Acceleration of online adaptive workflows for state-of-the-art and emerging radiotherapy treatment solutions

BENJAMIN ROBERFROID

A THESIS PRESENTED FOR THE DEGREE OF DOCTOR OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES

Molecular Imaging, Radiotherapy and Oncology (MIRO)
Université catholique de Louvain (UCLouvain)
Institut de Recherche Expérimentale et Clinique (IREC)

Acknowledgements

Cette thèse de doctorat représente un long parcours, jalonné de rebondissements et d'obstacles, qui est venu entrecouper un stage clinique de physicien médical. De nombreuses personnes ont contribué, chacune à leur manière, à l'édification de ce travail, que ce soit en partageant leurs connaissances, en m'aidant à prendre du recul sur d'autres, ou par leur soutien constant. Je tiens à les remercier chaleureusement.

Edmond, pour avoir su éveiller ma passion pour la physique médicale dès les premiers cours suivis ainsi que de m'avoir donné l'opportunité de pouvoir plonger dans ce domaine de recherche. Ton enthousiasme et ta capacité à transmettre cette passion ont été une véritable source d'inspiration tout au long de ce parcours.

John, pour ta sérénité contagieuse, et ta capacité à apporter la lumière dans les brainstormings les plus complexes, le tout saupoudré de références … assez « geeky ». Les réunions n'auraient définitivement pas été les mêmes sans tes interventions éclairantes.

Ana, pour être un vrai phare dans les ténèbres d'une thèse de doctorat ainsi qu'un pilier de cohésion d'équipe. Honnêtement, que ce soit au niveau technique, professionnel ou social, cette thèse n'aurait pu aboutir sans ton aide et ta bienveillance. Il va maintenant falloir apprendre à faire sans une Ana dans ma future aventure canadienne.

I also wish to express my deep appreciation to the jury members for their considerable investment of time and expertise in evaluating this work. Their insightful contributions and careful analysis have been instrumental in shaping these various studies into a cohesive work.

Aux collègues de MIRO, pour votre bienveillance quotidienne, les afterworks mémorables (qui aurait pu parfois être reclassifiés en "before work"), sans oublier les nombreux gâteaux pour toute occasion qui s'y prêtait de près ou de loin.

Ariane, pour ces tranches de rire partagées : les "food ranking", les parties de Sutom, les "un film, un jour" (, les canaux du monde entier), parmi tant de choses qui ont rendu cette aventure doctorale plus légère et joyeuse.

L'équipe physique du -4, pour votre bonne humeur et le partage généreux de vos connaissances, tant pendant mon stage que durant la thèse. Ce fut un véritable privilège d'avoir pu travailler à vos côtés et d'apprendre de votre expertise au quotidien.

À tous mes amis, sur qui j'ai pu compter tout au long de ces années, qui avez toujours été prêts à aider et à trouver des solutions, même sans exactement comprendre de quoi cette thèse retournait.

Enfin, mes parents, pour votre soutien indéfectible tout au long de mes études et de mes premiers pas dans la vie professionnelle.

Abstract

Online adaptive radiotherapy (OART) represents a significant advancement in cancer treatment, enabling plan adaptation to account for daily anatomical variations. However, its clinical implementation faces significant challenges, particularly in achieving time-efficient workflows while maintaining treatment quality. This thesis investigates different but complementary aspects of OART to enhance its clinical viability.

This thesis first investigates how to leverage the best value out of two critical steps of current adaptive workflow: the automatic segmentation and the automatic planning. Then it investigates how to bring efficient online adaptive workflow for two emerging treatment techniques the proton arc therapy and the conformal FLASH therapy, each presenting unique challenges for adaptation.

Through these investigations, this work contributes to both current clinical practice and emerging radiotherapy technologies, aiming to bridge the gap between OART theoretical potential and practical implementation. The findings emphasize the importance of balancing dosimetric benefits with temporal constraints, particularly for proton therapy-based treatments where adaptation time significantly impacts clinical viability.

PhD committee

Promoter

Prof. Edmond Sterpin, PhD, Ir

Université catholique de Louvain – Center of Molecular Imaging, Radiotherapy and Oncology (MIRO), Brussels, Belgium
KU Leuven – Department of Oncology, Laboratory of Experimental Radiotherapy, Leuven, Belgium
Particle Therapy Interuniversity Center Leuven – PARTICLE, Leuven, Belgium

Co-promoters

Prof. John A. Lee, PhD, Ir

Université catholique de Louvain – Center of Molecular Imaging, Radiotherapy and Oncology (MIRO), Brussels, Belgium

Prof. Ana M. Barragan Montero, PhD

Université catholique de Louvain – Center of Molecular Imaging, Radiotherapy and Oncology (MIRO), Brussels, Belgium

President of the jury

Prof. Xavier Geets, PhD, MD

Université catholique de Louvain – Center of Molecular Imaging, Radiotherapy and Oncology (MIRO), Brussels, Belgium
Cliniques universitaires Saint-Luc, Department of Radiation Oncology, Brussels, Belgium

Members of the jury

Dr. Geneviève Van Ooteghem, PhD, MD

Université catholique de Louvain – Center of Molecular Imaging, Radiotherapy and Oncology (MIRO), Brussels, Belgium
Cliniques universitaires Saint-Luc, Department of Radiation Oncology, Brussels, Belgium

Dr. Jennifer Dhont, PhD

Université libre de Bruxelles, Hôpital Universitaire de Bruxelles, Institut Jules Bordet, Department of Medical Physics, Brussels, Belgium
Université Libre De Bruxelles, Radiophysics and MRI Physics Laboratory, Brussels, Belgium

Invited members

Dr. Francesca Albertini, PhD

Center for Proton Therapy, Paul Scherrer Institute PSI, Villigen, Switzerland

Dr. Rudi Labarbe, PhD

Ion Beam Applications S.A. (IBA), Louvain La-Neuve, Belgium

Peer reviewed publications

Comparison of ethos template-based planning and ai-based dose prediction: General performance, patient optimality, and limitations.

Benjamin Roberfroid, Ana M Barragán-Montero, David Dechambre, Edmond Sterpin, John A Lee, and Xavier Geets. *Physica Medica*, 116:103178, 2023.

DIVE-ART: A tool to guide clinicians towards dosimetrically informed volume editions of automatically segmented volumes in adaptive radiation therapy.

Benjamin Roberfroid, John A Lee, Xavier Geets, Edmond Sterpin, and Ana M Barragán-Montero. *Radiotherapy and Oncology*, 192:110108, 2024.

Towards faster plan adaptation for proton arc therapy using initial treatment plan information

Benjamin Roberfroid, Margerie Huet Dastarac, Elena Borderías Villarroel, Rodin Koffeing, John A. Lee, Ana M. Barragán-Montero and Edmond Sterpin. Accepted in *Physics in Imaging in Radiation Oncology*.

Anticipating Potential Bottlenecks in Adaptive Proton FLASH Therapy: A Ridge Filter Reuse Strategy

Benjamin Roberfroid, Macarena Chocan Vera, Camille Draguet, John A. Lee, Ana M. Barragán-Montero and Edmond Sterpin. Revision in *Physics in Medicine and Biology*.

Conference communications

BHPA 2023, Brussels, BE
Evaluation of the automatic optimization engine of Varian Ethos: specificity and fine-tuning investigation, *Benjamin Roberfroid, Ana M. Barragán-Montero, Edmond Sterpin, John A. Lee, and Xavier Geets*, Oral presentation.

ESTRO 2023, Vienna, AT
Specificity and limits of Ethos template-based intelligent dose optimization engine, *Benjamin Roberfroid, Ana M. Barragán-Montero, Edmond Sterpin, John A. Lee, and Xavier Geets*, e-poster.

AAPM 2023, Houston, US
Towards Minimal Manual Corrections of Automatically Segmented Contours in Adaptive Radiotherapy: A Heatmap Guiding Clinicians to Dosimetrically Relevant Contour Corrections, *Benjamin Roberfroid, John A. Lee, Xavier Geets, Edmond Sterpin, and Ana M. Barragán-Montero*, e-poster discussion.

ESTRO physics workshop 2023, Torino, IT
Towards Minimal Manual Corrections of Automatically Segmented Contours in Adaptive Radiotherapy: A Heatmap Guiding Clinicians to Dosimetrically Relevant Contour Corrections, *Benjamin Roberfroid, John A Lee, Xavier Geets, Edmond Sterpin, and Ana M. Barragán-Montero*, Oral pitch.

BHPA 2024, Antwerp, BE
DIVE-ART: towards Dosimetrically Informed Volume Editions of automatically segmented volumes, *Benjamin Roberfroid, John A. Lee, Xavier Geets, Edmond Sterpin, and Ana M. Barragán-Montero*, Oral presentation.

ESTRO 2024, Glasgow, UK
DIVE-ART: towards Dosimetrically Informed Volume Editions of automatically segmented volumes, *Benjamin Roberfroid, John A. Lee, Xavier Geets, Edmond Sterpin, and Ana M. Barragán-Montero*, Oral presentation.

PTCOG62 ,2024, Singapore
Towards fast plan adaptation for Proton Arc Therapy, *Benjamin Roberfroid, John A. Lee, Edmond Sterpin, and Ana M. Barragán-Montero*, Oral presentation.

List of abbreviations

AI ..... Artificial Intelligence

ART ..... Adaptive Radiation Therapy

APT ..... Adaptive Proton Therapy

CBCT ..... Cone Beam Computed Tomography

CNN ..... Convolutional Neural Network

CT ..... Computed Tomography

CTV ..... Clinical Target Volume

DIR ..... Deformable Image Registration

DL ..... Deep Learning

DM ..... Dose Mimicking

DMF ..... Dose Modifying Factor

DP ..... Dose Prediction

DR ..... Dose restoration

DVH ..... Dose Volume Histogram

GAN ..... Generative Adversarial Network

GPU ..... Graphics Processing Unit

GTV ..... Gross Tumour Volume

HNC ..... Head and Neck Cancer

IGRT ..... Image Guided Radiation Thrapy

IMPT ..... Intensity Modulated Proton Therapy

IMRT ..... Intensity Modulated RadioTherapy

IOE ..... Intelligent Optimization Engine

KBP ..... Knowledge Based Planning

MC ..... Monte Carlo

ML ..... Machine Learning

MLC ..... MultiLeaf Collimator

MRI ..... Magnetic Resonance Image

MR-Linac ..... Magnetic Resonance Linear Accelerator

MU ..... Monitor Unit

NTCP ..... Normal Tissue Complication Probability

OAR ..... Organ(s) At Risk

OART ..... Online Adaptive Radiation Therapy

PAT ..... Proton Arc Therapy

PBS ..... Pencil Beam Scanning

PET ..... Positron Emission Tomography

PT ..... Proton Therapy

PTV ..... Planning Target Volume

QA ..... Quality Assurance

rCT ..... Repeated Computed Tomography

RF ..... Ridge Filter

ROI ..... Region Of Interest

RT ..... Radiation therapy

RTT ..... Radiation Therapy Technologists

sCT ..... Synthetic Computed Tomography

SGRT ..... Surface Guided Radiation Therapy

SPR ..... Stopping Power Ratio

TCP ..... Tumor Control Probability

TPS ..... Treatment Planning System

VMAT ..... Volumetric Arc Therapy

Contents

The author acknowledges the use of artificial intelligence tools. The initial draft was written entirely by the author. Subsequently, AI language models were used as writing assistants to refine sentence structure, improve clarity and correct the grammar.

Chapter 1: Introduction

1. Context and objectives

1.1. Context

From the Global Cancer statistics 2022, it is estimated that one in five humans develop cancer in their lifetime. One in nine men and one in twelve women die from it [1]. While various treatment modalities exist to combat cancer, radiation therapy (RT) is utilized in more than half of all cases.

RT is a cornerstone of modern cancer treatment that harnesses the power of ionizing radiation to combat malignant cells. Since Wilhelm Röntgen's groundbreaking discovery of X-rays in 1895, RT has undergone a remarkable evolution, transforming from a rudimentary tool into a sophisticated, precise weapon against cancer. The fundamental principle of RT remains unchanged: to deliver a lethal dose of radiation to tumor cells while sparing healthy tissues. However, the methods to achieve this goal have advanced dramatically.

Since the beginning of RT, focus has been on improving dose conformity to target. Initial approaches relied on simple two-dimensional ballistic treatments, where radiation beams were delivered from fixed angles with minimal shape adaptation and limited information from two-dimensional images. This evolved into three-dimensional conformal radiotherapy, which enabled better conformation of the radiation beam to the tumor shape with arrival of volumetric images [2]. The subsequent development of intensity-modulated radiotherapy (IMRT) [3] introduced the ability to modulate radiation intensity within each beam, significantly improving dose conformity. This progression culminated in volumetric arc therapy (VMAT) [4, 5], which exploits continuous beam rotation around the patient to create highly conformal dose distributions through optimal utilization of the whole set of beam angular positions.

