosis, ML algorithms are applied. This process is often referred to as image fusion, matching, or wrapping. The objective of this procedure is to identify the optimal transformation that best aligns the structures of interest in the input images [48]. Typically, to solve the image registration issue, image-based and biomechanical methods are created. In the image-based method, the original image is iteratively morphed to achieve a desirable match with the target image using a deformation engine. The biomedical method involves first segmenting images into organs and assigning known elasticity coefficients to each one [58].

Cao et al. [59] provided a non-rigid inter-modality registration framework based on deep learning, in which the intra-modality image similarity metric is skillfully transferred to train an inter-modality registration network. A flexible and practical alternative to conventional optimization-based algorithms is offered for the difficult non-rigid inter-modality registration problem, specifically without iterative optimization and parameter modification in the testing phase. Additionally, the dissimilarity loss is calculated in a dual method on the MR modality and the CT modality, respectively, to train the network more robustly and take advantage of the complementary anatomies from both modalities. CT and MR registration is performed, and the results showed promise in terms of accuracy and efficiency.

3.4.1.2 Image segmentation

In medical image analysis, image segmentation is crucial. Inverse optimization in Intensity-Modulated Radiation Therapy (IMRT) requires the delineation of all the volumes of interest. One of the most laborious tasks in RT is manual segmentation. In addition, manual segmentation has been demonstrated to have significant intra- and inter-observer variability [60] due to the non-uniform training and time constraints for planning. For example, with adaptive RT, when a new IMRT plan must be quickly developed, the time-consuming segmentation process is incompatible with clinical practice [58].

The fundamental method for evaluating AI for segmentation is centered on training a ML system, then comparing its performance to a gold standard using established criteria for overlapping comparison. According to numerous studies, auto segmentation considerably decreases contouring time while maintaining accuracy to routine inter-observer variability.

Two popular techniques for segmenting images are the snake model and the level set method (LMS). LMS can be further separated into two groups: edge-based models and region-based models. The region-based model controls the motion of the active contour using region information as opposed to the edge-based model, which uses edge information [48].

Segmentation can be applied to many structures, such as bones, organs, muscles and fractures. In recent research, the tree-based segmentation approach has been studied intensively for brain imaging. There are multiple methods discussed in [56], involving supervised and unsupervised segmentation. Compared to deep learning techniques, traditional ML algorithms like Markov random fields, k-means clustering, random forests, etc., are frequently less accurate, but they are frequently more sample efficient and have simpler structures. Several deep-learning networks provide excellent results for the segmentation of medical images. Additionally, the outcomes of deep learning are on par with those of expert manual segmentation. Deep learning achieves 97.31% accuracy compared to 96.29% from active contouring, and 96.74% from the graph cut.

3.4.2 Treatment planning

The ability to predict *a priori* acceptable dose distributions is one of the most addressed problems in the literature regarding AI implementation in RT planning.

RT treatment planning is typically a time-consuming process. Despite the consistent planning objectives, the planning outcomes vary due to anatomical patient diversity. Early in the planning phase, the outcomes cannot be predicted. The dosimetrist tunes a large number of optimization parameters during the planning stage without knowing the endpoint. Different institutions and individual planners frequently use inconsistent and suboptimal plan dosimetry. To solve these issues, knowledge-based planning (KBP) and automated planning techniques have been created [58].

KBP is motivated by the observation that the feasible patient dose is strongly correlated with the anatomy. For instance, a critical organ receives a larger dose the closer it is to the tumor. To learn the correlation between patient anatomies and planning dose, Wu et al. [61] presented the concept of the overlap volume histogram and established its relationship with the dose-volume histogram (DVH). The dose was predicted using a variety of ML techniques, including support vector regression. Artificial neural networks, which mimic the information flow and processing of biological systems, are employed in addition to direct regressional learning methods to predict dose distributions, showing similar performance only for simple cases. However, the prediction performance worsens for large regions of interest and complex cases [58].

These traditional methods frequently suffer from slow performance, sensitivity to parameter adjustment, low accuracy in complex scenarios and increasing requirements on the training dataset with more features included. With very straightforward training procedures, deep learning can learn implicit anatomical, imaging and dosimetric features, making this particular RT problem a perfect research opportunity [58].

The predicted dose can be used to partially or completely automatically guide treatment planning. The predicted dose can be used to extract DVH constraint points for use in commercial planning tools, or the optimization can be driven by 3D voxel doses [58].

DNNs can also be employed to automate planning. A DNN was developed in [62] without inverse planning to produce beam fluence maps directly from organ contours and volumetric dose distributions. Organ contours, including planning target volume (PTV) and organs at risk (OAR), and dose distributions, as seen from a single beam's-eye-view (BEV), were used as input data for the DNN's training phase, and the fluence map for each corresponding beam direction was employed as the intended output data. The trained network provided fluence maps within a second. Producing fluence maps directly from organ contours and dose distributions will improve the efficiency of the treatment planning phase by reducing the time needed to obtain optimal fluence maps and helping to preserve the quality of treatment plans.

Dias et al. [16] suggested an interesting method for optimizing beam angles. This work considered the number of beams to be determined *a priori*. It is necessary to evaluate the quality of each set of beams. Only once the Fluence Map Optimization (FMO) problem has been solved can this assessment be made so as to take into account the optimized beamlet intensities for each beam. In order to determine the optimal beamlet intensities, a voxel-based nonlinear model was applied. The FMO problem is always computationally expensive to solve for each set of beams, thus to get around this problem, the research suggests a surrogate model that will be able to estimate the actual objective function value in a very small fraction of the time it takes to calculate its true value. A patient-specific NN is developed, which will take sets of angles as inputs and provide, for a single patient, the value of the FMO objective function as output. The true value of the objective function, f, is determined for each set of k randomly generated angles by determining the optimal solution of the FMO problem. The neural network is then trained using these samples. The trained neural network is then prepared to calculate a surrogate function with the expectation that the result will be as near to the objective function as possible. Instead of using a single NN, 20 different NNs are used, in order to decrease the error with the increase in the number of available training samples. A genetic algorithm that views the Beam Angle Optimization (BAO) problem as a combinatorial optimization problem is used, where the interval of all possible angles is discretized into 360 possible degrees. After that, the algorithm repeats a sequence of operations until m new samples have been produced. Genetic algorithms are not well suited for clinical practice due to the computational time required to evaluate each individual (solution), especially when there are few computational resources available. Due to its goal of determining the fitness of most individuals in the population, the surrogate model (a trained neural network) is utilized to get around this issue. The computational results demonstrate that combining genetic algorithms with surrogate models can be a captivating path to follow.

The success of algorithms such as CNN in image processing and the learning capability of modern ML techniques enable treatment planners to provide patient-specific plans by leveraging the patient's anatomical features and learning from the optimization methods or behaviors of physicians. A fast and adaptable solution for the BAO problem was suggested in [15], employing DNN, which provides a solution in a matter of seconds, making it feasible to use it in clinical settings for cancer patients to fasten the treatment planning process. The proposed DNN approach recognizes the link between the patient's anatomy and the optimal set of beam orientations, based on anatomical features and an optimization algorithm, having the desirable ability to anticipate a set of beam orientations without prior knowledge of dose influence matrix values.

3.4.3 Radiotherapy delivery

High precision requirements are needed for modern RT and techniques that are able to anticipate dose distribution variations that may occur during therapy may boost delivery assurance and raise the overall quality of treatments administered. In simple terms, AI could be used to create a prediction of the dose that is actually supplied to the patient. As an example, the disparity between the planned and delivered movements of multileaf collimators is a significant cause of deviation that, if predicted, can be taken into account during treatment planning. To forecast these disparities from the plan data (such as leaf position and velocity, movement towards or away from the isocenter of MLC, etc.), a machine-learning approach has been created. Results revealed that a more accurate depiction of plan delivery would lead to a closer agreement in terms of dose volumetric parameters between the planned and the delivered treatments [49].

Another recent study examined log file data of 10 patients who had dynamic intensity-modulated RT [63], examining variables including leaf planned position, dose fraction, leaf velocity, leaf moving status, and leaf gap. To forecast *a priori* leaf positional deviations, a ML methodology was developed. With a perfect correlation coefficient (R = 0.999), the results showed that predicted leaf positions at control points closely matched delivered positions [49].

Discrepancies in the movements of multileaf collimators (MLCs) during RT can introduce errors in the distribution of radiation dose. To provide a more accurate representation of the actual dose delivered to the patient, a method was proposed in [64]. This method incorporates predicted MLC positional errors into the treatment planning system, giving the treatment planner a realistic view of the dose distribution. To predict these errors, planned and delivered MLC positions from a series of volumetric modulated arc therapy (VMAT) plans were collected. The differences between the planned and delivered positions were calculated. Additionally, leaf motion parameters, which were hypothesized to contribute to MLC errors, were computed for the plans. A ML model was developed using these parameters as inputs to predict the errors between the planned and delivered MLC positions. Improving optimization routines for encoding MLC leaf positions can enhance the calculation of dose distributions, providing a more realistic representation of the actual dose delivered to the patient.

3.4.4 Radiotherapy verification and patient monitoring

The precision and placement of each radiation beam are crucial to IMRT. The most common technique for evaluating the fidelity of IMRT is gamma analysis. The measured dose distribution and the planned dosage distribution are compared using the gamma statistic. Gamma analysis is insensitive to minor inaccuracies in multileaf collimator positioning and does not correspond to many clinically significant variations in delivered dosage.

A technique was created to identify particular errors using image features in gamma image. The gamma distributions are treated as images and use feature evaluation on the patient image to predict prognoses, therapeutic response and other outcomes [48].

Two popular graph models are the Markov random field and the Bayesian network. The Bayesian network model was employed by Smith et al. [65] to find faults in RT treatment planning. The network, a set of initial clinical data and a radiation oncology-based clinical database system were used to calculate the likelihood of getting specific RT parameters. When the network's performance was compared with the work of human specialists, the network outperformed them in the case of brain cancer. Physician order errors in external beam radiation can be found by the Bayesian network method [48].

3.4.5 Quality Assurance (QA)

Artificial intelligence can also play a vital role in quality assurance (QA). One way of looking at the relationships between QA and AI is to distinguish between the function of AI in enhancing QA practices and the requirement to utilize QA practices to ensure the security of machine-learned processes.