However, RT has now reached a pivotal transition point. While previous developments focused primarily on spatial optimization of dose delivery through increasingly sophisticated planning techniques, the RT community has emphasized the critical importance of considering temporal patient evolution [6]. The fundamental principle of treatment fractionation in RT, while essential for balancing tumor control with normal tissue toxicity, introduces inherent uncertainties into the treatment process. The extended treatment duration, typically spanning several weeks, exposes the treatment delivery to various

anatomical variations. These changes, which may include tumor regression, weight fluctuations, and organ motion, can significantly compromise the validity of the initial treatment plan [7] (for more details, see section 6. on treatment uncertainties). Uncertainties become particularly critical for hypofractionated treatments, which deliver larger doses per fraction and are increasingly prevalent in modern RT [8-10].

Conventional RT workflows, constrained by time and resource limitations, try to maintain the same treatment plan throughout the entire course of therapy. Conventionally, this treatment plan is only updated when large anatomical changes on daily patient images are observed. This approach, while pragmatic from an operational, does not account for a large part of dynamic anatomical variations and may lead to suboptimal treatment delivery.

Adaptive radiation therapy (ART) has emerged as a solution to comprehensively address these anatomical variations with adaptation of the treatment plan [6, 11]. The concept of ART encompasses a range of strategies and technologies to efficiently modify the treatment plan according to the patient's new anatomical state. The goal is to deliver a truly personalized treatment that evolves with the patient's changing anatomy and response to therapy. ART was first proposed as "offline" ART and evolved towards what is now called "online" ART (OART) [12]. In offline adaptation, images acquired during previous treatment sessions are analyzed to identify potential anatomical variations. The treatment adaptation process occurs between treatment sessions, not necessarily the same day of the adapted treatment fraction. In contrast, OART performs plan adaptation while the patient remains on the treatment couch, thus, immediately before treatment delivery. This approach enables the treatment plan to be modified based on the patient anatomy of the day, captured through on-couch imaging. While this offers the potential for precise treatment delivery by accounting for daily anatomical variations, OART exerts significant time pressure on the treatment workflow. All steps, including image acquisition, auto-segmentation, contour verification, dose optimization, and quality assurance, must be completed within a clinically acceptable timeframe, typically 15-30 minutes, to maintain patient comfort and treatment efficiency [13, 14]. Consequently, despite its potential benefits, the clinical implementation of OART faces significant challenges that need to be

addressed to make it both efficient and reliable across various treatment modalities.

In the pursuit of minimizing radiation-induced toxicity and side effects, research has also increasingly focused on using protons instead of X-rays, resulting in what is called now the proton therapy (PT) [15]. The fundamental advantage of PT lies in the distinctive dose-depth profile of protons, characterized by the Bragg peak, a sharp peak of energy deposition at a specific depth. This stands in marked contrast to photon radiation, which exhibits an initial energy build-up followed by exponential attenuation through tissue.

As for conventional RT, PT has experienced significant technological advances enabling high-precision treatments. Notable among these is intensity-modulated proton therapy (IMPT), the proton equivalent of IMRT, which combines modulated inhomogeneous beams to achieve highly conformal dose distributions [16, 17]. IMPT is frequently implemented using pencil beam scanning (PBS), where a narrow proton beam rapidly scans a target volume layer [18]. Complete target coverage is achieved by repeating this two-dimensional scanning pattern at different proton energies, enabling target depth coverage.

While protons physical characteristics enable superior sparing of healthy tissues both proximal and distal to the target volume, it also introduces important sensitivity to treatment uncertainties. The sharp dose gradient inherent to PT Bragg peak, while therapeutically advantageous, makes treatments particularly vulnerable to anatomical variations [19, 20]. Even minor changes in tissue density or thickness along the beam path can significantly alter the range of protons, potentially leading to severe target underdosage or unexpected exposure of organs at risk. This inherent sensitivity necessitates the implementation of sophisticated elements of planning, including Monte Carlo dose calculation algorithms [21-23], comprehensive robustness analysis [20, 24], and volumetric images able to precisely provide protons energy loss through the patient [25, 26]. While these requirements already make the initial treatment planning process more challenging than for conventional RT, online adaptation (OA) also adds the important time constraint. This results in a complex situation: while PT demands adaptive approaches more urgently than conventional RT, implementing such adaptation is inherently more complex.

Nonetheless, recent years have witnessed significant progress in the implementation of OA, with solutions emerging for both conventional RT and PT. For RT, commercial platforms such as Ethos [27], Unity [28], and MRIdian [29] now offer integrated environments that combine the necessary hardware and software components to enable adaptive workflows. For PT, research groups have proposed several methods for OA and recently, for the first time one of these proposed workflows was used in clinical environment to enable high-precision OA PT treatments [30].

However, the current state of OA still falls short of its full potential. The tools, methodologies, and workflows developed to enable adaptation within clinically acceptable timeframes often represent compromises between speed and optimality. When putting aside inter-fractions errors and comparing current OART methods to conventional treatment planning approaches, where time constraints are less stringent, these adaptive solutions may produce imperfect or suboptimal results. In order to reach the full potential of OART, it is necessary to address these limitations by identifying, analyzing, and resolving the constraints that currently prevent OA from achieving its maximum potential.

Although efforts continue to establish OA in clinical practice, parallel research keeps advancing on techniques for improved conformality and reduced toxicity by exploiting machine maximal degree of freedom or leveraging recent radiobiological discoveries. However, as such new techniques emerge, it is not always guaranteed that they keep pace with the evolution of RT towards OA. Yet, the potential synergy between OA and these emerging modalities remains underexplored, despite its promise to achieve unprecedented levels of treatment quality in cancer care.

Among these, proton arc therapy (PAT) [31], which is approaching clinical implementation, represents an evolution in PT analogous to the transition from IMRT to VMAT in conventional RT. By utilizing a vast range of beam angles, PAT aims to achieve superior dose distributions with enhanced sparing of healthy tissues. However, even in its non-adaptive form, PAT faces significant computational challenges due to the complexity of determining optimal beam energy across numerous beam angles and the multitude of beam parameters.

These inherent computational demands add another layer of complexity to the already challenging field of adaptive PT.

Conformal proton FLASH therapy represents another emerging modality that introduces distinct considerations for adaptive implementation. This technique leverages the FLASH effect, a phenomenon re-discovered a few years ago where ultra-high dose-rate radiation delivery results in reduced normal tissue toxicity while maintaining tumor control [32]. The requirement to maintain ultra-high dose rates throughout treatment delivery introduces additional constraints to the planning process. Any adaptive workflow for FLASH therapy must not only account for anatomical variations but also ensure that the adapted plans maintain the necessary dose-rate characteristics to preserve the FLASH effect.

These emerging modalities exemplify how technological advances in RT and PT continue to introduce new challenges for adaptive planning. Developing efficient adaptive workflows for these techniques requires innovative solutions that can address both their unique technical requirements and the fundamental challenges of OA.

1.2. Rationale of this thesis

With OA advancing as an inevitable trend in RT and PT, there is a pressing need to develop highly efficient adaptive techniques and to generalize them to emerging techniques. However, while bringing OA for PAT or FLASH could lead to substantial improvement of treatment quality from an individual patient perspective, it is necessary to ensure that such benefits is worth the required adaptation time from the point of view of a population of patients. Considering the cost of PT facilities [33, 34], maintaining adequate patient throughput remains crucial for its economic viability and ensuring patient access to this technology [35]. This critical balance between treatment time and dosimetric benefits has recently been illustrated for OA of IMPT lung patients, where a recent study indicated that adaptive treatment sessions should be contained within 6 to 20 minutes, depending on the considered setup margins, to maintain clinical relevance with limited machine capacity [35]. Exceeding this time window would prevent many patients from accessing PT treatment quality forcing their redirection to sub-optimal alternatives, thereby reducing the average treatment

quality across the patient cohort and compromising both the clinical and economic viability of the technique.

Consequently, such a constraint adds another layer of complexity for PT-based OA techniques and highlights the necessity of optimizing at best the treatment adaptive workflow: maximizing dosimetric benefits from adaptation to justify potentially extended treatment sessions, while containing the additional adaptation time within clinically relevant windows.

Moreover, dosimetric benefits achieved through improved dose conformality with existing OA strategies for IMRT/VMAT photon treatments may mitigate the expected benefit of emerging techniques. It is therefore essential that those emerging techniques are implemented with OA solutions quickly available to maximize their potential.

Therefore, OA must be feasible and efficient for emerging techniques to ensure their viability from clinical and economical points-of-view. This means that it is necessary to accelerate OA workflows at every step when possible, whether those are specific to those emerging techniques or not. As a consequence, optimizing current state-of-the-art OA workflow components and extending them to emerging modalities are both essential. This thesis addresses these challenges through four complementary chapters.

The first two chapters focus on improving and accelerating key components of current adaptive workflows: automatic planning and contour verification. These components are critical bottlenecks in the adaptation process, where improvements in efficiency and reliability could significantly impact clinical implementation. Automatic planning often demonstrates limitations compared to manual planning [36-38], potentially compromising dosimetric benefits. Similarly, contour verification frequently requires extensive time for reviewing and manually correcting auto-segmented structures to ensure treatment quality [39, 40]. Finally, the latter two chapters explore the extension of adaptive concepts to emerging treatment modalities, PAT and FLASH therapy, which present unique challenges and opportunities for adaptation.

This comprehensive approach aims to contribute to the broader goal of making OA for emerging techniques viable, ultimately working toward improved

treatment outcomes for cancer patients. The following sections of this introduction will lay the groundwork for understanding this challenge, explaining the clear objectives of this thesis and where we currently stand in terms of technological capabilities and clinical implementation.

1.3. Objectives

The research of this thesis is motivated by the challenges associated with implementing OART quickly and efficiently for the emerging treatment solutions, seeking to improve patient outcomes through more personalized and adaptive RT and PT approaches. It is structured around four interconnected chapters (chapter 2 to 5), each addressing a critical aspect of adaptive radiotherapy.

Chapter 2 focuses on automatic planning, comparing two different but popular approaches for automatic planning in OART. The first one being the use of a wish-list of objectives embedded in a template, implemented into the commercial system Varian Ethos. The second one being using artificial intelligence (AI) to predict the three-dimensional optimal dose distribution for the patient (i.e. dose prediction), followed by an optimization process to find the machine parameters that reproduce this dose, known as dose mimicking. This investigation aims to assess the performance, patient optimality, and limitations of these approaches, while exploring the potential for more efficient and effective automatic dose optimization. Automatic planning based on a template of clinical constraints is an appealing proposition that already showed to be able to provide plan of clinical quality in a few minutes (in the case of Ethos). However, in the context of reaching the full potential of OA and maximizing the ratio of dosimetric benefits over adaptation time, satisfying clinical goals might not be sufficient. Consequently, it is legitimate to evaluate the dosimetric optimality of these automatically optimized plans. This is why we aim at challenging such approach by another very well-known automatic method: the dose prediction. To provide a complete evaluation of optimality we aimed at evaluating the potential gap of quality of such automatic methods with regard to the current gold standard: manual planning.

Chapter 3 delves into automatic segmentation correction. The automatic delineation of anatomical volumes on daily treatment images represents a critical challenge in OART. Clinicians must verify and potentially edit the automatically

segmented contours, introducing significant time delays in the workflow without clear indication of which corrections are clinically meaningful. Substantial efficiency gains could be achieved by guiding clinicians to focus on regions where manual corrections would have genuine clinical impact. We propose here to develop a novel visualization tool that will help the clinician to focus on areas where editions will have a significant impact.

Chapter 2 and 3 deal with state-of-the-art methods essential for OART and try to provide ways to tend toward a maximized value of OART. Chapter 4 and 5 focus on how to provide adaptation workflow dedicated to the two considered emerging treatment modalities.

Chapter 4 addresses the challenges of PAT emerging treatment modality, to reach compatibility with OA conditions. As OA in PT has primarily focused on IMPT, we drew inspiration from these existing techniques to develop our approach for PAT. As previously introduced, PAT presents unique technical challenges due to the numerous plan parameters to optimize, typically resulting in extended planning times. Given PAT unique computational demands, this chapter proposes reusing selected parameters from the initial treatment plan rather than fully re-optimizing them on daily images, thereby reducing calculation time.

Chapter 5 investigates adaptive workflows for conformal FLASH PT, an emerging modality that would probably require a few more years than PAT to reach clinical environments. Since FLASH effect requires high dose-rate to be triggered, current proposed implementations of FLASH treatment necessitate to make the dose delivery as fast as possible. Consequently, some elements of conventional PBS cannot be used, e.g. energy switch-ups and -downs are too slow. While existing approaches achieve high dose-rates under standard conditions, those seem hardly compatible with timing pressure of daily adaptation. Consequently, this chapter investigates a simplified workflow that make possible to achieve high-dose rates treatment and relevant plan optimization time compatible with OA.