According to medical physicists, "quality assurance" refers to more than just lowering the incidence of obvious errors in radiation planning or delivery. QA is a holistic process that must include components of a patient's treatment requiring expertise regarding the patient and her/his condition as well as the best way to administer therapeutic radiation [18].

The most frequent QA tasks entail checking the results of a creative process and fit into the workflow of "human creates $\longrightarrow$ human verifies". When ML algorithms are incorporated into the process, computational machines can be used to construct a "human creates $\longrightarrow$ machine verifies" workflow, for example at the conclusion of a planning process, linac operation or dose computation. A point has been reached where software systems are capable of performing the creative task on their own, for instance, "machine creates $\longrightarrow$ human verifies", acting under human supervision [18].

Largely, the goal of assessing a variety of treatment plan components prior to radiation delivery is to seek obvious errors and the appropriateness of the plan. While rule-based techniques are effective in finding errors, they have drawbacks in terms of adaptability, efficiency and the capacity to reason in grey areas. Thus, although rule-based systems are efficient at detecting some faults, the difficulty of coding the rules for these exceptions quickly limits their capacity to manage anomalies.

Several ML-based approaches have been explored to catch some of the outliers that cause errors in the appropriateness of the plan. A K-means clustering algorithm was used in one application to achieve potential error identification by learning from historical prostate patient plans [18].

KBP can be used to lessen the burden of planning by offering a preliminary knowledge base of clinically treated and/or approved prior plans as a basis for comparison to a current patient of interest. The aim to enhance planning frequency in order to adapt to daily changes in physiological features as they happen throughout a treatment course is an important factor for automated KBP. Convolutional neural networks, principle component analysis and other ML techniques have recently been developed to investigate how to achieve this more individualized approach. The need for more frequent QA checks leads to a third paradigm where one machine creates and another performs validation, as in “machine creates ⟶ machine verifies” [18].

The algorithmic approach to QA evaluation and error detection can be used to assess machine performance to identify linac failure modes or identify random or systematic errors in delivery. AI-based QA techniques can estimate or classify ambiguous or potentially inaccurate segmentations and then provide them to experts for revision [55].

The assessment of ML product quality is a captivating and continuously evolving matter. Certain fundamental principles have emerged when evaluating models, regardless of whether they were created using ML techniques or not. Various modeling approaches have been developed, typically involving the division of data into separate modeling and testing datasets. Furthermore, addressing incomplete data, as mentioned earlier, is crucial as it signifies a fundamental gap in understanding the domain or the problem at hand.

ML models have shown significant potential in various aspects of radiation oncology, offering opportunities for substantial improvements. These ML applications are anticipated to lead to several advancements in the near future:

1. Development of Third-Party and Integrated QA Tools: The success of ML applications will likely drive the creation of third-party QA tools as well as integrated QA tools. These tools will enhance the utilization of resources and improve the quality of existing QA tasks, such as automated plan creation, plan checking, and plan measurement evaluations.

2. Integration of ML QA into Machine-Integrated Applications: ML QA techniques will be incorporated into machine-integrated applications, enabling the transition to new QA paradigms required for adaptive planning.

In the long run, it is expected that ML researchers will contribute to the realization of goals aligned with the principles of precision medicine in medical treatment.

One of the challenges of the use of AI is when AI is interpreted as a “black box” in which operators may only comprehend input and output predictions [49]. This drawback can be even more important for QA.

4

Materials and methods

4.1 Materials

This work is focused on the treatment of prostate cancer patients with Intensity-Modulated Radiation Therapy (IMRT). In the computational tests, the focus was on cases involving patients with the tumor localized within the prostate gland, post-surgery. CT images of five patients were used, including the structures of interest, that were previously delineated. The delineated structures include the rectum, bladder, left and right femoral head, planning target volume (PTV) and body. The resolution of the computed tomography (CT) images for each patient is presented in Table 4.1.

	Patient 0	**Patient 1**	**Patient 2**	**Patient 3**	**Patient 4**
X(mm)	3	0.98	0.98	0.97	1.27
Y(mm)	3	0.98	0.98	0.97	1.27
Z(mm)	3	3	3	3	3

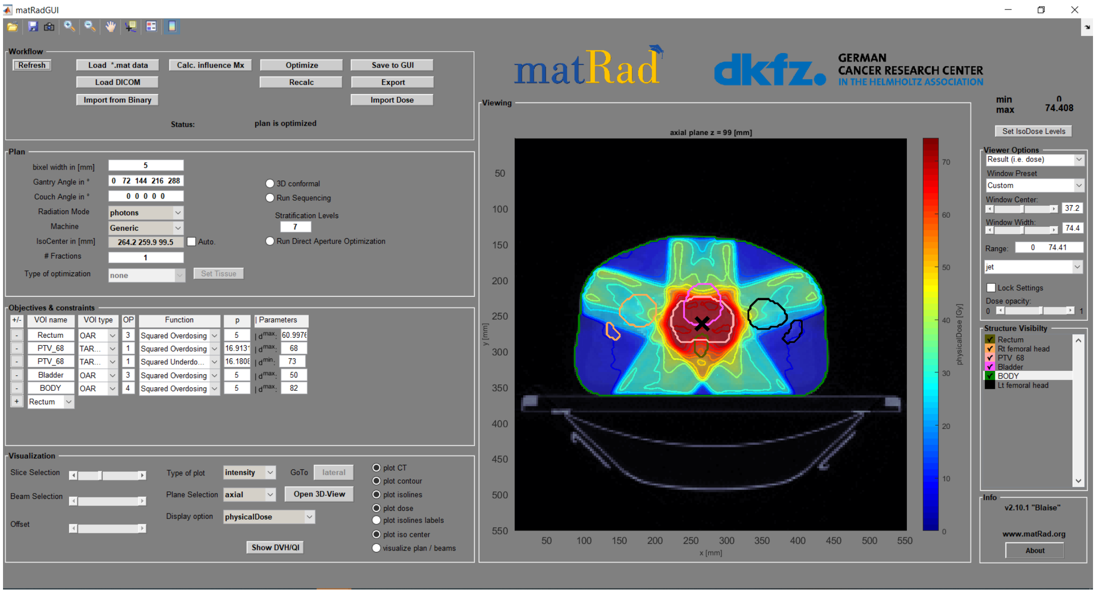
Table 4.1: CT resolution for each five patients

This work was developed using matRad, an open-source cross-platform toolkit developed entirely in Matlab, that provides most of the existing functionalities present in treatment planning systems for academic research, exhibited in Figure 4.1. It allows 3D IMRT treatment planning for photons, scanned protons, and scanned carbon ions [66]. All work developed, namely the new models and algorithms considered, were programmed in Matlab, version R2022b, and incorporated in matRad.

With the implementation of IMRT, a variety of dose metrics were considered. D_{95}, used for the PTV, represents the volume of structure that receives 95% of the prescribed dose [15]. As discussed in chapter 2, for the organs at risk (OAR) a maximum or mean dose can be used as metric, regarding the type of OAR that is being considered. For example:

- $D_{95} \geq$ 66,64 Gray (Gy) means the minimum dose should be greater than 66,64 Gy, in at least 95% of the structure's volume.

- $D_{mean} \leq$ 50 Gy means the medium dose should be lower than 50 Gy in all the structure's volume.



Status: plan is optimized

bixel width in [mm]: 5
Gantry Angle in °: 0 72 144 216 288
Couch Angle in °: 0 0 0 0 0
Radiation Mode: photons
Machine: Generic
IsoCenter in [mm]: 264.2 259.9 99.5 □ Auto.
Fractions: 1
Type of optimization: none
○ 3D conformal
○ Run Sequencing
Stratification Levels: 7
○ Run Direct Aperture Optimization

+/-	VOI name	VOI type	OP	Function	p	| Parameters
-	Rectum	OAR	3	Squared Overdosing	5	| d^{max}: 60.9976
-	PTV_68	TAR...	1	Squared Overdosing	16.9131	| d^{max}: 68
-	PTV_68	TAR...	1	Squared Underdo...	16.1808	| d^{min}: 73
-	Bladder	OAR	3	Squared Overdosing	5	| d^{max}: 50
-	BODY	OAR	4	Squared Overdosing	5	| d^{max}: 82
+	Rectum					

Slice Selection
Beam Selection
Offset
Type of plot: intensity
GoTo: lateral
Plane Selection: axial
Display option: physicalDose
● plot CT
● plot contour
● plot isolines
● plot dose
○ plot isolines labels
● plot iso center
○ visualize plan / beams

min 0
max 74.408
Result (i.e. dose)
Window Preset: Custom
Window Center: 37.2
Window Width: 74.4
Range: 0 74.41
jet
□ Lock Settings
Dose opacity: 0 1

✓ Rectum
✓ Rt femoral head
✓ PTV 68
✓ Bladder
✓ BODY
Lt femoral head

v2.10.1 "Blaise"
www.matRad.org

Figure 4.1: matRad interface

- $D_{max} \leq 70$ Gy means no voxel should receive more than 70 Gy in all the structure's volume.

Table 4.2 lists the plan restrictions for each structure that must be met in accordance with earlier studies for a plan to be considered admissible.

Bladder	Rectum	PTV	Rt femoral	Lt femoral	BODY
		$D_{95} \geq 66{,}64$			
$D_{mean} \leq 50$			$D_{mean} \leq 45$	$D_{mean} \leq 45$	
	$D_{max} \leq 70$				$D_{max} \leq 82$

Table 4.2: Medical prescriptions for each structure

The treatment plans were obtained for patients with prostate cancer considering a dose prescription of 68 Gy. To that end, five coplanar equidistant beam directions were used (0°, 72°, 144°, 216°, 288°).

4.2 Methods

In the current Radiotherapy (RT) workflow, the planning is done manually by the planner, which can be a lengthy and inaccurate process, as mentioned in Chapter 2. One of the main focuses of this work is to find an automated solution that facilitates this process and represents accurate and clinically applicable solutions, which means complying with the prescribed dose for the PTV and OAR, without a manual adjustment being necessary.

The iterative algorithm developed in [28] is used for this work. This algorithm solves the Fluence Map Optimization (FMO) problem in an iterative way using fuzzy logic, automatically setting the optimization model parameters to achieve the desired doses for each of the volumes of interest. The objective function that guides the optimization process aims to find the best treatment plan but its value does not have any clinical meaning. In this work, the quadratic penalty is considered for each voxel that receives a dose greater/lower than the dose limits defined by the medical prescription. The objective function used in this work is described in 2.3.2, as well as its full equation 2.1.

For each OAR, upper dose deviation is considered, meaning that only upper limits that should not be surpassed were taken into account and, for the PTV, under and upper deviations were considered. The goal is to find, in a totally automated way, weights and upper and lower bounds that lead to a solution that complies with the medical prescription. Let us consider, for example, the rectum. In FMO it can make sense to consider an upper limit in the objective function that is different from the upper limit defined by the medical prescription: the lower this upper limit is the higher the importance this structure has during FMO. If, in any iteration, the delivered dose in any of the voxels of the rectum is superior to the upper bound considered, the objective function will be penalized, indicating the algorithm to search for a solution that lowers the dose of all voxels regarding that structure's volume. Equally, if an inferior limit for the PTV is being considered, the objective function will be penalized if that minimum value is not being respected for every voxel, indicating the algorithm to search for a solution that rises the dose for every voxel of the structure's volume.

As previously discussed in 2.4.1, FIS is divided into two phases. However, in this work, only the first phase will be considered, meaning that we are only considering the goal of trying to achieve the desired PTV coverage, guaranteeing at the same time proper OAR sparing according to the medical prescription. This is firstly done by changing the upper and lower bounds only, and afterwards changing the structures' weights if the bounds are insufficient. Although the existing FIS approach is already capable of reaching an admissible treatment plan without any human intervention, it starts from scratch for every new patient, meaning that no learning is taking place, and the way in which the FMO model parameters are updated in each iteration is always the same. With the approach developed in this work, the aim is to integrate the use of the FIS approach with a learning capability so that the calculation of admissible treatment plans is done in an automatic and fast way (faster than the original FIS approach).

4.2.1 Reinforcement Learning: Q-learning

In this work, a new approach was developed based on the idea of having Q-learning define the fuzzy rules that are using during the running of the iterative FIS approach.

As far as the authors know, the approach presented in this work is novel and has not been addressed before. There is already some research regarding Q-learning applications in RT, but none of them involve the integration of optimization models and Artificial Intelligence (AI) as it is being proposed here. Research presented in [67] tried to combine the power of an agent-based approach with reinforcement learning for simulating and optimizing complex biological problems such as RT. In agent-based modeling, the interactions between agents and their environment are considered, providing a more detailed representation of the system under study. This approach offers a natural way to describe the dynamics of the system. The Q-learning algorithm was used as a model-free technique to optimize the RT treatment plan based on agent-based simulation.

In this context, the intensity of radiation was treated as an action variable, while the tumor size served as the state Q-table. Consequently, each element in the *Q(s, a)* table represented a combination of radiation intensity for a specific tumor size. Initially, the algorithm identified the current state of the tumor. During the early stages of the algorithm, actions were selected randomly to encourage exploration rather than focusing on maximizing rewards. As the Q-table accumulated rewards and converged, the exploitation part of the algorithm became more prominent, resulting in the selection of actions based on the maximum expected rewards.

In this work, optimization models and algorithms, fuzzy inference systems and Q-learning are used in a complementary way to develop an automated tool for RT treatment planning.

The FMO problem was solved using a quadratic optimization model and fuzzy rules, within a FIS, to tune the FMO model parameters. As the fuzzy rules themselves have parameters that can be tuned, and that define the mathematical representation of concepts, in this work we want to see if it is possible to learn what are the best rules within FIS that should be used in every step of the RT treatment optimization process. So, we have defined two different sets of rules: one set of rules considers the concepts of small, medium and large as being smaller than the other set of rules. One set of rules will lead to smaller changes in the FMO model parameters in each iteration, so convergence will be probably slower but more certain. The other set of rules will consider larger steps in each iteration, taking the risk of not converging because of these larger steps, but fastening the process if convergence is achieved. The optimal situation would be to take the most out of these rules in different steps of the FIS FMO process, and this is what we expect Q-learning to learn.

The idea is to use Q-learning so that it can choose the best action, in this case, the best set of fuzzy rules to use at each iteration of the FIS FMO algorithm so that the maximum reward possible is achieved, which translates into having the FIS FMO approach reaching an optimal solution in the least amount of time (minimizing the total number of iterations).

As previously mentioned, Q-learning begins to explore the environment, which means randomly choosing the actions and testing all possibilities, and then exploiting what it already knows, meaning choosing the course that obtains the best reward possible [53].

Following that reasoning, this study was developed using two sets: first, a training set is created to build a Q-table with each state best possible course of action. In this set, the actions are randomly chosen when training the algorithm. Then, according to the next state achieved by choosing a given action, a reward is given. The Q-learning algorithm is supposed to encounter as many different states as possible to get a complete and deep knowledge of the environment.

After training is complete, a final Q-table is obtained, ready to be used in the test set. In this set, the algorithm is supposed to choose the best action given the current state, according to the Q-table. This means choosing the action that represents the best Q-value for that current state. The objective is to have the FIS FMO approach using this Q-table to minimize the number of iterations until the treatment plan is calculated.

For illustration purposes, let us consider a Q-table with 3 states, such that each state is defined by what is happening in the current solution to a given structure:

1. The structure is complying with the medical prescription.

2. The structure is not respecting dose constraints by less than 10%.

3. The structure is not respecting dose constraints by more than 10%.

Considering, also, two actions, corresponding to:

1. Use fuzzy rules defined by set 1 to change the parameters of the FMO model.

2. Use fuzzy rules defined by set 2 to change the parameters of the FMO model.

Let us consider that the Q-table, after training, will look like this:

	Action 1	**Action 2**
State 1	4	3
State 2	2	1
State 3	1	3

Table 4.3: Example of a Q-table

This information can be interpreted as follows: whenever the fuzzy inference system (FIS) ends an iteration, a correspondent state is achieved that is defined by what is happening with the structure of interest. Based on this current state, the algorithm then chooses which action to take based on the information already gathered in the matrix. For example, if the correspondent state is state 2, the algorithm will choose action 1, given that it displays the biggest Q-value for that state. It will decide what is the set of fuzzy rules that it should use to change the parameters of the FMO model so that the treatment plan can be improved in the next iteration. The main purpose is to use this reasoning for IMRT treatment planning, considering time reduction as a main goal. To that end, two different strategies were tested.

4.2.1.1 First strategy - single Q-table

The Q-learning algorithm is incorporated into the fuzzy inference system (FIS). FIS, given the chosen action at each iteration, will run the algorithm until an admissible solution is found taking into account the prescribed doses for each structure.

The first strategy involves considering a Q-table composed of 27 lines (corresponding to 27 different states) and 2 columns (corresponding to the 2 courses of action). The 27 states correspond to the 3 possible states for each structure, which gives us a total of $3^3 = 27$ states. The 3 possible states for each structure and their corresponding rewards are defined as follows:

- the structure complies with the prescribed dose (S_2- reward +2)

- the structure does not comply with the prescribed dose by 1 Gy (S_1- reward +1)

- the structure does not comply with the prescribed dose for more than 1 Gy (S_0- reward +0)

As for the actions, two are being considered: one with smaller and the other with larger deviations associated with the concepts considered, as described before.

At each iteration, a state is activated in the Q-table depending on each structure's current state, according to the following expression:

$$state(IT) = admin(1, IT) * (3^2) + admin(2, IT) * (3^1) + admin(3, IT) * (3^0) + 1;$$

For example, let us consider that the current action (choosing one out of the two possible fuzzy rules) leads to a state where the bladder and the PTV satisfy the prescribed dose, but the rectum does not (considering, for example, that the dose is not being delivered for less than 1 Gy). In this expression, IT corresponds to the current iteration, and $admin(x, IT)$ refers to the reward corresponding to the state of a specific structure in that iteration. In this code, $admin(1, IT)$ corresponds to the bladder, $admin(2, IT)$ corresponds to the rectum, and $admin(3, IT)$ corresponds to the PTV. So, in this case, the bladder and PTV will activate state 2 and the rectum will activate state 1, and their value in $admin(x, IT)$ will be the corresponding rewards.

The training set is used to fill the matrix according to the reward, which will be a direct sum of each structure's reward at the current iteration. This table is supposed to guide the optimization process in the test set, in such a way that the algorithm converges quickly and creates an admissible treatment plan. The goal is to develop a single Q-table that can be applied to any patient that presents similar conditions to the ones used for this study.

4.2.1.2 Second strategy - a Q-table for each structure

In the second part of the work, three tables were considered, one for each structure of interest. So, instead of considering the overall state of the system including all the structures in the definition of this global state, each structure is considered independently from one another regarding the state that it is in, although the calculation of the rewards considers the interaction that can exist between the different strategies. Each volume of interest can be in one of the following states:

- The delivered dose complies with the prescribed dose (state 0- S_0)
- The delivered dose does not comply with the prescribed dose by 1 Gy or less (state 1- S_1).
- The delivered dose does not comply with the prescribed dose by more than 1 Gy but less than 5 Gy (state 2- S_2).
- The delivered dose does not comply with the prescribed dose by more than 5 Gy (state 3- S_3).

In this second approach, each structure can be in a different state, and different fuzzy rules can be learned for each structure. This means that three Q-Tables will exist, and three learning processes will be simultaneously in place. This will allow different fuzzy rules to be applied to different structures.

	S_0	S_1	S_2	S_3
S_0	1	0	0	0
S_1	2	1	0	0
S_2	3	2	1	0
S_3	4	3	2	1

Table 4.4: Matrix regarding the current and previous state

As for rewards, they are calculated considering a table that relates the previous state with the current one, considering the chosen action. This means that, differently to what occurs in the first strategy, where the reward was automatically established given the current state, the reward takes into account the immediately preceding state. The matrix that guides the reward calculation is described in Table 4.4. This matrix will be used to calculate the rewards for each structure. As each structure will be in its own state but, in reality, they are all interconnected, the calculation of the reward for each structure will take into account all structures, regarding their values on Table 4.4. An action chosen for one structure will also influence what happens in the other structures. This is what motivates this choice of reward calculation. The reward is calculated as follows:

$$\begin{aligned}\text{reward}_1(IT) = {} & \text{matrix}(state_1(IT-1), state_1(IT)) + \text{matrix}(state_2(IT-1), state_2(IT)) \\ & + \text{matrix}(state_3(IT-1), state_3(IT))\end{aligned} \tag{4.1}$$

There were two types of tests made regarding this strategy, both using all five patients.