Through these four technical chapters, this thesis aims at enabling OA for emerging PT techniques while ensuring their clinical and economic viability. This underscores a crucial principle: dosimetric gains must be weighed against their temporal requirements, highlighting the fundamental importance of accelerating OA workflows. Moreover, technological progress continues in conventional

photon-based treatments. This progress cannot be overlooked and must be integrated in emerging techniques as well to maintain their comparative advantage.

2. Principles of radiation therapy

Radiation therapy (RT) is a treatment modality that relies on fundamental principles of biology and physics. The following subsections quickly remind the specific mechanisms by which radiotherapy operates, including its effects on cancerous and healthy cells, the fractionation of treatment as well as the current techniques and workflows.

2.1. Ionizing radiation and dose fractionation

As its name implies, ionizing radiation is able to ionize matter. In RT, by directing these radiations toward tumor volume, this capability is used to kill cancerous cells through DNA damages.

Two kinds of DNA damage are generally reported: direct damage where the radiation interacts directly with the DNA strands; and indirect where ionizing radiation interacts with water molecules in the cell, leading to radiolysis. Highly reactive free radicals created through radiolysis can in turn damage DNA [41]. Nonetheless, there exist mechanisms that allow the cells to repair themselves below a reasonable damage threshold.

While ionizing radiation can break the DNA strands and kill cells, the biological response to these ionizations is not the same in healthy tissue and cancerous ones. Indeed, healthy cells have usually the faculty to repair these DNA alterations to some extent. This allows thus a significant part of the healthy cells to recover after a certain amount of time. Such repair mechanism is usually deficient for cancerous cells leading to a low proportion of them to be able to repair these deathly alterations [41].

By making use of such differential response between healthy and cancerous cells it is thus possible to deliver large total dose to the tumor while nonetheless preserving an important proportion of the healthy tissue. Consequently, a key RT concept consists in fractionating the total dose in several fractions to let a recovering time to the healthy tissue between the delivering of ionizing radiation. The challenge lies in optimizing the time interval between fractions: it must be sufficient for healthy tissue recovery and repopulation while preventing surviving tumor cells from replicating and repopulating the irradiated area.

Usually in conventional RT, a delay of one day is observed between the different fractions but it can be sometimes a bit longer or shorter without being too detrimental for the treatment quality [42]. The total delivered dose as well as the number of fractions describe the treatment fraction scheme. Such fractionation scheme is mostly determined on the basis of radiobiological parameters.

2.2. Therapeutic ratio: balancing tumor control and normal tissue toxicity

The therapeutic ratio concept in radiotherapy describes the balance between the probability of Tumor Control Probability (TCP) and the Normal Tissue Complication Probability (NTCP) [41].

TCP designates the probability of controlling the tumor as its name implies. As radiation dose increases, the probability of killing all cancer cells (tumor control) increases. This is represented by the red curve in Figure 1.

NTCP indicates the probability of inducing a complication, a “toxic” effect to the healthy tissue with the received dose. The goal in radiotherapy is thus to control the tumor volume with a chosen treatment dose while ensuring that toxicity remains under reasonable level.

The gap between these two curves designates the therapeutic index. The larger the gap, the easier it is to define a treatment to control the cancerous cells without leading to unreasonable toxic effect. RT research aims at increasing this therapeutic window. This therapeutic window is dependent on several factors, such as the tumor radiosensitivity (some tumors can be more sensitive to ionizing radiation than others), normal radiosensitivity (healthy tissues have a large varying level of tolerance to ionizing radiation), and tumor location: tumor volume near highly radiosensitive organs can reduce the therapeutic window. Nonetheless it is also possible to move the blue curve to the left and displace the red one to the right to increase the therapeutic window. The blue curve can be moved to the left with, for example, radiosensitizers: drugs that make tumor cells more sensitive to radiation; and the red curve can be moved to the right using radioprotectors agents that reduce the effect of ionizing radiation in normal tissue [43]. Consequently, RT is often jointly used with chemotherapy.

Moreover, techniques that allow to gain in precision when delivering dose to the tumor result also in a shifted red curve, since normal tissue will receive less ionizing dose [44, 45]. Understanding and optimizing the therapeutic ratio is at the heart of radiation oncology. It drives decisions about total dose, fractionation schemes, and the adoption of new technologies and techniques.

Fundamentally, the fractionation of RT treatment is one of the most important factors creating a real treatment window. Without fractionation, blue and red curves would in most cases almost overlap and the therapeutic windows would almost be non-existent.

Nonetheless, as it will be highlighted a little further in this introduction, treatment fractionation has a cost on treatment uncertainties.

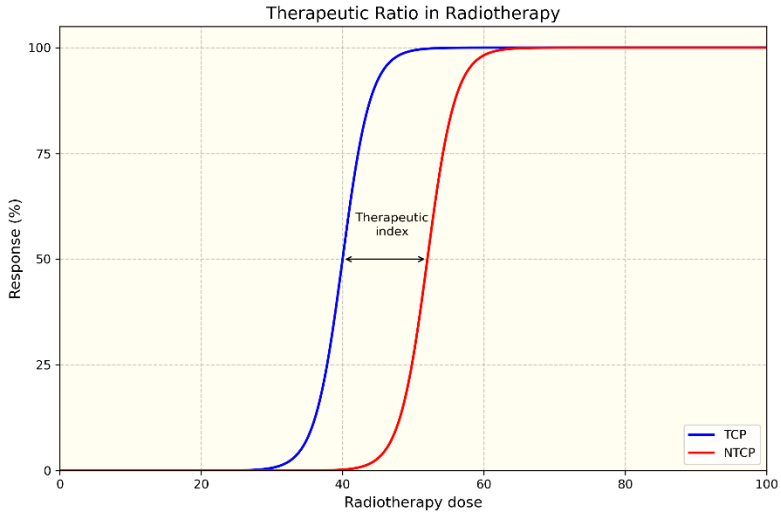


Figure 1 : TCP-NTCP curves. The blue curve is the TCP curve, the red one is the NTCP. The larger the gap between them, the higher the therapeutic index.

3. Conventional radiotherapy workflow: overview

The creation of a high-quality radiotherapy treatment plan necessitates a rigorous and multifaceted workflow. This process aims to ensure precise dose delivery to the target volume while maximizing healthy tissue sparing. The workflow comprises several critical steps, each demanding significant time and expertise.

The complexity and duration of this planning process underscore the rationale behind the conventional approach of reusing the same treatment plan for each fraction of a patient's treatment course. This practice has been a pragmatic response to the substantial time constraints inherent in the planning process.

The subsequent sections will briefly introduce the specific steps of the conventional workflow, highlighting the time-consuming nature of each component as well as the exigent quality required.

3.1. Planning imaging

The first step is the acquisition of high-quality imaging data. This serves two critical functions: providing essential information for both anatomical delineation and providing electronic density for the dose calculation. The quality of this imaging is paramount, as it forms the foundation upon which the entire treatment plan is built.

CT imaging offers precise spatial information, allowing for accurate localization of organs at risk (OARs) and target volumes. Moreover, CT data provides electron density information crucial for calculating how ionizing radiation will deposit dose within the patient's body [46, 47]. To enhance the accuracy of region of interest (ROI) delineation, clinicians often supplement CT data with additional imaging modalities, primarily magnetic resonance imaging (MRI) and positron emission tomography (PET). Each of these modalities offers unique advantages:

MRI provides superior soft tissue contrast and higher spatial resolution compared to CT, enabling more precise delineation of both tumors and critical structures [46-48], particularly in areas with complex soft tissue anatomy as illustrated in Figure 2.

PET, especially when using fluorodeoxyglucose-18F (FDG) as a tracer, offers functional imaging capabilities. By visualizing glucose metabolism, PET can highlight areas of increased cellular activity, which is particularly valuable for identifying tumor extent and potential areas of spread [46-48].

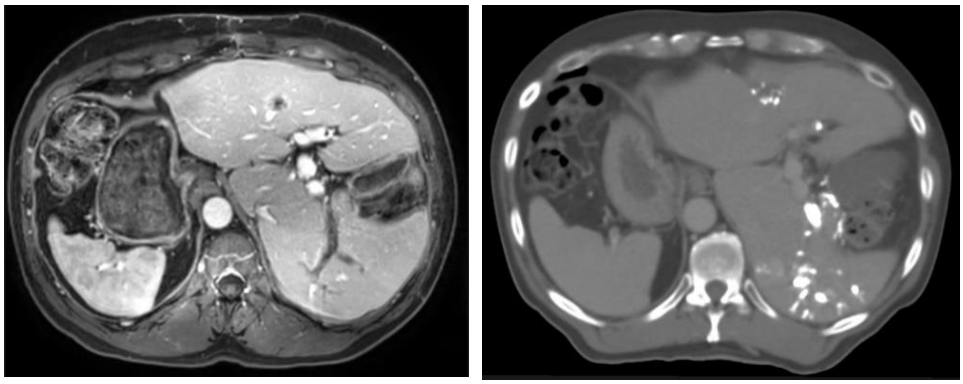


Figure 2 : Example of different imaging modalities, transversal view. Top left: MR image; top right: CT image.

3.2. Image Registration

The acquisition of multiple imaging modalities presents a significant challenge in radiotherapy planning on how to accurately visualize and compare these diverse datasets. Each imaging modality may have different origins, dimensions, and voxel values, making direct comparison problematic. Image registration algorithms address this challenge by aligning these disparate images based on spatial similarity. Two main different registration possibilities are often used: Rigid and non-rigid [49, 50].

Rigid registration provides a rotation matrix that applies 3D translation and rotation of the whole image. Initial spatial distances are thus preserved (illustrated in Figure 3).

Non-rigid registration generates a deformation field that warps the moving image to achieve optimal overlap with the reference image, without necessarily preserving initial spatial relationships. Unlike rigid registration, this process allows local deformations where different parts of the image can be transformed independently.

Note that the accuracy of such algorithms relies a lot on the initialization of the images position, i.e. if the algorithm is launched with the images disposed in a highly different position, it might never converge to a satisfying solution. Moreover, when significant anatomical changes occur in the second image, such

as the appearance of large gas pockets in the bowel or imaging artifacts, the algorithm may produce unrealistic displacement vectors, particularly when using non-rigid algorithms for local deformation [49].

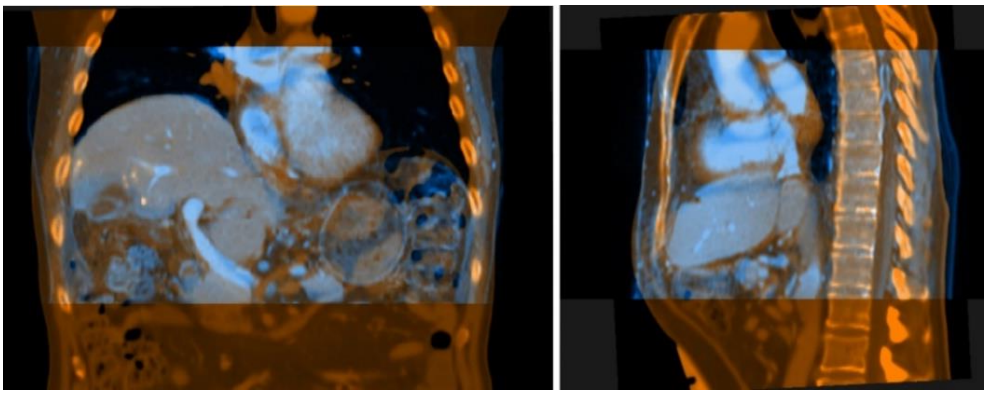


Figure 3: Illustration of a rigid registration process between the previous MR and CT images. Orange signal: CT; Blue signal: MR. Left: coronal view; Right: sagittal view.

3.3. Segmentation

Once the CT is acquired, it is necessary to precisely localize all the important anatomical volumes such as the OAR and the tumor volume. To do so, physicians generally manually segment these different volumes on the planning image. Since a CT is a volumetric image, they proceed slice by slice: they review each transversal slice of the 3D matrix and segment each of these volumes on every slice (illustrated in Figure 4). This is thus a time-consuming step necessitating knowledge about anatomy and interpretation of medical image [51, 52]. Though, the contrast of a CT is limited and sometimes some soft tissue might not be clearly visible on it. In such case, other image modalities such as MRI or PET can be used [46].