Firstly, we tested all five patients with five different sets of Q-tables, obtained from each patient training set. Therefore we obtained 5 trained Q-tables, which were then tested in all 5 patients, giving us a total of 25 tests. This means that each case was tested considering its “optimal” Q-table (the one built by learning with that same case), and also the other Q-tables created by learning with the other available cases. The objective of these tests was to assess whether there were differences in the Q-tables created by using different learning contexts and whether it would be possible to find a pattern for why a specific Q-table resulted in better outcomes, no matter what patient was tested on. Furthermore, it is also important to understand if the reduction achieved in the number of iterations was only observed when testing a patient with his own trained Q-table or if the reduction was always observed, regarding other Q-tables created with other cases. Depending on the results, it is possible to conclude whether this strategy is too patient-specific, or if it really is a viable and generalizable approach, with promising results.

After this strategy was tested, and considering the results obtained, a second approach was created with the use of cross-validation within this second strategy. The tests were performed under a leave-one-out cross-validation method to create Q-tables by excluding one patient at a time from the group, only to be later employed for evaluation on the very same patient. Four patients were used to create the Q-tables, using their individually trained ones. The average, maximum and minimum values were calculated, resulting in three groups of Q-tables, each group containing a Q-table for the rectum, bladder and PTV . Subsequently, they were tested on the patient who was left out of the initial group of patients. This method was applied to all patients, meaning that the tables were calculated five times, every time excluding a different patient that was then used for the testing set. These tests were carried out with the goal of achieving consistent results that revealed improvement in all five patients.

Another important difference between strategies 1 and 2 is the way in which FIS FMO is used for learning: whilst in the first strategy FIS FMO considers the simultaneous change of the parameters of all the structures that are not complying with the medical prescription, in this case, it was chosen to change the parameters of one single structure at a time. This structure is randomly chosen. Furthermore, learning capability is also leveraged by randomly initializing the optimization model's parameters whenever the algorithm converges: when a treatment plan complying with the medical prescription is achieved, the process is restarted considering a different set of initial parameters, until a given maximum number of iterations is reached.

All the results obtained regarding both strategies are presented next in Chapter 5.

5

Results

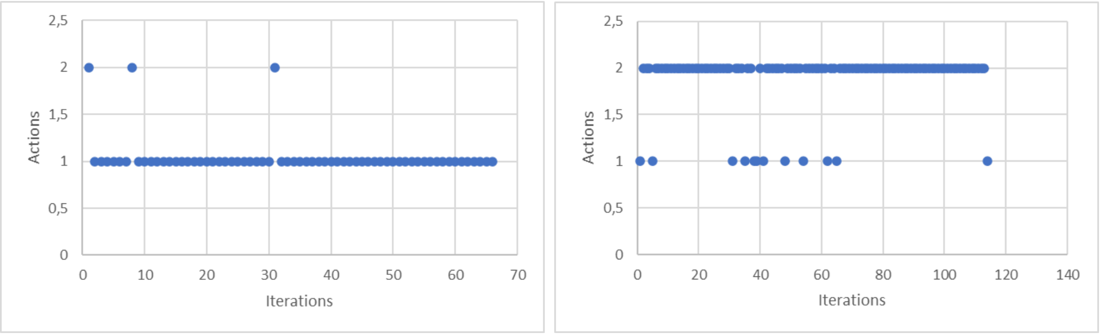
The main goal of this study is to use the Q-learning algorithm, along with an optimization model, as a new tool for Radiotherapy (RT) treatment planning.

5.1 Q-learning table with all three structures

In the first part of this work, a Q-learning matrix involving all three structures (bladder, rectum and planning target volume (PTV)) was created, trained and tested.

Prior to conducting this study, it was crucial to ensure that there was no single action consistently deemed the preferable choice, as such a scenario would undercut the core purpose of employing Q-learning. To that end, we verified the actions chosen by Q-learning in two test sets, to further understand if one action was always better regarding the state, which would withdraw Q-learning of its purpose. The graphics 5.1a and 5.1b below correspond to two test sets used, where the Q-learning algorithm chooses the best action for the current state.

To further validate this, three different scenarios were considered using this first strategy. One scenario consistently employs action 1, another scenario employs action 2, while the remaining scenario randomly selects between the two actions. This last arrangement effectively fulfills the exploration phase of Q-learning. It is crucial to describe the differences between the two actions. First, the actions vary concerning the outputs. On one hand, action 1 is more demanding regarding the outputs, meaning the algorithm will give smaller steps until converging. Action 2, on the other hand, will have more loose bounds for the outputs, meaning the algorithm can give bigger steps in each iteration. Either way, in all three tests the algorithm always con-



(a)

(b)

Figure 5.1: Two different cases, (a) and (b), with respective action choices

verged, no matter how loose or demanding the bounds were, which reflects the robustness of the fuzzy inference system used. The purpose of conducting three distinct tests was to understand whether or not randomly choosing between actions helped the algorithm converge faster, which was verified, as it can be observed in Tables 5.1 and 5.2. Thus, we concluded that Q-learning qualified as a relevant tool for optimization.

Training Iterations with action 1	Training Iterations with action 2
92	131
135	152
131	131
201	201
131	141

Table 5.1: Training sets for each action with various patients

Training Iterations with both actions	Testing Iterations
88	92
135	125
116	131
201	201
129	66

Table 5.2: Comparison between training and testing sets with various patients

In the training set, a Q-table with a random choice of actions is filled. Table 5.3 represents one final Q-learning matrix after being trained, in this case for patient 0. One can notice that there are two states that are more activated than the rest. These correspond to the state where both the bladder and PTV are complying with the prescribed dose but the rectum is not (state 21), and to the state where both the bladder and PTV are complying with the prescribed dose but the rectum is not by the difference of 1 Gray (Gy) (state 24). This tendency was observed in every trained matrix, again regardless of the scenario considered.

In the testing set, the trained Q-table is used, which means that there is no random choice of actions and no updates to the matrix. For the test set, at each iteration, the algorithm will choose the action that has the bigger Q-value, calculated using the training set. A bigger Q-value means a bigger reward and ultimately, a better result. Theoretically, Q-values were calculated having the best possible outcome in mind, so choosing the actions according to the matrix would mean a faster convergence, with the best possible dose for each structure.

It is worth noting that the use of the Q-table never led to worse outcomes than using only one of the defined fuzzy rules set. However, looking at the number of iterations, it is observed that the testing results are worse than the training results in 3 out of the 5 cases, as can be observed in Table 5.2. Despite some cases presenting fewer iterations in the test set, those are not sufficient to be considered relevant or consistent. So, by observing these results, we can

	Action 1	**Action 2**
State 1	0	0
State 2	0	0
State 3	0	0
State 4	0	0
State 5	0	0
State 6	0	0
State 7	0	0
State 8	0	0
State 9	0	0
State 10	0	0
State 11	0	0
State 12	0	0
State 13	0	0
State 14	0	0
State 15	0	0
State 16	0	0
State 17	0	0
State 18	0	0
State 19	3.1562	0
State 20	0	2.0000
State 21	7.4339	8.0500
State 22	0	0
State 23	0	3.9520
State 24	8.3647	8.7344
State 25	0	0
State 26	0	0
State 27	0	4.5912

Table 5.3: Q-learning matrix with two actions used for patient 0

conclude that it can be advatageous to use more than one set of fuzzy rules to update the FMO model parameters, but using the trained Q-table to do this does not present a clear advantage.

The dose-volume histogram (DVH) represents the percentage of volume that is being irradiated and the corresponding dose, for each structure. Although our Q-table only concerns three structures, the left and right femoral head and the overall body are also considered for dose calculations. For illustration purposes, the DVH with all structures regarding patient 0 is displayed in Figure 5.2. In the ideal scenario, PTV would be 100% irradiated with the prescribed dose and the other structures would have volume doses percentages close to 0%.

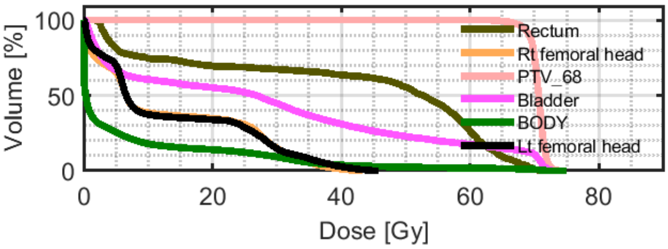


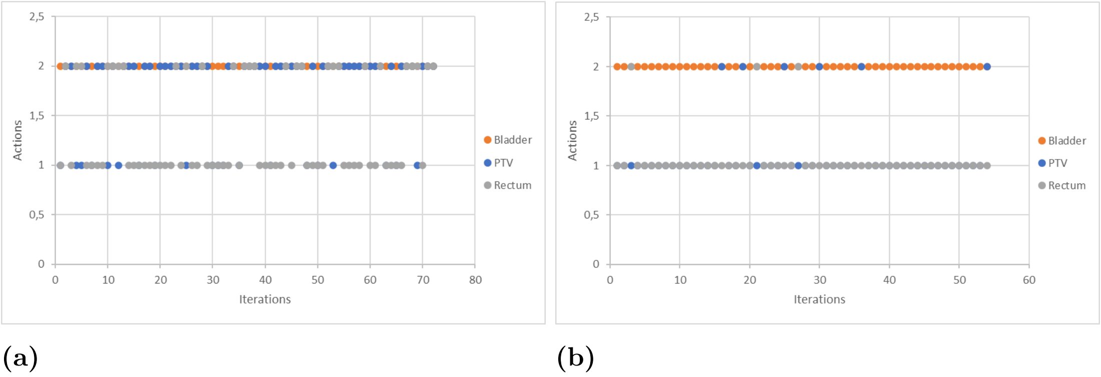
Figure 5.2: DVH for patient 0 regarding the first strategy

5.2 Q-learning table for each structure

In the second part of the work, three matrices were considered, one for each structure of interest. This approach intends to find better and more specific optimal solutions for each structure. However, it is important to point out that no action is chosen without taking into account the other structures. As each structure is being evaluated individually, it is possible to create more specific and varied conditions, hence the four new conditions created and displayed in Chapter 4.

Again, it was crucial to ensure that there was no single action consistently deemed the preferable choice. The results observed are presented in figures 5.3a and 5.3b, corresponding to two test sets used, where the Q-learning algorithm chooses the best action for the current state.

As already discussed, this strategy looks at each structure through an individual matrix,



(a)

(b)

Figure 5.3: Two different cases, (a) and (b), with respective action choices

meaning that it is structure-specific. By creating individual Q-tables for each structure and incorporating rewards that take into account the previous step, the strategy exhibits promising improvements aimed at making the algorithm more robust, boosting the overall performance.

Observing the Q-tables for each structure (an example is presented in Figure 5.4), one can observe multiple important results. Firstly, the Q-table for the bladder only has one state activated (one line), which is the one where the structure is complying with the prescribed dose. This was verified in 3 out of the 5 patients tested, which means that the bladder, most of the time, does not represent a concerning structure in terms of planning, since it is possible to comply with its dose limits, minimizing its risk of being compromised. Regarding the PTV and the rectum, the results are very different. Both cases present values for every state meaning that during training the doses acquired were sometimes compromising other organs or were simply not complying with the prescribed ones, and therefore needed adjustment. Nonetheless, a feasible solution was always found, due to the robustness involving fuzzy logic. The two most common states, for both structures, were the third and fourth state (line), which corresponds to the structures not complying with the prescribed doses for more than 1 Gy but less than 5 Gy and for more than 5 Gy, respectively. This possibly means that reaching the prescribed doses was not an easy task to perform while finding a compromise between the two structures (again, taking into consideration that the bladder is not much cause for concern).

1.4413	2.8110
0	0
0	0
0	0

(a) Q-table: bladder

0.3504	1.7617
1.8658	1.8126
2.1774	5.3633
4.5797	2.0300

(b) Q-table: PTV

2.2747	1.4574
0.8137	1.9794
4.7943	2.2328
2.8691	4.8816

(c) Q-table: rectum

Figure 5.4: Example of Q-tables for the second strategy

The iterations concerning each patient's training set are presented in Table 5.4.

Patient	Iterations
Patient 0	425
Patient 1	125
Patient 2	501
Patient 3	501
Patient 4	501

Table 5.4: Iterations of each training set

After training each patient, the correspondent Q-table was tested. After this first test, the trained Q-table of one patient was used on other patients, as mentioned before. This second strategy revealed itself with more diverse results, but very relevant nonetheless. These results, regarding the Q-table obtained for each patient, applied to the same and other patients, as well as the corresponding iterations in the test set, are presented in Table 5.5.

Patient tested Trained Q-table	Patient 0	Patient 1	Patient 2	Patient 3	Patient 4
Patient 0	72	42	114	200	200
Patient 1	54	54	200	200	200
Patient 2	63	78	200	200	200
Patient 3	45	38	72	200	200
Patient 4	63	78	133	200	149

Table 5.5: Iterations of each test set regarding each trained Q-table

The DVH presented in Figure 5.5 represents, once again, the delivered doses for each structure regarding patient 0, regarding the percentage of volume irradiated and the corresponding dose.

When comparing the DVHs of both strategies, the disparities are most evident in the bladder and rectum structures. The second strategy is undoubtedly better for both structures since they have a lower percentage of volume being irradiated. Regarding the remaining structures, the outcomes are similar, indicating that, overall, the second strategy outperforms the first when comparing the percentages of volume being irradiated.

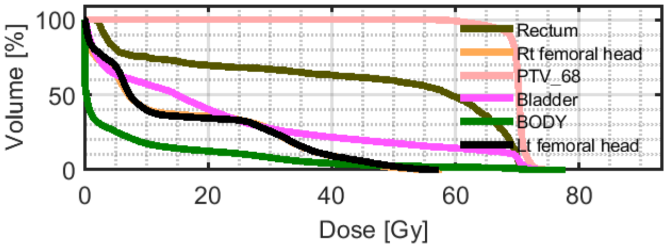
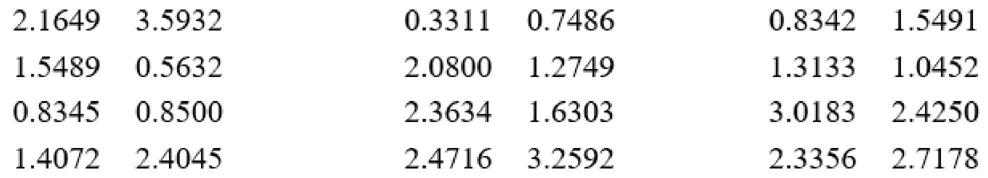


Figure 5.5: DVH for patient 0 regarding the second strategy

Some Q-tables definitely reduce computational time for all patients and converge to a feasible solution. This means that, ultimately, it is possible to achieve a global Q-table that can be used in all patients in the same conditions, and that can be applied for all types of conditions, taking into account the structures that are relevant for each case. For example, when training with patient 3, the Q-table presented a huge decrease in computational time when applied to all other patients.

However, the reduction expected from the training to the test set was inconsistent and not always verified, leading to the last test, described in Chapter 4, using a leave-one-out cross-validation technique. As a result, five sets of three Q-tables are generated, each set omitting one patient from consideration. Three sets of Q-tables are presented in Figures 5.6, 5.7 and 5.8, in this case excluding patient 0.

As we can observe, the minimum values' Q-table stands out due to its large number of empty entries. This is particularly noticeable in the bladder's matrix, where a significant portion of the matrix is zeros. This absence means that these particular states were inactive throughout the original Q-tables. This outcome aligns with our earlier observations, given that the bladder



2.1649	3.5932
1.5489	0.5632
0.8345	0.8500
1.4072	2.4045

(a) Q-table: bladder

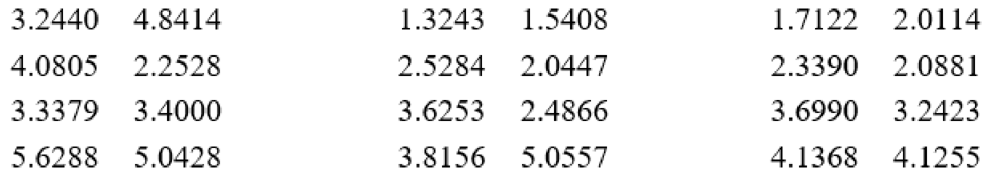
0.3311	0.7486
2.0800	1.2749
2.3634	1.6303
2.4716	3.2592

(b) Q-table: PTV

0.8342	1.5491
1.3133	1.0452
3.0183	2.4250
2.3356	2.7178

(c) Q-table: rectum

Figure 5.6: Q-tables with average values



3.2440	4.8414
4.0805	2.2528
3.3379	3.4000
5.6288	5.0428

(a) Q-table: bladder

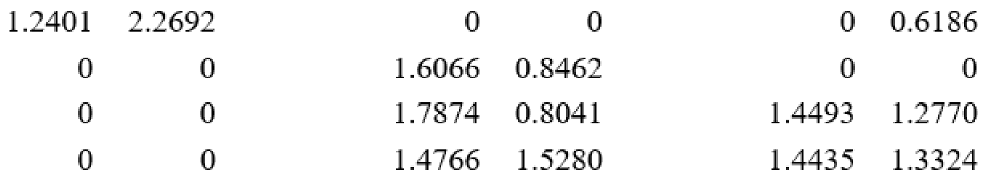
1.3243	1.5408
2.5284	2.0447
3.6253	2.4866
3.8156	5.0557

(b) Q-table: PTV

1.7122	2.0114
2.3390	2.0881
3.6990	3.2423
4.1368	4.1255

(c) Q-table: rectum

Figure 5.7: Q-tables with maximum values



1.2401	2.2692
0	0
0	0
0	0

(a) Q-table: bladder

0	0
1.6066	0.8462
1.7874	0.8041
1.4766	1.5280

(b) Q-table: PTV

0	0.6186
0	0
1.4493	1.2770
1.4435	1.3324

(c) Q-table: rectum

Figure 5.8: Q-tables with minimum values

predominantly complies with the recommended doses. As a result, values relating to this specific state are largely present in the first line.

When we look at the average and maximum value matrices, we see that all states are filled with their corresponding values. Notably, distinct states take prominence within each configuration. In the PTV, the state associated with being the farthest distant from a feasible prescription has the highest level of activation. The rectum shows a well-balanced distribution of values across many states. Nonetheless, the state furthest from a solution has the largest level of activation, as seen in the PTV. In the bladder, once again, the state associated with dose compliance takes precedence.

These Q-tables are referent to one patient (patient 0), meaning that they were created excluding that patient's Q-table from the equations. They were then applied to that same patient, in the test set. This leave-one-out method was applied to all patients and the results obtained regarding the test sets are presented in Table 5.6.

	Average	Maximum	Minimum
Patient 0	56	56	60
Patient 1	74	51	61
Patient 2	53	52	61
Patient 3	500	500	500
Patient 4	52	52	62

Table 5.6: Iterations of each test set

These tests aimed to find consistent results, that showed improvement in all five patients. As we can observe, the results are consistent throughout the three test sets (regarding maximum, minimum and average values) when compared to the training sets presented in table 5.4. Still, the Q-tables that concern maximum values consistently presented the best results. Regarding the DVHs, one can easily observe that there is no evident difference comparing this approach to the first one used in this second strategy, leading us to the conclusion that although this approach presents evident time reduction, its feasibility and effectiveness do not present relevant differences.