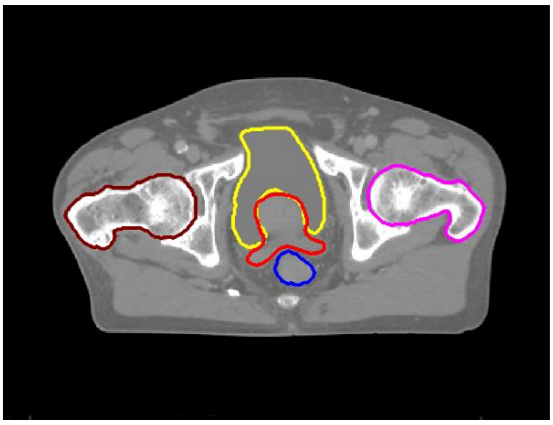


Figure 4 : blue: rectum; pink: left femoral head; dark red: right femoral head; yellow: bladder; red: target.

Besides the OAR delineation, it is important to note that different kinds of target volume exist for RT treatment. We define the gross tumor volume (GTV) as the visible or palpable extent of malignant growth. The clinical target volume (CTV) encapsulates the GTV plus adds a margin to account for potential microscopic spread of cancerous cells that are not clearly visible [53]. The planning target volume (PTV) encapsulates the CTV and adds a treatment security margin to mitigate potential treatment errors [54] as illustrated in Figure 11. Unlike the CTV, the PTV does not represent an anatomical structure but rather a geometric tool for treatment planning. Consequently, when treatment margins exceed the spatial separation between the CTV and nearby organs at risk (OARs), the PTV may overlap with these critical structures. This geometric expansion and its potential implications for treatment planning are further elaborated in the section "Van Herk formalism."

3.4. Plan set-up and optimization

The dose planning is generally a manual step where a physician and/or a dosimetrist define the ballistic of the treatment, i.e., the beam angles, the energy of the beam, the optimization constraints [55, 56]. They generally use a treatment planning system (TPS) where the treatment machine is fully modelized and where several optimization tools help them to generate a high-quality treatment plan that meets at best the clinical constraints. For modern treatments inverse planning is typically used. With inverse planning, the planner defines the set of constraints to be met, then the TPS iteratively optimizes treatment machine

parameters that allows to get the closest to these input constraints. The problem can be defined as followed:

$$\min \sum_{i=1}^{n} w_i f_i\,(d) \tag{1}$$

With *n* cost functions *f* dependent of the dose distribution *d* and weighted by non-negative importance parameters *w*. In practice, the clinicians try different cost functions *f,* e.g. constraints on minimum or maximum dose on a specific volume, and importance *w* and evaluate how the optimizer manage to minimize this weighted sum of cost function, i.e. to get close to the inputted clinical goals.

Consequently, this is also an iterative process for the human planner since the initial set of used constraints cannot be ensured to be the most optimal one. The planners are usually required to iteratively refine the optimization constraint they use to ensure that the dose distribution is close to optimality. Such a process usually takes from tens of minutes to several hours.

3.5. Quality Assurance

Before delivering the treatment, complex plans have to pass through a patient-specific QA. It usually involves measuring the real dose in phantom to verify the accuracy of the dose delivery [57].

Another approach is a second, independent dose calculation where the dose is recomputed with a different approach than the initial one [58, 59]. For example, it can be using another algorithm of dose calculation, and/or using the log files of the machine to reconstruct the dose distribution from it.

Since treatment planning system models may not account for all dose-influencing parameters, the accuracy of calculated dose distributions might differ from actual delivery, particularly in complex cases. Therefore, quality assurance measurements are necessary to verify the agreement between calculated and delivered dose distributions.

3.6. Treatment delivery

Once the treatment plan has gone through all these steps, it can be delivered to the patient. There are usually several days between the planning imaging step and

the first day of treatment delivery. Nonetheless, the patient is placed in the same position as for the imaging planning step, and after some adjustments (see section about in-room imaging) the treatment can start. The patient will come back to the treatment center several days until the last treatment fraction, with usually a maximum of 1 fraction per day.

4. Conventional types of external radiotherapy

4.1. Conventional X-ray therapy

External photon beam RT (so-called conventional RT) is by far the most predominant treatment by ionizing radiation [60]. It is an approach where the tumor is treated with a source of high energy X-rays placed outside of the patient. In state-of-the-art treatment units, the X-rays are produced with a linear accelerator (linac) that can fit inside a treatment room. Linac accelerates electrons until they reach high energies to then produce X-rays through the collision of them onto a material of high-atomic-number (called “bremsstrahlung”). A Modern linac can usually accelerate electrons to different energies producing X-rays of different spectral range. Such energy usually ranges from 6 to 18 MV. These X-rays are then directed to the target volume, i.e. the tumor volume.

To precisely conform the delivered dose to the target, modern linacs are equipped with a multi-leaf collimator (MLC) in the gantry snout [61, 62]. The MLC consists of numerous independently movable leaves, typically made of tungsten for its high atomic number and density. These leaves are often from 2 to 15 mm width [63] and can move in and out of the ionizing beam, effectively shaping the beam to conform to the tumor's outline from any given angle as illustrated in Figure 5. The precise control of these MLC leaves is crucial in achieving highly conformal dose distributions. Several techniques of MLC and gantry angle positioning have been developed over the years to enable highly conformal treatments.

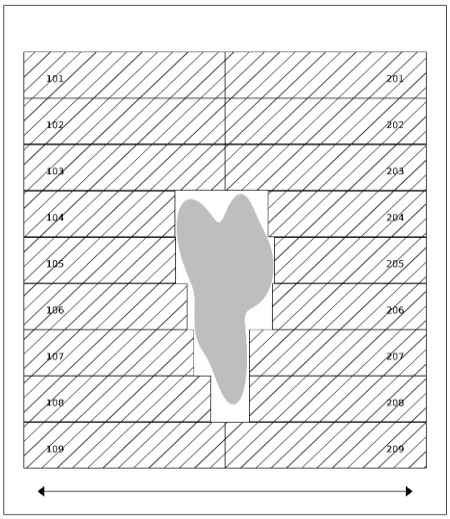


Figure 5 : Illustration of the MLC from a beam-eye-view. The gray shape is the target volume that is going to be irradiated; the dashed rectangles are the different leaves constituting the MLC. Notice how they can be easily disposed around the target to conform the dose to such volume.

It is possible to shape the beam with fixed MLC for different beam angles, such treatment is called 3D-conformal treatment since by distributing the dose to several angles and by collimating the beam on the tumor it enables a 3D conformity around the target [2].

It is possible to go further by combining several beam angles that have several different MLC position sequences not necessarily collimated to the target to create an intensity modulated radiation therapy (IMRT) treatment [3]. With such IMRT treatment, the different beams alone only provide inhomogeneous dose not usable for treatment, while when all combined together it is then possible to achieve a highly conformal treatment.

Volumetric modulated arc therapy (VMAT) extends the IMRT concept by enabling continuous gantry rotation while simultaneously modulating dose rate, gantry speed, and MLC positions [4]. This technique provides more degrees of freedom than fixed-field IMRT, potentially achieving superior dose conformity and reduced treatment times.

Consequently, the process of optimizing a photon therapy treatment involves determining the optimal set of MLC positions for each beam angle or arc segment, given a chosen photon energy.

4.2. Proton therapy

Proton therapy (PT) is another kind of treatment with ionizing particles. While less widespread than conventional RT with X-rays, it is more and more used as a powerful alternative treatment modality. It delivers treatment using high-energy protons generated by synchrotrons or cyclotrons to irradiate target volumes.

The main strength of this treatment modality is its dose deposition pattern, which differs from that of photons. Protons deposit their energy in a distal peak, the Bragg peak, in contrary to photons where the dose deposition consists in a smoother curve presenting no real ending point [64]. Consequently, while it is possible to cover a target volume with a monoenergetic photon beam, PT necessitates a distribution of the Bragg peak within target region. To do so, several protons energies can be combined to achieve what is called a "spread-out Bragg peak" where the summation of these different Bragg peaks can lead to a dose plateau along the tumor depth as illustrated in Figure 6. This spread-out-Bragg peak is usually much more advantageous than dose deposition pattern of X-rays in term of dose received by the normal tissues.

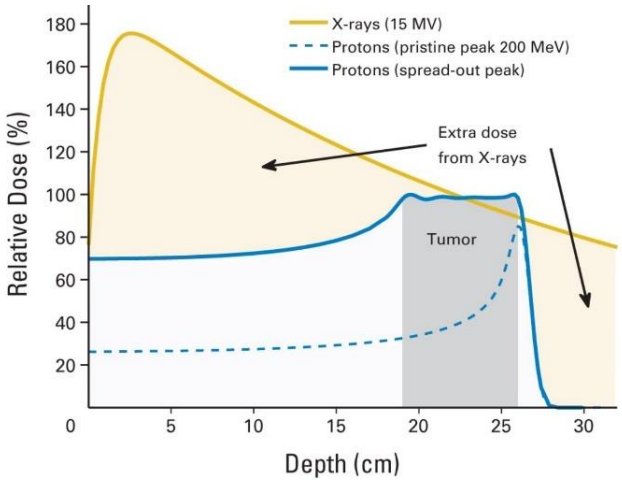


Figure 6 : Illustration of the depth-dose profile for X-rays, in yellow, and protons, in blue. The dashed line represents a monoenergetic Bragg peak, the solid blue line a spread-out Bragg peak achieved with the cumulation of several Bragg peaks of different energy. Adapted from [64].

There are two conventional approaches to generate spread-out Bragg peaks in proton therapy: passive scattering and pencil beam scanning (PBS) [65-67]. Passive scattering employs a patient-specific range modulator with a field collimator and broad proton beam. In contrast, For PBS a thin proton beam is used to laterally scan different layers of the target. The lateral covering of the target is thus ensured by deflecting the proton beam under magnetic control while target depth covering is achieved by modulating the energy of the proton beam. More precisely, the target volume is covered by a grid of treatment spots that are sequentially scanned. Classical PT plan optimization involves determining the optimal weight of each spot, representing the number of protons to be delivered at each position.

PBS has emerged as the current state-of-the-art technique, surpassing passive scattering due to its superior target conformity and elimination of patient-specific hardware requirements [68].

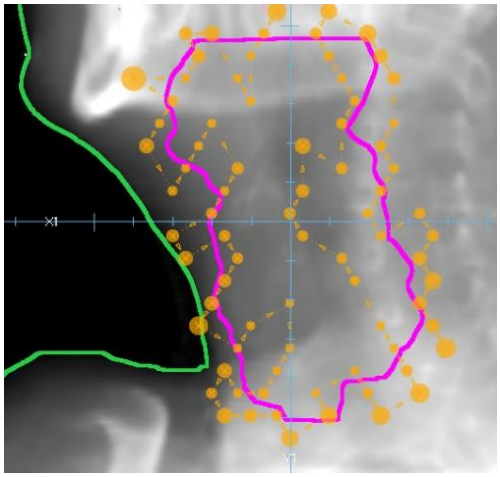


Figure 7: Example of proton treatment spots in beam-eye-view. Orange dots represent the treatment spots of a specific energy layer. The size of the spot represents its weight (which is proportional to the number of protons to be sent to that spot). Pink contour is the target, green contour is the body contour. Dotted orange lines with arrows indicate the spots scanning path.

4.2.1. Intensity modulated proton therapy

Following a similar reasoning than for IMRT, it is possible in PT to combine several inhomogeneous beams from different angles into a highly conformal treatment plan [16, 17]. Such an approach is called intensity modulated proton therapy

(IMPT) and is nowadays the gold standard of PT. Conventionally, IMPT is jointly used with PBS. Consequently, the different beam angles generally use several treatment spots and several energy layers.

4.2.2. Proton Arc Therapy

As IMPT followed a similar reasoning than IMRT, proton arc therapy (PAT) tends to follow a similar idea than VMAT. PAT goal is to efficiently deliver proton beams from a vast range of beam angles around the patient with a beam rotation [31]. However, in contrary of VMAT, PAT is not yet deployed in clinical conditions and is still considered as an emerging treatment technology. Indeed, the problem with PAT is to determinate the best energy layers per angle while for VMAT the question of beam energy does not exist. On the one hand, to let the treatment time remains in reasonable timing range, it is necessary to be cautious on how many energy layers per angle and how many energy switch-ups (those are much slower due to magnetic hysteresis effect in the delivery system) will be used. On the other hand, to ensure the high quality of the dose distribution it is tempting to use as many energy layers as possible to benefit from a high number of degrees of freedom. PAT energy layer selection is thus a complicated balance between beam delivery time and plan dosimetric quality. This problem has recently been classified as NP-hard [69], indicating that it becomes exponentially more difficult to solve as its size increases, to the point where finding a perfect solution quickly becomes impossible in practice. Consequently, it may require significant time and effort before an efficient near-optimal solution could be developed and implemented. Nonetheless some research groups have already tried to tackle the problems from different perspectives and achieved interesting results [31, 70].

The potential of such modality is illustrated in Figure 8 with the dedicated PAT optimization algorithm embedded in the commercial TPS RayStation (Stockholm, Sweden). On this figure, we can clearly observe the dose conformality to the target (blue and red contours) of the PAT plan in comparison to an IMPT plan for the same head and neck patient.