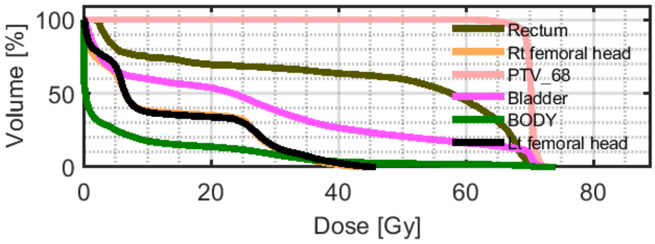


Figure 5.9: DVH from maximum Q-table values

6

Discussion

Q-learning emerged as an attractive tool for Radiotherapy (RT) treatment planning, working as a supplement to the optimization process, especially in fuzzy logic employed in previous works. This study conducted multiple tests in order to discuss its importance in the optimization process, with a primary focus on reducing computational time. This chapter will now discuss the results obtained and presented in Chapter 5.

6.1 First strategy

This first strategy was created keeping in mind that only three structures were being considered, therefore it would theoretically be possible to gather all information regarding the three in one Q-table. The rewards were always calculated considering the current state of each and every structure (if they were complying with the doses or not, and if they were far from their dose objectives).

Q-values were calculated having the best possible outcome in mind, so that choosing the actions according to the Q-table in the test set would mean a faster convergence. However, that was not observed. When looking at Table 5.2, it is easily noticed that there was no time reduction, which in this case is translated into the number of iterations, from the training set to the test set. Since the main goal is to achieve a Q-table capable of being applied in multiple cases that present the similar conditions, one can quickly conclude that this cannot be done given these results.

This led to the inevitable conclusion that Q-learning was not improving the algorithm on its main goal: finding the best possible actions to converge faster and reduce computational time. In this context, it could even be argued that FIS reliance on the chosen action is minimal, eliminating the necessity for multiple action options and consequently eliminating the need for Q-learning in the optimization process since in most cases FIS converges regardless of the action or actions that are being considered. For instance, when looking at graphics 5.1a and 5.1b, that derive from distinct patients, we can easily observe that one action dominates over the other in both scenarios.

Nevertheless, neither action 1 nor 2 is equally chosen in all patients, meaning that there is no single action that proves to be globally preferable in all cases when comparing the same structures. Depending on the patient being examined, looser or more demanding bounds can be favored. This means that despite FIS being very flexible regarding the action chosen and normally being able to converge, having different actions being chosen in different iterations of the FIS FMO algorithm is reasonable and can be advantageous.

There are, however, some cases that do not find convergence even with FIS. When training the algorithm with patient 3, convergence was not achieved and the treatment plan was obtained by reaching the maximum number of iterations (which was 200 in this case), which did not always allow for a feasible solution. This could be due to computational capacity, the complexity of the case or the algorithm needing more time to achieve a feasible plan.

In light of these results, there is still a significant data point worth mentioning. Whatever action is chosen, regarding the current state, the bladder always complies with the prescribed dose. This was verified in all patients tested. This is a very important result because it gives the planner more freedom when adjusting parameters. Knowing that the treatment plan will most likely not affect the bladder leads to the conclusion that the true compromise will always be between the rectum and the planning target volume (PTV). Additionally, it was observed that the rectum was the most difficult structure to achieve the prescribed doses, regardless of the action chosen. Most of the time, convergence happens without the rectum having any alterations in its state.

This represented another problem for this approach, given that the goal is to have all structures comply with the prescribed doses. This outcome highlights the widely held view that finding a balance between structures is difficult, if not impossible, even when only two structures are considered. It is important to keep in mind that all structures are dependent on one another, so it is possible to jeopardize the rectum whilst bettering the results for the PTV .

Observing these results altogether, it was concluded that the approach taken does not improve treatment planning and therefore does not achieve substantial improvements. The only important result was the observation that using Q-learning would not worsen the results conpared with the situation of using only one set of rules. Nonetheless, it represented a first step towards incorporating Q-learning into the treatment planning optimization process, which led us to the second part of our work.

Although we are considering only three structures where one of them always complies with the prescribed doses, therefore not interfering with a compromise between structures, gathering all information in one Q-table may not be the best approach. In this case, the rectum and the PTV have very different objectives, that are not always well reflected when considering a global reward, since it is not possible to observe if a certain action that represents an improvement for the PTV and for the overall reward, is not actually worsening the rectum in its goals. The reward is calculated according to all three structures, so it is possible to have a positive reward for the PTV and no reward for the rectum and the overall reward being positive, meaning that the Q-value will be positive, misleading the algorithm to choose an action that actually does not benefit all structures.

This was the main idea that motivated us to create an alternative strategy concerning fuzzy logic and Q-learning, focusing on addressing each structure's individual demands.

6.2 Second strategy

This second strategy emerged as a solution to the problem encountered in the first one: lack of specific objectives for each structure. The results were far more satisfying. Since this strategy is more structure-specific and thus more demanding, it requires a longer computational training time. Nonetheless, the improvements observed were far more relevant regarding treatment planning.

Before discussing the results of the test sets, it is relevant to analyze the first graphics, which concern the choice of actions for each structure. The importance of having both actions is easily understandable, even more so than in the first strategy. Since now we are considering the three structures individually, their actions will also be chosen individually, with careful deliberation regarding that structure's goals. Looking at the graphics, we observe that an action that is repetitively chosen by one structure is hardly chosen by another. This could only draw us to the conclusion that neither of them is essentially better than the other, confirming our reckoning, already disclosed in the first strategy.

When looking at the obtained results, in almost all cases there was an evident decrease in the number of iterations (meaning less computational time) from the training set to the test set, when using the same patient. This is already an improvement from what was observed in the first strategy, where no consistent results were found when it came to the reduction of iterations.

As it was described before, every Q-table created in the training set was tested in all patients, in order to conclude if there were a Q-table that could be used globally. That goal was not achieved since there was not a Q-table with consistent reduction regarding all patients. Looking at patients 0 or 1 in Table 5.5, the results were promising since reduction was always visible and substantial, contrary to what had happened in the first strategy. However, we can see that patient 3 never converged and always achieved the maximum number of iterations permitted, which can be explained by it being a more complex case. As for patient 4, we can observe it also achieved the maximum number of iterations permitted, except when it was tested with its trained Q-table. This can lead to the conclusion that, for some cases, Q-tables need to be more case-specific, meaning that it is difficult to achieve a global Q-table that can be applied to a large number of cases. Nonetheless, these results demonstrate that it is possible to do so since the Q-table trained with data from patient 4 consistently reduced the number of iterations when applied to all other cases, which means that even when the Q-table is created based on a complex case, it still presents very favorable results when applied to other simpler cases. This patient's Q-table was the only one presenting positive results for all patients except patient 3.

Some new conclusions can be drawn contrary to what was observed in the first strategy, specially when looking at cases where the training and test set reach the maximum number of iterations. For instance, regarding the bladder, the Q-table presents values for more than one state, meaning that it does not always comply with the prescribed dose. This event did not occur consistently during each instance of Q-table training. When it did occur, it was always connected with situations in which the algorithm failed to find feasible solutions. In most cases, the bladder presented no difficulty with complying with the prescribed dose regardless of the patient being used, so we consider that overall it can still be considered the less concerning structure.

As for the rectum, the same as the first strategy is observed. It does not always comply with the prescribed dose, for less than 1 Gy.

As for the test sets, a clear prevalence of action 2 (it was chosen in every iteration) for the bladder and a mix of both actions for the PTV and rectum were observed in most patients. While it is possible to argue that the bladder does not need Q-learning, the same assertion cannot be made for the other structures. This is due to the absence of consistent outcomes, which stops us from drawing wide-ranging conclusions.

This second strategy revealed itself more promising since the number of iterations decreased in multiple patients given different Q-tables. However, we cannot overlook that a lot of test sets concerning patients 2, 3 and 4 did not reach a feasible solution. In light of these results, one cannot conclude that this strategy fulfilled our goal entirely. It is not possible to find a consistent reckoning that explains in what conditions the test set will or will not converge to a feasible solution since the results do not follow a clear train of thought. However, we cannot neglect the positive results obtained either, given that they do represent an improvement from the first strategy and provide us with a promising foundation for further investigation, indicating a clear path towards potential improvements.

These results led us to the last approach tested in our study, which consisted of building our Q-tables based on more than one patient, considering different parameters, in order to achieve a more robust foundation that could potentially be applied to a wider range of patients exhibiting similar clinical conditions, reaching feasible results in all of them.

The results obtained in this last approach were by far the most satisfactory. In 4 out of the 5 patients, a reduction in computational time was verified, in all three test sets. Furthermore, it was possible to obtain fewer iterations when compared to the other approach studied within this second strategy. Looking at patients 2 and 4, the results are even more positive since the reduction is drastically bigger. For all Q-tables that derived from training sets used to test both of these patients in the first approach, the reduction was either nonexistent or less meaningful, so using cross-validation presents as the most promising method.

Looking at Table 5.6 with a closer look, one can observe that the best results were always obtained when using the Q-table that was constructed from the maximum values obtained. This can presumably be due to the nature of a maximum value. In our case, a maximum value simply means that that action presented itself as the best option a greater number of times. Hence, it can make sense that this is the Q-table that presented the best result since it is a compilation of all the best courses of action. However, even when looking at the average and minimum Q-tables' results, we can still observe that they all present better results than when using a Q-table built only from the training set of one patient.

Calculating Q-tables based on multiple patients created more accurate and complete ones and led us to the conclusion that it is of good reasoning to build them based on a more robust batch that displays the same conditions. The use of the leave-one-out cross-validation method for the test sets proved to be a highly effective approach in ensuring both robust and trustworthy results. This method capacity to systematically exclude one patient during testing not only contributed to the reliability of the outcomes but also increased the credibility of the results.

By employing this approach, the research was able to yield results that exhibit a high level of consistency, which can be regarded as a solid foundation for drawing meaningful conclusions.

Overall, the results obtained verify that Q-learning can be used as a tool to optimize treatment planning in RT. In our study, only five patients were used, and while using the cross-validation method, only four patients were used to create our Q-tables. This leads us to the assumption that if a bigger database was used, better results would probably be obtained. Nevertheless, these outcomes represent significant accomplishments, as reduction was observed even when using a small dataset.

In conclusion, from the results gathered, we can already state that our second strategy is more promising since it presented more accurate and feasible results, which were overall positive. Considering specific objectives for each structure will always mean better accuracy, which should always be one of the main concerns. Nevertheless, combining that with computational time reduction is not always easy or even possible. Overall, it is very important to keep in mind that there are a lot of variables that can influence the quality of the treatment plan. While the results show progress, they also highlight the need for more refinement and improvement in our ongoing pursuit of quality performance, so the possibility for advancement remains.