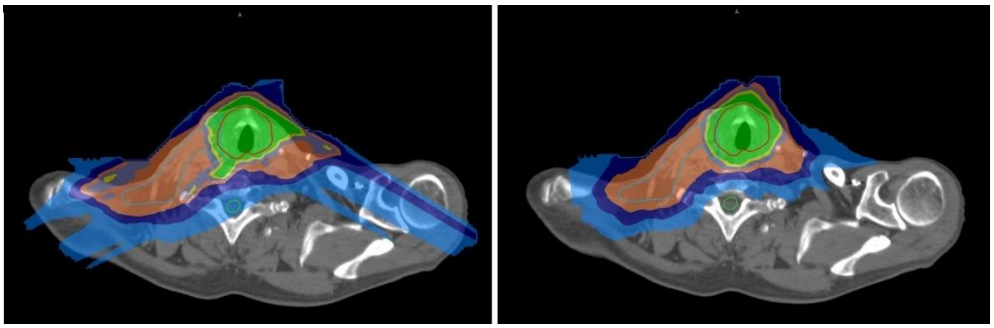


Figure 8 : Illustration for a head and neck patient of the difference of the distribution of the dose between IMPT and PAT. The highest dose level is depicted in green, then yellow, then lavender, then orange, then dark blue. The lowest level is in light blue.

4.2.3. FLASH therapy

In 2014 Favaudon et al. [32] rediscovered a radiobiological phenomenon, the FLASH effect. FLASH effect refers to a sparring of normal tissue for the same tumor control when using high dose-rate, i.e. >40Gy/s. The effect of the FLASH effect is often summarized with the dose modifying factor (DMF) indicating which dose is necessary to achieve a same effect in conventional dose-rate and in ultra-high dose-rate conditions:

$$\mathrm{DMF} = \frac{D_{UHDR}}{D_{CONV}}\Big|_{isoeffect} \tag{2}$$

with D_{UHDR} being the dose at ultra-high dose-rate and D_{CONV} being the dose at conventional dose-rate. A DMF >1 representing thus a sparing of normal tissue due to FLASH effect.

Such effect has been demonstrated with various types of ionizing radiation, including electrons, photons, and protons, and across different normal tissues such as brain, lung, and skin. While electrons remain limited to superficial treatments due to their low penetration depth, both photons and protons offer the capability to treat deeper targets. Nonetheless, achieving FLASH effect at human isocenter depth seems currently challenging for photon systems [71, 72] while already feasible for PT systems [73, 74]. Moreover, the possibility of beneficiating from a synergistic combination of toxicity reduction through FLASH effect and PT reduced integral dose generates strong interest.

While such treatment modality is not close yet to be included in clinical routine, it has the potential to become one of the main breakthroughs in the following decades and might potentially shift the paradigm of maximization of dose conformity to a paradigm of maximization of the dose-rate.

Nonetheless, clinical delivering is already being studied, especially for PT (for the previous reasons). For FLASH PT, PBS with switching of energy layers cannot be used due to the energy switching time being too long to ensure a dose-rate compatible with FLASH. Consequently, two alternative modalities employing monoenergetic PBS have been proposed: transmission mode and conformal mode.

Transmission approach employs high-energy beams such that the Bragg peaks are positioned distal to the patient's anatomy. This configuration facilitates the optimization of treatment plans while maintaining high dose rates.

Conformal approach achieves superior conformity by precisely positioning treatment spots within the target volume range. It incorporates a patient-specific Ridge filter (RF), a device presenting a grid of thin peaks placed in the gantry snout allowing to generate a spread-out Bragg peak within the target region with a monoenergetic beam. The RF is a 3D-printed device placed in the proton beam path consisting in a grid of thin peaks that modulate the range of incoming protons. However, the manufacturing of such RF is time-intensive which might not well fit in a clinical routine. Figure 9 illustrates a modelized RF in 3D and its placement with regards to target volume in a Beam-eye-view.

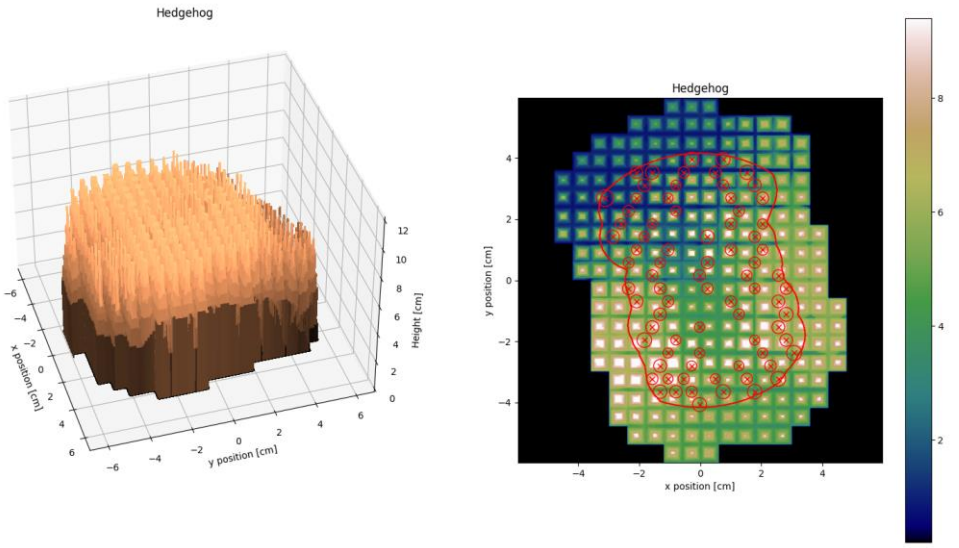


Figure 9 : Illustration of the ridge filter, also called "hedgehog", used in conformal FLASH therapy. Left: 3D spatial visualization; Right: 2D Beam-eye-view visualization. The red contour is the target, the red circles with cross are the treatment spots and colormap refers to the peak height, black being 0 cm, white being 10cm.

5. Dose calculation

To deliver high quality RT and PT treatment it is essential to know how the dose will distribute in the human body. To do so, there exists two main methods: Monte-Carlo simulation and analytical dose algorithm.

5.1. Monte-Carlo

The core concept of Monte-Carlo simulation is to use a random number generator in order to simulate a large number of particle histories [21, 23, 75]. These particle histories are modeled through incremental steps of traveled distance. At each step, potential physical interactions are sampled from their respective probability density functions. Such methods are known to present excellent accuracy when a sufficient number of individual particles are simulated. In terms of accuracy, they are often considered as a gold standard.

While being a gold standard of dose distribution accuracy, MC usually necessitates the simulation of a very high number of particles in order to reach a reasonable

level of dose uncertainty. Such numerous simulations of particles can be too slow in specific clinical conditions.

5.2. Analytical algorithms

Due to the inherent computationally intensiveness of MC engines research quickly focused on providing a way to quickly compute an accurate dose computation. It resulted in a rather indirect approach based on certain assumptions about the dose deposition that we often depict as “analytical algorithms” [76]. These algorithms model the dose distribution process using simplifications of the complex physical interactions involved in radiation transport and energy deposition. While generally considered less accurate than MC simulations, they offer a significant advantage in computational efficiency while being able to provide sufficient accuracy for RT clinical conditions. Such speed being often crucial in clinical settings where time constraints are a critical factor in treatment planning and delivery.

Most photon analytical algorithms lie on a superposition principle where TERMA (total energy released per mass unit) is first determined, e.g., with a raytracing of primary radiation through the body, and then followed by a superposition of dose spread kernel. We can for example cite the Varian Analytical Anisotropic Algorithm [77] (AAA) applying lateral scaling of the beam according tissue density or the collapsed-cones convolution algorithm discretizing the point-spread kernel into cones collapsed on a same point and scalable with the tissue density [78, 79]. These are still commonly used in clinical conditions for conventional RT.

Since a few years, another kind of analytical model is commonly used in clinical conditions: the Linear Boltzmann Transport Equation-solvers (LBTE) [80]. Such algorithms consist in discretizing the patient volume and numerically solving a set of differential equations for the transport of photons and electrons. These approaches usually benefit from an accuracy closer to the MC ones.

Such analytical algorithms also exist for PT. As the pencil beam analytical algorithm or the ray-casting one approaching the problem by a range dilution effect [81]. While enabling a fast calculation of a PT dose, such analytical algorithms are considered as less accurate than the ones for RT, especially for the distal Bragg peak. Lot of PT centers prefer thus to compute the dose through a MC engine even if it represents a significantly slower approach.

6. Treatment uncertainties

As previously explained, fractionation is a fundamental principle in RT and PT, necessitating treatment delivery over multiple days. During this extended period, various anatomical changes can occur between the initial CT simulation and subsequent treatment days, as illustrated in Figure 10. These changes can impact treatment accuracy and effectiveness [82].

- Organs motion: organs are subject to motion over different time scale, e.g. as for lung or cardiac motion, bladder filling.
- Target motion: due to motion of close organs, the target position can also deviate from the initial observed position.
- Target shrinkage: during the course of treatment, it is possible to observe a shrinkage of the tumor, due to the treatment itself.
- Patient evolution: global body change can also be observed over the course of the treatment, e.g., fat and/or muscle loss.

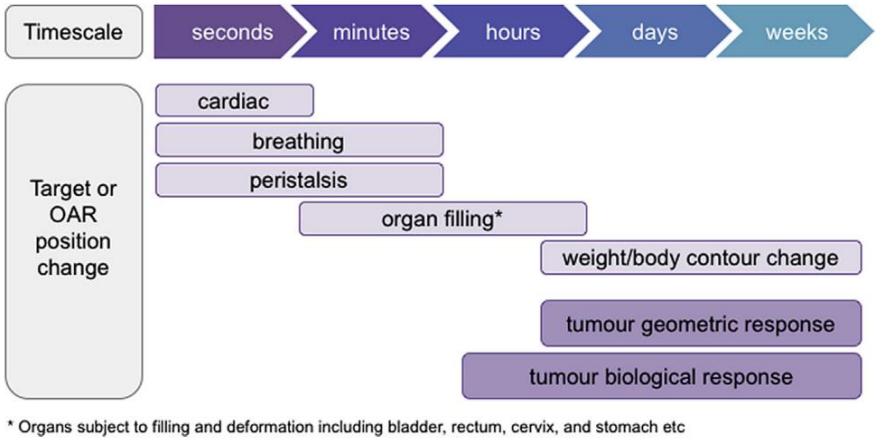


Figure 10 : Illustration of anatomical changes occurring at different time scales. Adapted from [7]

Nonetheless, anatomical variations are not the only kind of uncertainties that can impact the treatment delivery:

- Errors of target and OAR delineation: When delineating on the CT image, the limited resolution and contrast of the image itself can impact the process. It produces an uncertainty where it is usually assumed that contour cannot be determined more precisely. Moreover, the contour

and the interpretation of the image is subject to inter-observer variability [83]. It means that if several physicians with a similar background have to delineate a same contour on a same image, variations can often be observed between both contours. There exists also uncertainty in defining CTV boundaries, as precisely determining the extent of microscopic disease spread remains challenging.

- Setup errors: as previously explained the whole treatment is dependent on the very first step of the workflow: the simulation image. Consequently, if the patient simulation setup is not perfectly reproduced for each of its treatment fractions this creates setup discrepancies. This could be a consistent shift of the couch or consistent mis-positioning on the treatment couch.

As these errors and changes occur at different time scales, it is often convenient to either classify them as systematic (error systematically repeated during the whole treatment) or as random (whose influence randomly changes over the treatment duration). For example, segmentation errors are considered as pure systematic errors in a conventional RT workflow (without adaptation) since their impact is determined at the planning step, while organ motion entails a random contribution since its impact is generally different from a fraction to another [54, 82]. Table 1 displays some examples of random and systematic components of treatment errors.

Table 1: Examples of random and systematic errors in (non-adaptive) RT treatment.

Error Type	**Random Component**	**Systematic Component**
Patient Setup	Daily variations in patient positioning	Consistent misalignment of patient
Organ Motion	Breathing, peristalsis, bladder filling, cardiac movements	Organ position shift between planning and the rest of treatment course
Structure Delineation	N/A	Error of delineation at the planning step
Dosimetric	Fluctuations of machine output between fractions	Beam calibration offset

Solutions and work-around exist to deal with the noxious effects of these various uncertainties and are described in the following sections.

6.1. Management of uncertainties

6.1.1. Van Herk margin formalism

To account for these errors, the Van Herk treatment margin formalism, introduced by Marcel van Herk and colleagues in 2000 [54] is often used in RT. It provides a systematic approach to determining Planning Target Volume (PTV) margins in radiotherapy. This formalism addresses the need for a rigorous, quantitative method to account for uncertainties in RT planning.

The core principle of the Van Herk formalism is the distinction between systematic and random errors in RT delivery, described in the previous section.

The formalism is expressed mathematically as:

$$M = \alpha\Sigma + \beta\sigma \tag{3}$$

Where:

- M is the margin around the Clinical Target Volume (CTV) to create the PTV;
- Σ is the standard deviation of the distribution of systematic errors;
- σ is the standard deviation of the distribution of random errors;
- α and β are coefficients derived from the desired confidence level and coverage probability.

Van Herk et al. proposed values of α = 2.5 and β = 0.7 to ensure that 90% of patients receive a minimum cumulative CTV dose of 95% of the prescribed dose. These values assume a 3D Gaussian distribution of errors and consider both translational and rotational uncertainties.

The strength of this approach lies in its ability to separate the impacts of systematic and random errors, recognizing that systematic errors have a more significant effect on the required margin. This distinction allows for more precise margin calculations and potentially smaller PTVs compared to simplistic, uniform expansion methods.

Implementation of the Van Herk formalism requires careful measurement and analysis of setup errors and organ motion specific to each institution and treatment site.

Due to its simplicity of implementation the Van Herk margin is a widespread technique to deal with treatment errors. Currently it is mostly used for conventional RT. Indeed, in particle therapy due to the high distal dose gradient in dose deposition, such errors do not simply lead to a rough displacement of the cloud dose as it could be assumed in conventional RT. In PT, such errors often lead to significant dose distortion preventing use of PTV margin in numerous cases [84, 85]. Nonetheless, for PT, robust optimization represents a valid alternative. Such tool is described in the following section.

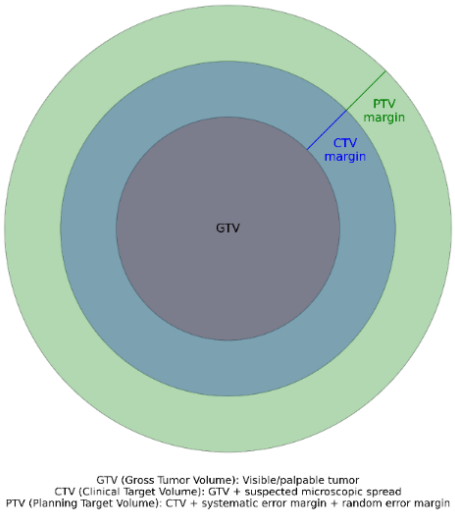


Figure 11 : Illustration of the different target volumes.

6.1.2. Robust planning

The core tenet of robust optimization is the explicit incorporation of uncertainties into the treatment planning process [19, 86]. Unlike conventional methods that optimize for a single nominal scenario, robust optimization considers a range of possible error scenarios, ensuring that the treatment plan remains clinically acceptable across various potential outcomes. Equation (2) can be slightly more precisely defined as follow:

$$\min_{x \in X} \sum_{i=1}^{n} w_i f_i \left(d(x)\right) \tag{4}$$

With *n* functions *f* dependent of the dose distribution *d* weighted by non-negative importance parameters *w* and *X* the set of possible machine variables, e.g. MLC positions and segment weight for IMRT or spot weight for IMPT.

It is possible to extend such optimization problem to several error scenarios. Such a problem can be treated with a minimax optimization of n_r functions required to be robust [24].

$$\min_{x \in X} \max_{s \in S} \sum_{i=1}^{n_r} w_i f_i \left(d(x, s)\right) \tag{5}$$

With *S* the set of possible scenarios, e.g. image translation in different directions or image intensity variations.

Such an approach allows to consider the dose distortion due to treatment error at the planning level [24, 85, 86]. Consequently, it is very often used in PT to ensure the target coverage or maximum dose to critical OAR e.g. the spinal cord. Nonetheless, such techniques can also be convenient in RT [87, 88], for example, to deal with large anatomical changes, e.g. breast treatments where significant anatomical deformations could occur.

It is however important to notice that this necessitates to evaluate the objective function over the different error scenarios, thus, to compute dose for these (even if some approximations can sometimes be used to reduce the number of dose evaluations). Consequently, robust optimization often leads to prolonged optimization time when compared to regular optimization (especially if the dose is calculated with a MC dose engine).

6.2. Mitigating uncertainties in clinical routine

While treatment errors can be accommodated with treatment margins and robust optimization, it must be noted that such methods lead to a reduction in target conformity. Indeed, to accommodate these errors, margin and robust optimization lead to extra-dose to healthy tissue around the target. Consequently, the larger the error one wants to deal with, the lesser the target conformity, thus

the larger the extra-dose received by healthy tissue. Therefore, it is necessary to choose reasonable treatment uncertainties when using such approaches. This choice of reasonable uncertainty can be eased by the use of treatment uncertainties reduction techniques. This section provides a description of a few techniques used in RT and PT.

A first way of reducing setup errors is to provide an easily reproducible patient position between the simulation step and the different days of treatment. To do so, immobilization devices can be used. These devices can take very different forms. For example, it can be a thermoplastic mask forcing the head and upper part of body to be in a similar position or a vacuum cushion that can be fixed on the initial treatment couch or even arms or foot holder. There exist numerous solutions sharing this similar goal.

The Image Guided Radiation Therapy (IGRT) is another technique that allows to reproduce at best the simulation setup of the patient on the treatment couch [89]. The concept lays on the acquisition of an image just before the start of the treatment. This imaging is typically performed using on-board imaging systems such as cone-beam computed tomography (CBCT), planar kV X-rays, or even integrated MRI. Using this image, clinicians can operate translational shifts of the treatment couch. These shifts, applied in three dimensions (lateral, longitudinal, and vertical), aim to align the current patient position with the planned treatment position. This process effectively compensates for setup errors, which are inevitable due to the challenges of reproducing exact patient positioning day after day. The capabilities of modern IGRT extend beyond simple translational corrections. Advanced systems can also account for rotational errors, applying pitch, roll, and yaw corrections to further refine patient positioning.

Another technique is the Surface Guided Radiation Therapy (SGRT). SGRT follows a similar goal than IGRT but rather than using a volumetric image it uses a surface one. It uses a combination of a projector and several high-resolution cameras to register a real-time 3D surface [90]. SGRT aids in the initial positioning of the patient on the treatment couch. By comparing the current surface model to the reference model, the system can guide clinicians in making precise adjustments to achieve the planned treatment position. It offers a radiation-free alternative to traditional IGRT methods. However, it only allows to account for surface change

since it is not a volumetric image. The main advantage of such technology is that the acquisition is constant and real-time. It allows thus to detect intra-fraction surface changes. This allows to pause the treatment when large discrepancies between the initial surface image and the current surface image are observed.

To account for breathing movement, there also exist techniques such as audio/video coaching or mechanically assisted ventilation that enable a reproducible breathing pattern during treatment [91-94].

7. The appeal of plan adaptation

7.1. The different forms

While treatment uncertainty reduction techniques mitigate some potential treatment errors, they provide incomplete compensation for the full range of uncertainties encountered during RT. A more comprehensive approach involves modifying the treatment plan based on the patient's current anatomical configuration, a process known as adaptive radiotherapy (ART). Three distinct categories of ART have been established in the literature: offline adaptation, online adaptation, and real-time adaptation. Their operating timescales are depicted in Figure 12.

In offline adaptation, the treatment plan is adapted by considering a more recent patient image than the planning one. Specifically, adaptation is operated while the patient is not on the treatment couch and not necessarily with an image obtained the same day than the following treatment fraction. Consequently, there is a usual gap of a few hours or days between offline adaptation and delivery of adapted fraction, during which anatomical changes can still occur.

In online-adaptive radiation therapy (OART) the treatment is adapted with an image of the patient already on the treatment couch, i.e. this is an image-of-the-day acquired a few minutes before the delivery of the adapted treatment fraction. While much closer to the current patient anatomy than offline adaptation, OART adds an extra timing constraint, since the patient is waiting on the treatment couch during the adaptation process.

In real-time adaptation, the goal is to consider any intra-fraction errors that could occur and almost instantly adapt to it. While full real-time adaptation is not yet feasible, intermediate concepts such as gating and tracking are already clinically

used to mitigate intra-fraction movement, e.g. target and OAR displacement due to breathing. Gating consists in pausing the treatment when discrepancies with the reference patient set-up is observed and continuing when current patient setup is deemed to be back in the reference conditions [95, 96]. Continuous patient monitoring for gating can be done with SGRT for example. Tracking consists in following in real-time the target displacement to ensure correct treatment delivery to the tumor volume [97-99]. This can be done with SGRT continuous monitoring or with fluoroscopy. It should be noted that coaching and mechanically assisted ventilation techniques discussed previously are also sometimes considered as real-time adaptation techniques.

Note that these adaptive approaches are designed to complement rather than replace the previously cited uncertainty mitigation techniques, working together to minimize treatment uncertainties and enhance target dose conformity.

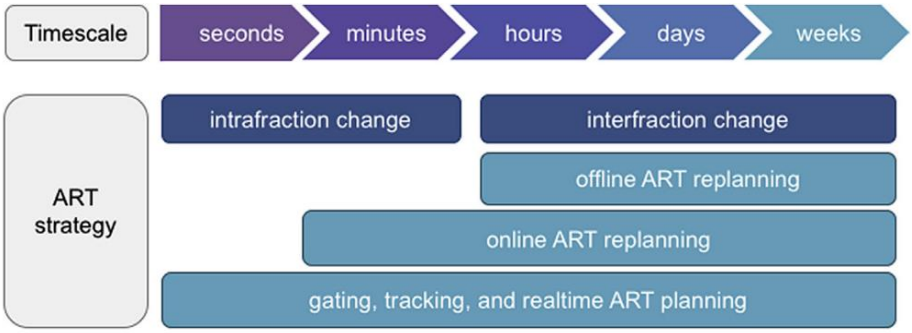


Figure 12 : Illustration of the different adaptive techniques according to the time scale. Adapted from [7].

7.2. The current challenge: online adaptation

Among these different approaches, OART currently generates the greatest interest due to its potential benefits for treatment outcomes, since this promises to be a more powerful approach than offline adaptation and since full real-time adaptation does not yet seem technically feasible.

As previously introduced, the implementation of OA allows to reduce the magnitude of treatment uncertainties (both systematic and random) compared to conventional non-adaptive approaches, where treatment plans are designed several days in advance of delivery. This reduction should occur through two

primary mechanisms. First, adaptation based on daily imaging naturally minimizes geometric uncertainties (uncertainties in target and organ positioning) by accounting for the patient's current anatomical state. Second, the adaptive process effectively transforms certain systematic errors into random ones, as exemplified by image segmentation uncertainties, which vary between fractions under OA conditions. This conversion is particularly significant since random errors typically tend to average out over the treatment durations, while systematic errors affect all fractions uniformly, maintaining their full magnitude throughout the treatment course. Note that other types of errors, for instance treatment machine calibration, or errors induced by cardiac motion that cannot be reduced with OA.

When considering these combined effects, it is then possible to reduce the PTV margin and/or values of robustness parameters to ultimately decrease the total irradiated volume. This reduction in treated volume directly translates to lower radiation exposure to healthy tissues, potentially leading to decreased normal tissue toxicity and improved therapeutic outcomes [100, 101]. Such decrease in normal tissue toxicity can in turn justify longer adaptation time for PT-based techniques, as long as the overall NTCP for a patient cohort remains favorable despite extended treatment times [35].

Nonetheless, as presented in the workflow section, the different steps to produce a high-quality plan become extremely tricky in a timeframe where clinicians only dispose of few minutes to treat a patient:

- First, OART requires an image of the day of sufficient quality to allow to precisely segment anatomical volume and an accurate spatial representation of the electron density in order to compute the dose distribution. One possibility is to have access to an in-room CT (a CT scanner in the same room as the treatment couch). However, this requires having a disposal a treatment bunker of a consequent size. Moreover, this represents a significant additional cost in comparison to classical in-room imaging such as CBCT imagers as well as a more consequent workload.
- Second, the anatomical volumes segmentation cannot be done manually as such step represents several hours of work. Since manual

segmentation is still considered as a gold standard in RT, any automation to speed such process have to be of a highly convincing quality, so that review and corrections of such automatically segmented contours would only last a few minutes.

- Third, the planning process is also another highly time-consuming step. Again, the solution is to automatize at most the process or at least initialize treatment parameters to value close to optimal ones.
- Fourth, one needs at least a minimal assurance of treatment quality to be sure that what has been achieved with in silico computation is relevant when the moment of delivering such treatment comes.
- Fifth, since the treatment image, the contours, the optimization constraints will be updated all along the treatment, it is necessary to have a way of accumulating the dose on a reference image. Such accumulation should clearly indicate how the different daily contributions sum all together in order to assess the quality of the treatment. Nonetheless, the main method of doing this is to use the registration process between the initial image and the different images that served as adaptation image. Such registration algorithms present limitations, especially when large anatomical changes occur between the two images.