7

Conclusion

This study focuses on Q-learning and its potential benefits in Radiotherapy (RT). RT remains one of the most used techniques for cancer treatment, therefore the search for innovation and refinement in RT techniques and technologies is still a constant pursuit. Researchers and medical professionals continuously seek novel strategies to enhance the precision, effectiveness and safety of radiotherapy.

This study tries to create a more automated approach that would improve plan accuracy while reducing susceptibility to human errors, all in a significantly reduced time frame. To that end, two main strategies whose goal was to achieve feasible treatment plans were conducted using Q-learning as a new tool.

Q-learning holds a lot of potential for RT. To our knowledge, it had never been used as an instrument in RT. Since it has the ability to learn and develop with the environment, we saw a great window of opportunity in using it for an already developed fuzzy inference system that builds treatment plans.

In summary, our study revolved around two main objectives: validating the potential advantages of employing Q-learning in radiotherapy, in a manner that resulted in feasible plans, and assessing whether this approach achieved such results while concurrently decreasing computational time.

After testing the two strategies and considering the two approaches directed in the second strategy, we can conclude that Q-learning yields potential benefits for treatment planning and can be used in order to reduce computational time. Although our results were not 100% positive in any approach, 80% of our results in the last approach, which utilized cross-validation, were positive. Given that our database consisted of only 5 patients and this approach tested 3 different sets on each patient, and only one patient revealed no improvement, we consider this to be an optimistic and fruitful result, as discussed in Chapter 6.

Our second strategy was motivated by the results obtained from the first strategy. Therefore, we consider that further investigation could lead to better and more robust approaches, having our approach as a foundation.

This study aims to improve RT in the near future, creating plans that exhibit high accuracy in delivering radiation to the tumor whilst sparing the organs at risk, all in a quicker and more efficient way. Through these efforts, we aspire to positively impact RT treatments, thereby contributing to the improved well-being of cancer patients. Given that cancer remains one of the most pressing global causes of mortality, this pursuit remains crucial and of utmost importance.

It is important to identify the obstacles to this work as well as the approaches used. Our study only takes into consideration three structures, which are the most relevant ones. Nonetheless, that can lead to less accurate results, given that all structures are interrelated. Furthermore, only 5 patients were used, which leads to little data variance. Altogether, these factors can influence the results obtained, specially regarding the cases where feasible plans were not reached.

As for future work, as already discussed in Chapter 6, there are a lot of factors that can be looked into. Our study was conducted on prostate cancer patients and our results suggest that this approach should also be tested in other types of cancer. Applying this work to all types of cancer would be a huge step in the RT field, potentially yielding significant advancements.

It is vital to mention that creating structure-specific goals is not the only possible approach for improving our work and reaching better results regarding Q-learning. Q-learning was used solely to choose between actions. Therefore, exploring those actions could be an interesting angle for future work. Searching for actions that lead to faster convergence, better and more demanding results, and using more than just two actions are all valid points. These can be explored since our actions and corresponding bounds were chosen based on a trial-and-error procedure alone. It could be interesting to test this approach with other bounds, obtained using a different procedure that is less prone to errors.

Another possible future study is related to the structures that are being considered. In our case, we rule out some structures that are normally also taken into consideration, such as the left and right femoral heads and the overall body. Using more structures means creating more Q-tables, which will automatically contribute to an increase in computational time. Although considering them in the algorithm can make it slower, it is possible to create more accurate results. However, compromise is difficult to obtain, and increasing the number of structures can make it even more so, at least considering our FIS approach. Furthermore, more structures and consequent Q-tables mean more margin for errors and wrong assumptions.

Bibliography

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] H. Rocha and J. M. Dias, "On the optimization of radiation therapy planning," *Inescc Research Repor*, no. 15, 2009.

[3] R. Baskar, K. A. Lee, R. Yeo, and K.-W. Yeoh, "Cancer and radiation therapy: current advances and future directions," *International journal of medical sciences*, vol. 9, no. 3, p. 193, 2012.

[4] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep learning in medical image registration: a review," *Physics in Medicine & Biology*, vol. 65, no. 20, p. 20TR01, 2020.

[5] D. M. Shepard, M. C. Ferris, G. H. Olivera, and T. R. Mackie, "Optimizing the delivery of radiation therapy to cancer patients," *Siam Review*, vol. 41, no. 4, pp. 721–744, 1999.

[6] W. Owadally and J. Staffurth, "Principles of cancer treatment by radiotherapy," *Surgery (Oxford)*, vol. 33, no. 3, pp. 127–130, 2015.

[7] A. Ahnesjö, B. Hårdemark, U. Isacsson, and A. Montelius, "The imrt information process—mastering the degrees of freedom in external beam therapy," *Physics in Medicine & Biology*, vol. 51, no. 13, p. R381, 2006.

[8] Y.-P. Liu, C.-C. Zheng, Y.-N. Huang, M.-L. He, W. W. Xu, and B. Li, "Molecular mechanisms of chemo-and radiotherapy resistance and the potential implications for cancer treatment," *MedComm*, vol. 2, no. 3, pp. 315–340, 2021.

[9] P. Carrasqueira, M. Alves, J. Dias, H. Rocha, T. Ventura, B. Ferreira, and M. Lopes, "An automated bi-level optimization approach for imrt," *International Transactions in Operational Research*, vol. 30, no. 1, pp. 224–238, 2023.

[10] H. Rocha, J. M. Dias, B. C. Ferreira, and M. C. Lopes, "Beam angle optimization for intensity-modulated radiation therapy using a guided pattern search method," *Physics in Medicine & Biology*, vol. 58, no. 9, p. 2939, 2013.

[11] J. Dias, R. Jaganathan, and S. Petrovic, "Determining the number of beams in 3d conformal radiotherapy: a classification approach," *Procedia Technology*, vol. 9, pp. 958–967, 2013.

[12] M. Hoffmann, J. Pacey, J. Goodworth, A. Laszczyzk, R. Ford, B. Chick, S. Greenham, and J. Westhuyzen, "Analysis of a volumetric-modulated arc therapy (vmat) single phase prostate template as a class solution," *Reports of Practical Oncology and Radiotherapy*, vol. 24, no. 1, pp. 92–96, 2019.

[13] E. M. Quan, X. Li, Y. Li, X. Wang, R. J. Kudchadker, J. L. Johnson, D. A. Kuban, A. K. Lee, and X. Zhang, "A comprehensive comparison of imrt and vmat plan quality for prostate cancer treatment," *International Journal of Radiation Oncology* Biology* Physics*, vol. 83, no. 4, pp. 1169–1178, 2012.

[14] P. Carrasqueira, H. Rocha, J. Dias, T. Ventura, B. Ferreira, and M. Lopes, "An automated treatment planning strategy for highly noncoplanar radiotherapy arc trajectories," *International Transactions in Operational Research*, vol. 30, no. 1, pp. 206–223, 2023.

[15] A. Sadeghnejad Barkousaraie, O. Ogunmolu, S. Jiang, and D. Nguyen, "A fast deep learning approach for beam orientation optimization for prostate cancer imrt treatments," *arXiv e-prints*, pp. arXiv–1905, 2019.

[16] J. Dias, H. Rocha, B. Ferreira, and M. d. C. Lopes, "A genetic algorithm with neural network fitness function evaluation for imrt beam angle optimization," *Central European Journal of Operations Research*, vol. 22, no. 3, pp. 431–455, 2014.

[17] D. Yan, F. Vicini, J. Wong, and A. Martinez, "Adaptive radiation therapy," *Physics in Medicine & Biology*, vol. 42, no. 1, p. 123, 1997.

[18] A. M. Kalet, S. M. Luk, and M. H. Phillips, "Radiation therapy quality assurance tasks and tools: the many roles of machine learning," *Medical physics*, vol. 47, no. 5, pp. e168–e177, 2020.

[19] K. L. Moore, "Automated radiotherapy treatment planning," in *Seminars in radiation oncology*, vol. 29, pp. 209–218, Elsevier, 2019.

[20] J. Dias, H. Rocha, P. Carrasqueira, B. Ferreira, T. Ventura, and M. Lopes, "Operations research contribution to totally automated radiotherapy treatment planning: A noncoplanar beam angle and fluence map optimization engine based on optimization models and algorithms," *Operations Research for Health Care*, p. 100378, 2023.

[21] W. D D'Souza, R. R. Meyer, and L. Shi, "Selection of beam orientations in intensity-modulated radiation therapy using single-beam indices and integer programming," *Physics in Medicine & Biology*, vol. 49, no. 15, p. 3465, 2004.

[22] D. M. Aleman, H. E. Romeijn, and J. F. Dempsey, "A response surface-based approach to beam orientation optimization in imrt treatment planning," in *IIE Annual Conference. Proceedings*, p. 1, Institute of Industrial and Systems Engineers (IISE), 2006.

[23] E. K. Lee, T. Fox, and I. Crocker, “Simultaneous beam geometry and intensity map optimization in intensity-modulated radiation therapy,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 64, no. 1, pp. 301–320, 2006.

[24] D. Djajaputra, Q. Wu, Y. Wu, and R. Mohan, “Algorithm and performance of a clinical imrt beam-angle optimization system,” *Physics in Medicine & Biology*, vol. 48, no. 19, p. 3191, 2003.

[25] H.-M. Lu, H. M. Kooy, Z. H. Leber, and R. J. Ledoux, “Optimized beam planning for linear accelerator-based stereotactic radiosurgery.,” *International journal of radiation oncology, biology, physics*, vol. 39, no. 5, pp. 1183–1189, 1997.

[26] Y. Li, D. Yao, J. Yao, and W. Chen, “A particle swarm optimization algorithm for beam angle selection in intensity-modulated radiotherapy planning,” *Physics in Medicine & Biology*, vol. 50, no. 15, p. 3491, 2005.

[27] X. Wu, Y. Zhu, J. Dai, and Z. Wang, “Selection and determination of beam weights based on genetic algorithms for conformal radiotherapy treatment planning,” *Physics in Medicine & Biology*, vol. 45, no. 9, p. 2547, 2000.