The adaptation problem is even more preponderant for PT due to the inherent sensitivity of the technique to proton range errors. Indeed, such an inherent advantage of PT is conversely its main constraint: any errors in the estimated position of the target or in the quantification of the necessary protons range can lead to a severe drop in treatment quality.

Moreover, PT presents additional challenges for adaptation due to the limitations of analytical dose calculation algorithms, often necessitating the use of slower MC engines. While conventional non-adaptive workflows can accommodate optimization times of several hours, adaptive workflows must complete dose optimization within minutes while maintaining clinical quality standards. The process also requires adequate image quality for both anatomical segmentation and PT dose computation, which can be particularly challenging in OA settings. Consequently, certain assumptions and practical compromises must be implemented to produce adapted plans within timeframes compatible with online adaptation.

8. Artificial intelligence: a strong ally for online adaptation

Artificial intelligence (AI) refers to computer systems designed to perform tasks that typically require human intelligence. These tasks include learning, problem-solving, pattern recognition, decision-making, and language understanding. AI systems use algorithms and large datasets to analyze information, make predictions, and adapt to new inputs. Most often AI in RT refers to Machine learning (ML) and deep-learning (DL).

ML usually refers to methods where the model performance relies on the extraction and engineering of features (the measurable properties or characteristics) from the training dataset that are then used to feed the model. Those features for non-DL models are usually human-interpretable and lowly abstracted, e.g. the volume of the tumor, the patient weight, the distance between tumor and nearest OAR, etc.

DL, on the other hand, refers to methods designed to automatically decompose training data into sub-components, effectively creating its own features. These features are generally less interpretable to humans and more abstract, being created through multiple layers of processing within the model. The more layers, i.e. the deeper the model, the more abstract the features become. This capability to efficiently decompose initial data into effective features is the primary strength of DL models. Another key distinction is that traditional ML often requires more human intervention in feature selection and engineering, while DL can learn relevant features automatically. This can be advantageous in complex domains where optimal features are not obvious to human experts. Consequently, DL is often considered as being a subsection of ML.

Given this capability, DL models are often assumed to have higher performance potential than traditional ML models for similar tasks. However, this assumption holds true primarily when sufficient data is available for training. The high data requirements represent one of the main challenges of DL compared to traditional ML approaches.

Looking at the broader perspective, AI aims to develop models that can emulate or surpass human cognitive capabilities in specific tasks. Consequently, AI has garnered significant attention in RT applications over the past decade. Its ability to

combine speed and accuracy has made it increasingly indispensable for state-of-the-art adaptive RT approaches.

The following sub-sections describe the common models that can be found in radiotherapy as well as their applications.

8.1. Common models behind the applications

8.1.1. Supervised deep-learning models

Different deep neural networks. Deep neural networks in radiotherapy encompass several architectures, each suited for specific tasks. Convolutional Neural Networks (CNNs) are widely used for image analysis, excelling in tasks like tumor segmentation and organ-at-risk delineation (see next section). They process 2D and 3D medical images by applying convolutional filters to extract relevant features hierarchically.

U-Net, a specialized CNN architecture, has become particularly prominent in RT [102]. Its U-shaped structure, combining a contracting path and an expanding path with skip connections, allows for precise pixel-wise segmentation of medical images. U-Net is extensively used for diverse applications, offering high accuracy even with limited training data. It has created different U-Net-like structure such as the Hierarchically-densely connected U-Net (HD-Unet) [103] , illustrated in Figure 13, or the “no-new-U-Net” (NN-Unet) [104].

nnU-Net, short for "no-new-U-Net," is an advanced implementation of the U-Net architecture designed specifically for medical image segmentation tasks. It is a self-configuring system that automatically adapts its architecture and training process to the specifics of a given dataset, requiring minimal manual intervention. This versatility allows nnU-Net to handle various medical imaging modalities like CT, MRI, and PET, as well as different data dimensions including 2D, 3D, and even 4D [105-107]. The framework incorporates automated preprocessing steps tailored to medical imaging data, such as resampling, normalization, and handling of missing values. nnU-Net typically employs an ensemble approach, training multiple models with different configurations and combining their predictions for improved performance. In the field of radiotherapy, nnU-Net has demonstrated excellent results in tasks such as tumor segmentation, organ-at-risk delineation, and treatment planning assistance. It's often used as a strong baseline for comparing new segmentation algorithms in medical imaging research.

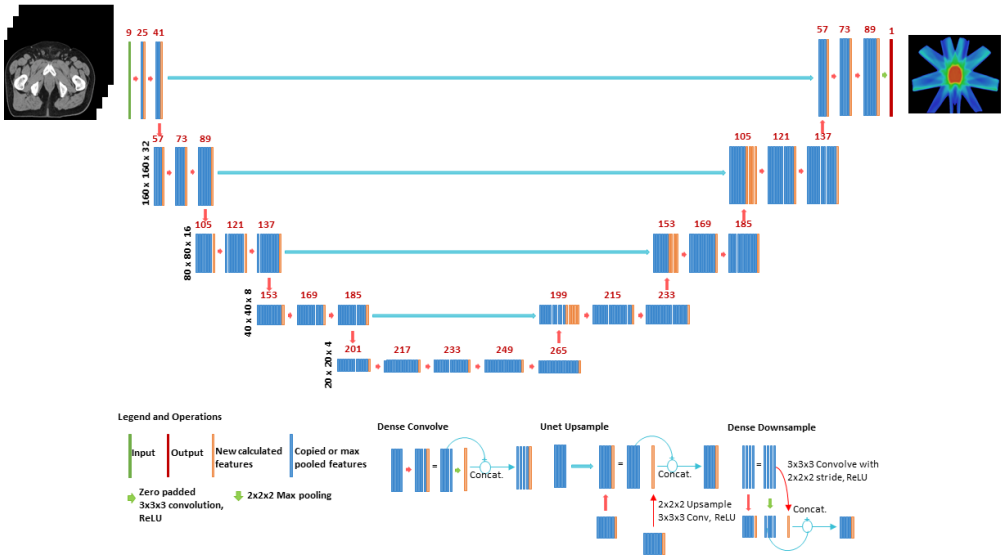


Figure 13 : example of a U-net derived structure, the HD-Unet. In this example, a CT and different 3D mask of OAR and target are used as input. The output is a predicted 3D dose distribution.

Generative Adversarial Networks (GANs) emerged as a promising approach in radiotherapy applications, particularly for synthetic data generation and image enhancement tasks [108]. With GAN, two neural networks (a generator and a discriminator) compete against each other as illustrated in Figure 14. The generator has to produce an output close to the data in the training set while the discriminator has to be able to discriminate which data is generated and which one is actually really coming from the training dataset. However, despite several years of active research and development, they have faced persistent challenges: training instability (mode collapse) [109, 110], difficulty in achieving convergence, and sometimes producing artifacts in medical images [111] that could be problematic for clinical applications. Moreover, their adoption has been further challenged by the recent emergence of diffusion models, which have demonstrated great performance and stability across various imaging tasks.

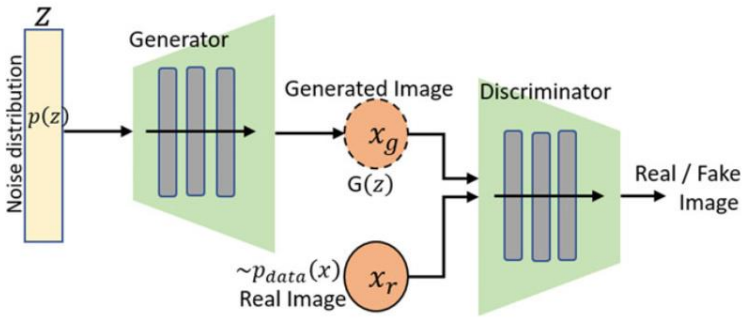


Figure 14 : original GAN model where a generator model tries to trick a discriminator model. Image from [112].

Since a few years the diffusion models achieved outstanding results in diverse research fields [110]. Diffusion models have begun to find applications in radiation therapy [113-115], offering promising new approaches to various imaging and planning challenges. At their core, diffusion models operate on a fundamental principle: they learn to reverse a gradual noise-addition process, progressively reconstructing high-quality images from noisy inputs through a series of denoising steps. This iterative denoising process allows the models to capture complex data distributions and generate highly realistic outputs. The step-by-step nature of diffusion models also makes them more interpretable than GANs, which is particularly valuable in medical applications where understanding the model's decision-making process is crucial. In radiotherapy specifically, these models show potential for various applications such as synthetic image generation, image enhancement, and anatomical structure prediction.

8.1.2. Reinforcement learning

Since a few years, reinforcement learning has also been used in RT. Reinforcement learning is a specific kind of ML algorithm where an agent is trained to make an optimal action on a given environment. The agent has to be rewarded according the state of the environment after having decided on an action that could maximize the total final reward. Consequently, the agent has to find a balance between the exploration of a new set of actions or exploiting well-known actions that provide local maximized rewards (with its current knowledge). The interest is that since such models are not part of the supervised model (they are not trained on a labelled pair of input-output data) they can learn optimal strategies that are sometimes unsuspected by humans. This raises interest in

automatizing some of the highly manual steps in RT where such agent could simply learn the manual human pattern and potentially even enhance it, e.g. for treatment planning [116].

8.2. Applications

8.2.1. Automatic segmentation

A first common application of AI is the automatic segmentation of medical images. The main input of such models is the image itself. Modern AI models can produce accurate 3D segmentation of diverse OARs and even target volumes. The implementation of AI-driven image segmentation in radiotherapy offers several advantages, including enhanced efficiency in treatment planning, reduced inter-observer variability [117, 118], and the potential for identifying subtle imaging features that may elude human perception. However, challenges persist in ensuring robust performance across diverse patient anatomies and in handling atypical or rare cases. Nowadays, numerous commercial companies now provide AI-based auto-segmentation across various image modalities e.g. CT, CBCT, MR…. Example of commercial providers are Raysearch, Varian, Siemens Healthineers Syngo. Via, MIM Contour Protégé or Therapanacea ART-plan.

8.2.2. Automatic plan optimization

Another popular application is the automation of the planning process. There exist several approaches that enable fast and automatic plan optimization in RT. Most of these solutions aim at predicting information or parameters of the treatment plans based on the patient image and contours.

A first approach is the prediction of the achievable dose-volume constraints, i.e. the model learned on previous high-quality plans what were the dependencies between the used optimization constraints and the patient anatomy. With such information, the human planner can initialize the optimization process with relevant constraints that are close to the optimally achievable constraints. This allows to save time and reduce the inter-observer variation [119].

A similar idea is to directly predict the 3D dose distribution. Here, the model outputs a 3D matrix containing the spatial dose distribution, i.e. the model learned the dependencies between spatial dose distribution and patient anatomy. For such a model, common inputs are the patient image and the patient contours (OARs and target) [120, 121]. Consequently, it provides more information than

the previous prediction of dose-volume constraints. To make best use of such information a common process following the dose prediction is “dose-mimicking”. Such a second step allows to convert this 3D matrix into a real deliverable plan [122, 123]. It generally consists in minimizing the voxel difference between what has been predicted and what achieves the optimizer at each iteration of the optimization process. It can also consist in optimizing maximum and minimum dose on isodose contours derived from the predicted dose. While powerful, it is nonetheless necessary to pay attention to the beam configuration on which the model has been trained, i.e. if a model was exclusively trained on 9 fields IMRT plans it will only be able to output predictions for such beam configuration.

Another approach of automatizing the planning consists in predicting the treatment parameters, such as the MLC position the necessary MU and/or the optimal beam configuration. This provides an initialization for the optimization process and can save a significant time [116, 124, 125].

Finally, some groups also proposed models aiming at reproducing the human-tuning of optimization constraints during the optimization of the dose [126, 127]. The input of such models is generally the state of the dose-volume constraints at a given optimization iteration, the output is a set of actions on optimization constraints, e.g. increasing/decreasing optimization weights on specific constraints. Consequently, the model automatically optimizes the treatment plan through several decisions it progressively makes for the values of the optimization constraints and their weights.

Automatic plan optimization has gained widespread adoption in RT, with vendors offering various solutions. For instance, Varian RapidPlan for DVH prediction and Raysearch dose prediction models.

8.2.3. Synthetic data generation

A third application is the generation of synthetic data. Synthetic data in RT refers to artificial data that mimic real patients. The model main input is generally an image of a given modality, e.g. CBCT, MR, the model common output is an image for the same patient but predicted for another modality e.g. CT [128].

A first application of such inter-modality image translation is usually used to enable the recovering of electronic density necessary for the dose computation

step when working with MR or CBCT images for example. Such an application has already been proposed by different vendors such as MR-box of Therapanacea, syngo.via RT from Siemens or MRCAT from Philips.