[28] J. Dias, H. Rocha, T. Ventura, B. Ferreira, and M. d. C. Lopes, “Automated fluence map optimization based on fuzzy inference systems,” *Medical physics*, vol. 43, no. 3, pp. 1083–1095, 2016.

[29] M. Alber and F. Nüsslin, “Optimization of intensity modulated radiotherapy under constraints for static and dynamic mlc delivery,” *Physics in Medicine & Biology*, vol. 46, no. 12, p. 3229, 2001.

[30] G. Bednarz, D. Michalski, C. Houser, M. S. Huq, Y. Xiao, P. R. Anne, and J. M. Galvin, “The use of mixed-integer programming for inverse treatment planning with pre-defined field segments,” *Physics in Medicine & Biology*, vol. 47, no. 13, p. 2235, 2002.

[31] C. Cotrutz and L. Xing, “Segment-based dose optimization using a genetic algorithm,” *Physics in Medicine & Biology*, vol. 48, no. 18, p. 2987, 2003.

[32] C. Men, H. E. Romeijn, Z. C. Taşkın, and J. F. Dempsey, “An exact approach to direct aperture optimization in imrt treatment planning,” *Physics in Medicine & Biology*, vol. 52, no. 24, p. 7333, 2007.

[33] H. E. Romeijn, R. K. Ahuja, J. F. Dempsey, and A. Kumar, “A column generation approach to radiation therapy treatment planning using aperture modulation,” *SIAM Journal on Optimization*, vol. 15, no. 3, pp. 838–862, 2005.

[34] T. Kalinowski, “A duality based algorithm for multileaf collimator field segmentation with interleaf collision constraint,” *Discrete Applied Mathematics*, vol. 152, no. 1-3, pp. 52–88, 2005.

[35] D. M. Shepard, M. A. Earl, X. A. Li, S. Naqvi, and C. Yu, “Direct aperture optimization: a turnkey solution for step-and-shoot imrt,” *Medical physics*, vol. 29, no. 6, pp. 1007–1018, 2002.

[36] X. Zhang, X. Li, E. M. Quan, X. Pan, and Y. Li, “A methodology for automatic intensity-modulated radiation treatment planning for lung cancer,” *Physics in Medicine & Biology*, vol. 56, no. 13, p. 3873, 2011.

[37] M. Zarepisheh, T. Long, N. Li, Z. Tian, H. E. Romeijn, X. Jia, and S. B. Jiang, “A dvh-guided imrt optimization algorithm for automatic treatment planning and adaptive radiotherapy replanning,” *Medical physics*, vol. 41, no. 6Part1, p. 061711, 2014.

[38] R.-P. Li and F.-F. Yin, “Optimization of inverse treatment planning using a fuzzy weight function,” *Medical physics*, vol. 27, no. 4, pp. 691–700, 2000.

[39] Q. Jia, Y. Li, A. Wu, F. Guo, M. Qi, Y. Mai, F. Kong, X. Zhen, L. Zhou, and T. Song, “Oar dose distribution prediction and geud based automatic treatment planning optimization for intensity modulated radiotherapy,” *IEEE Access*, vol. 7, pp. 141426–141437, 2019.

[40] J. Fan, J. Wang, Z. Chen, C. Hu, Z. Zhang, and W. Hu, “Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique,” *Medical physics*, vol. 46, no. 1, pp. 370–381, 2019.

[41] X. Li, J. Zhang, Y. Sheng, Y. Chang, F.-F. Yin, Y. Ge, Q. J. Wu, and C. Wang, “Automatic imrt planning via static field fluence prediction (aip-sffp): a deep learning algorithm for real-time prostate treatment planning,” *Physics in Medicine & Biology*, vol. 65, no. 17, p. 175014, 2020.

[42] B. W. Schipaanboord, M. K. Giżynéska, L. Rossi, K. C. de Vries, B. J. Heijmen, and S. Breedveld, “Fully automated treatment planning for mlc-based robotic radiotherapy,” *Medical Physics*, vol. 48, no. 8, pp. 4139–4147, 2021.

[43] R. Bijman, L. Rossi, A. W. Sharfo, W. Heemsbergen, L. Incrocci, S. Breedveld, and B. Heijmen, “Automated radiotherapy planning for patient-specific exploration of the trade-off between tumor dose coverage and predicted radiation-induced toxicity—a proof of principle study for prostate cancer,” *Frontiers in Oncology*, vol. 10, p. 943, 2020.

[44] S. Breedveld, P. R. Storchi, P. W. Voet, and B. J. Heijmen, “icycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar imrt plans,” *Medical physics*, vol. 39, no. 2, pp. 951–963, 2012.

[45] G. Wortel, D. Eekhout, E. Lamers, R. van der Bel, K. Kiers, T. Wiersma, T. Janssen, and E. Damen, “Characterization of automatic treatment planning approaches in radiotherapy,” *Physics and Imaging in Radiation Oncology*, vol. 19, pp. 60–65, 2021.

[46] S. Cilla, C. Romano, V. E. Morabito, G. Macchia, M. Buwenge, N. Dinapoli, L. Indovina, L. Strigari, A. G. Morganti, V. Valentini, *et al.*, "Personalized treatment planning automation in prostate cancer radiation oncology: a comprehensive dosimetric study," *Frontiers in Oncology*, vol. 11, p. 636529, 2021.

[47] H. Yan, F.-F. Yin, H. Guan, and J. H. Kim, "Fuzzy logic guided inverse treatment planning," *Medical physics*, vol. 30, no. 10, pp. 2675–2685, 2003.

[48] S. Siddique and J. C. Chow, "Artificial intelligence in radiotherapy," *Reports of Practical Oncology and Radiotherapy*, vol. 25, no. 4, pp. 656–666, 2020.

[49] G. Francolini, I. Desideri, G. Stocchi, V. Salvestrini, L. P. Ciccone, P. Garlatti, M. Loi, and L. Livi, "Artificial intelligence in radiotherapy: state of the art and future directions," *Medical Oncology*, vol. 37, pp. 1–9, 2020.

[50] I. Soares, J. Dias, H. Rocha, L. Khouri, M. do Carmo Lopes, and B. Ferreira, "Semi-supervised self-training approaches in small and unbalanced datasets: Application to xerostomia radiation side-effect," in *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016: MEDICON 2016, March 31st-April 2nd 2016, Paphos, Cyprus*, pp. 828–833, Springer, 2016.

[51] K. Singh and M. Xie, "Bootstrap: a statistical method," *Unpublished manuscript, Rutgers University, USA. Retrieved from http://www. stat. rutgers. edu/home/mxie/RCPapers/bootstrap. pdf*, pp. 1–14, 2008.

[52] M. Kuhn, "Futility analysis in the cross-validation of machine learning models," *arXiv preprint arXiv:1405.6974*, 2014.

[53] S. Richard, "Reinforcement learning: An introduction/richard sutton, andrew g. barto," *Cambridge, Massachusetts. London, England: A Bradford Book, Second edition.–2014-2015.–338 p*, 1998.

[54] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.

[55] L. Vandewinckele, M. Claessens, A. Dinkla, C. Brouwer, W. Crijns, D. Verellen, and W. van Elmpt, "Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance," *Radiotherapy and Oncology*, vol. 153, pp. 55–66, 2020.

[56] H. Seo, M. Badiei Khuzani, V. Vasudevan, C. Huang, H. Ren, R. Xiao, X. Jia, and L. Xing, "Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications," *Medical physics*, vol. 47, no. 5, pp. e148–e167, 2020.

[57] N. J. Nilsson, "Introduction to machine learning. an early draft of a proposed textbook (1998)," *Software available at http://robotics. stanford. edu/people/nilsson/mlbook. html*, 2020.

[58] K. Sheng, “Artificial intelligence in radiotherapy: a technological review,” *Frontiers of Medicine*, vol. 14, pp. 431–449, 2020.

[59] X. Cao, J. Yang, L. Wang, Z. Xue, Q. Wang, and D. Shen, “Deep learning based inter-modality image registration supervised by intra-modality similarity,” in *Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 9*, pp. 55–63, Springer, 2018.

[60] B. E. Nelms, W. A. Tomé, G. Robinson, and J. Wheeler, “Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 82, no. 1, pp. 368–378, 2012.

[61] B. Wu, F. Ricchetti, G. Sanguineti, M. Kazhdan, P. Simari, M. Chuang, R. Taylor, R. Jacques, and T. McNutt, “Patient geometry-driven information retrieval for imrt treatment plan quality control,” *Medical physics*, vol. 36, no. 12, pp. 5497–5505, 2009.

[62] H. Lee, H. Kim, J. Kwak, Y. S. Kim, S. W. Lee, S. Cho, and B. Cho, “Fluence-map generation for prostate intensity-modulated radiotherapy planning using a deep-neural-network,” *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.

[63] A. F. Osman, N. M. Maalej, and K. Jayesh, “Prediction of the individual multileaf collimator positional deviations during dynamic imrt delivery priori with artificial neural network,” *Medical Physics*, vol. 47, no. 4, pp. 1421–1430, 2020.

[64] J. N. Carlson, J. M. Park, S.-Y. Park, J. I. Park, Y. Choi, and S.-J. Ye, “A machine learning approach to the accurate prediction of multi-leaf collimator positional errors,” *Physics in Medicine & Biology*, vol. 61, no. 6, p. 2514, 2016.

[65] W. P. Smith, J. Doctor, J. Meyer, I. J. Kalet, and M. H. Phillips, “A decision aid for intensity-modulated radiation-therapy plan selection in prostate cancer based on a prognostic bayesian network and a markov model,” *Artificial intelligence in medicine*, vol. 46, no. 2, pp. 119–130, 2009.

[66] H.-P. Wieser, E. Cisternas, N. Wahl, S. Ulrich, A. Stadler, H. Mescher, L.-R. Müller, T. Klinge, H. Gabrys, L. Burigo, *et al.*, “Development of the open-source dose calculation and optimization toolkit matrad,” *Medical physics*, vol. 44, no. 6, pp. 2556–2568, 2017.

[67] A. Jalalimanesh, H. S. Haghighi, A. Ahmadi, and M. Soltani, “Simulation-based optimization of radiotherapy: Agent-based modeling and reinforcement learning,” *Mathematics and Computers in Simulation*, vol. 133, pp. 235–248, 2017.