Note that if there exist for a same patient two images of different modalities then registration methods already perform well at matching images. When applying the displacement field to the moving image, the resulting image might sometimes be considered as synthetic data. In this context, some groups started to investigate the prediction of displacement field between two images with DL models, mostly to speed up such a process that can sometimes take several minutes [129-132].

Data augmentation represents another valuable application of this method. Deep learning approaches typically require extensive datasets for optimal performance. In situations with limited data availability, generating synthetic data can artificially expand training sets to improve prediction quality [133]. However, this application remains nascent and has not yet been integrated into commercial systems.

While sCT has great interest in inter-modality translation, it can also be used to enhance the quality of the input image [134-136]. Consequently, the output modality is the same as the input one. This has been proposed specially to enhance CBCT quality such that anatomy could be better discerned and/or artifact could be corrected. For instance, Elekta has recently announced to soon provide commercial software for such purpose.

8.2.4. Anatomical changes and motion prediction

Predicting anatomical variations and organ motion has become an increasingly important focus in adaptive radiotherapy. Several research groups have developed AI-based approaches to anticipate anatomical changes during treatment courses. These predictive models serve multiple purposes: they can forecast the necessity of replanning by analyzing early anatomical variations and their likely progression [137, 138], identify optimal timing for treatment adaptation by predicting the most significant anatomical changes, and even enable near real-time motion prediction as demonstrated in the CyberKnife system. Such predictions are particularly valuable for tumors subject to respiratory motion or located in anatomical regions prone to significant interfractional variations. Early identification of patients likely to require

adaptation can improve resource allocation in clinical workflows while ensuring timely intervention when necessary. Moreover, the ability to predict tumor motion in real-time, as implemented in the CyberKnife system, enables more precise beam delivery by anticipating target position, thereby potentially reducing treatment margins and improving healthy tissue sparing.

8.2.5. Outcome Prediction

AI was also proposed as a tool to predict the toxicity of RT for various treatment locations, e.g. prostate, breast, head and neck [139, 140]. It has also been proposed as a predictor of treatment failure or response. It is also interesting to note that a similar objective can be reached through the prediction of a dose (see automatic planning section above) and the computation of normal tissue complications probability with it.

Despite its significant potential, this application has not yet been widely adopted or integrated into commercial systems.

9. State of the art of online adaptation for photon therapy

The implementation of AI tools in OART offers promising advancements. However, to ensure time efficiency, it is crucial to integrate the different automation tools within a comprehensive workflow in a unified treatment environment. This necessity often leads to the adoption of all-in-one commercial solutions that provide both the requisite hardware and software in an efficient ecosystem.

Over recent years, commercial vendors and research groups have demonstrated significant interest in addressing the challenges of OART, resulting in the development of various solutions. In conventional radiotherapy, three primary commercial solutions currently dominate the market, with two additional systems on the horizon. These include the Ethos (Varian, Palo Alto, US), the Unity (Elekta, Stockholm, Sweden), and the MRidian (ViewRay). While these systems employ different technologies, they follow similar adaptive workflows.

The general adaptive workflow begins with the generation of an initial treatment plan based on a planning CT, performed before the first treatment fraction. Subsequently, the adaptive process typically encompasses the following steps: patient setup, volumetric image acquisition, image segmentation, plan

optimization based on the initial plan, treatment quality assurance and treatment delivery.

It is worth noting that additional steps may be incorporated depending on the imaging modality and the anatomical location of the treatment. For instance, systems equipped with MRI capabilities may perform real-time image acquisition during treatment delivery, allowing for enhanced monitoring and potential intervention. Details about these different commercial approaches are described hereafter.

9.1. The Ethos

The Ethos therapy system, introduced by Varian Medical Systems in 2019, is an enhanced version of their Halcyon linear accelerator, augmented with comprehensive automation software to enable rapid adaptive workflows. This system aims to streamline the treatment adaptation based on daily anatomical changes in patients undergoing RT in less than 20 minutes. Its main features are an automatic planner, the Ethos Intelligent Engine (IOE), an AI-based automatic segmentation algorithm (for pelvis and abdomen) and kV iterative CBCT (iCBCT) [27].

A distinctive feature of this system is its ability to perform OA using the daily CBCT. This approach offers both advantages and limitations. While CBCT is a widely available and familiar on-couch imaging modality, it provides significantly lower image resolution and contrast compared to conventional CT imaging. To overcome these limitations, the Ethos system employs various approximations and technical solutions to efficiently enable CBCT-based OART.

The Ethos workflow begins with initial planning, which follows a similar process to classical simulation CT-based planning but utilizes the IOE for automatic planning. The IOE operates with a template of dose volume-constraints that users aim to achieve. It sequentially optimizes the different objectives of the inputted list of goals, starting from the most important ones and ending with the least important. Moreover, it can automatically add objectives to control the dose gradient around the target and the dose homogeneity inside the target. The photon optimizer algorithm proposes then IMRT and VMAT plans to the user.

Then, each day of treatment a CBCT is acquired and used as surrogate for the anatomy of the day. Contours are then automatically segmented on this CBCT. The most important OARs are segmented with AI, those AI-segmented OAR are named the “influencers”. The target is then propagated using a guided non-rigid registration based on the initial position of the influencers and target and their position of the day. Other remaining contours are simply registered on from the initial CT to the CBCT of the day. The clinicians can then check the different contours and correct them if necessary. Once the contours have been approved, an adapted plan is generated through a new launch of the IOE with the exact same list of clinical goals inputted at initial planning. The dose calculation is done on a synthetic CT (sCT) that consists in the initial CT non-rigidly registered onto the CBCT (illustrated in Figure 15) with the Acuros analytical dose engine. The physician can then choose between the initial plan or the adapted plan to treat the patient.

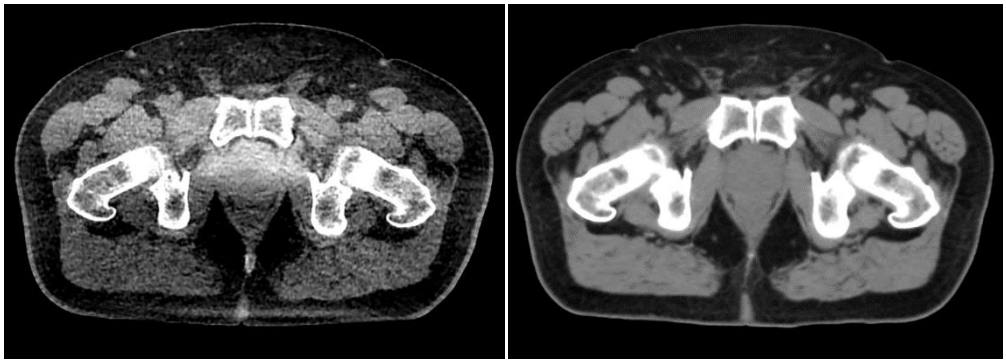


Figure 15 : Comparison of image quality in pelvic imaging. Left: conventional Ethos CBCT; Right: sCT generated to match the CBCT anatomy. The sCT is used for the dose computation to overcome the limited image quality of conventional CBCT.

The use of sCT is necessitated by the limited quality of CBCT entailing the direct dose computation on it. However, there is currently no real evidence to guarantee that daily generated sCTs remain valid throughout the entire treatment course for the complete patient population. Furthermore, CBCT is particularly susceptible to image artifacts, which can further complicate the generation of clinically usable sCT, as illustrated in Figure 16.

Finally, a second CBCT is usually acquired before delivering the treatment. This allows to correct for potential patient drift that could have occurred during the

adaptive workflow. In parallel to this second image, a second dose computation is launched as dose QA. This second dose is computed on the Mobius platform [141]. It uses a CCC algorithm and the log files of the produced plan to propose a second calculation of the dose. While recent study suggested that Mobius could serve as a reliable indicator of plan failure [142], another one suggested that it could not be necessary for every treatment fraction verification [143].

This workflow is generally reported to be completed within a 20-minute time slot [37, 144-148]. However, it also exists attempts to treat complicated cases such as head and neck simultaneous integrated boost where workflow average duration can extend to more than 30 minutes mostly due to prolonged manual corrections of the numerous registered anatomical volumes [149].

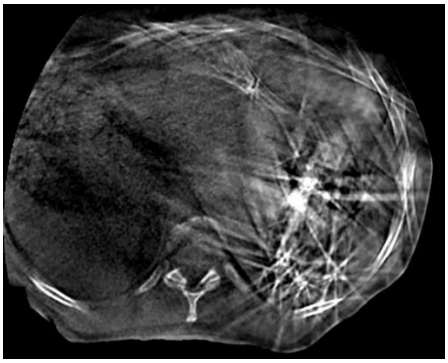


Figure 16 Example of a thoracic CBCT image, notice the low soft tissue contrast combined with important artefacts that strongly degrades the interpretability of the image.

Over recent years, studies showed various benefits of daily CBCT-adapting treatment plans such as improving target dose coverage and reducing dose in OARs [37, 144-146, 150-152].

While already performing quite well, the recent released of a higher resolution CBCT, "*HyperSight*" for the Ethos promises easier and more efficient adaptation due to its enhanced image quality [153-155]. Such new imager provides an enhanced image quality, easing the interpretation of the image for the clinicians [156] and enabling a direct computation of the dose on the image of the day without having to use any intermediary sCT [157]. This provides much more reliable information for the dose calculation than the sCT that is sensitive to large anatomical changes between planning image and image-of-the-day [158]. While it

could also be supposed that the accuracy of AI-based contour segmentation could be increased, the first investigation on prostate cancer patients did not show any clear improvements [159]. However, at the moment, the AI-model has not been updated and remains fully trained on conventional iCBCT. It is not yet clear if it would be re-trained one day with *HyperSight* CBCTs and if this will increase the accuracy of AI-segmentation.

It is noteworthy that such an automated adaptation workflow, with its capability for direct dose computation on daily images, could enable simulation-free treatments, eliminating the need for conventional planning steps prior to the first treatment day [160].

9.2. The Unity

Introduced in 2018, the Unity (Elekta AB, Stockholm) combines a linac with a 1.5T MRI. The unity proposes an adaptive workflow MR-based. As previously explained, the incorporation of MRI technology offers superior soft-tissue contrast compared to conventional computed tomography (CT) or cone-beam CT (CBCT), while also offering a non-ionizing imaging approach. This also enables image acquisition during beam delivery, allowing for intra-fractional motion monitoring. This capability provides the potential for beam interruption in cases of significant anatomical displacement, adding an additional layer of precision and safety to treatments [28].

The Unity comes with several different adaptation workflows where the MLC positions and treatment segment weights can be updated or kept as on the original planning. Complete re-optimization of treatment plans based on daily anatomical variations is also possible. These adaptive workflows are integrated in a special version of Monaco TPS (Elekta AB, Stockholm, Sweden).

As soon as a volumetric image of the patient has been acquired, the anatomical segmentation volume begins. Here such segmentation consists in either rigidly-registered or non-rigidly registered contours from initial CT to the one MR scan. While for rigidly registered-contours no corrections are authorized (it is part of a simplified workflow) for non-rigidly registered contours this often requires consequent human reviewing and corrections. Indeed, unlike for the Ethos no AI-based auto-segmenter is available for MR image segmentation inside the commercial software environment. However, research groups obtained

interesting result with in-house deep-learning model [161-163], and it can also be mentioned that commercial software such as ARTplan from Therapanacea (Paris, France) already proposed AI segmentation on MR images.

Concerning the planning, the optimization process takes also place in Monaco TPS where template-based optimization can be used. However, it seems that such an optimizer still needs human-intervention to achieve satisfying results [38]. Research have investigated the use of different approaches such as Swarm Particle optimization, Machine-learning based optimizers [38] or also mCycle [164], the Erasmus iCycle optimizer [165] (Erasmus University, Rotterdam, Netherlands) implemented in Monaco TPS and showed better dosimetric quality while unfortunately often being slower. Also note that for the moment, VMAT plans cannot be generated. Moreover, the electron density for dose computation is either assigned through a registration process for basic adaptation without recontouring (similar than for the Ethos) or through bulk density assignment for full re-optimization workflow (illustrated in Figure 17). For the second case the inhomogeneities inside the different tissues are thus lost while it is more robust for large anatomical changes that could produce inaccurate sCT. However, such a bulk density assignment naturally also results in some dose uncertainty.

It is also interesting to note that the presence of such a large magnetic field introduces complexities in dose distribution due to the Lorentz force acting on secondary electrons. To address an accurate modeling of the dose distribution, MC engine dedicated to the Unity has to be used.

Finally, before treatment delivery the acquisition of a new volumetric image can be performed as on the Ethos to compensate for patient shift.

Concerning online quality assurance, performance of a secondary dose calculation with a dedicated software considering the effect of transverse magnetic field is advised, e.g. a modified Clarkson algorithm with correction for transverse magnetic field [166-168].

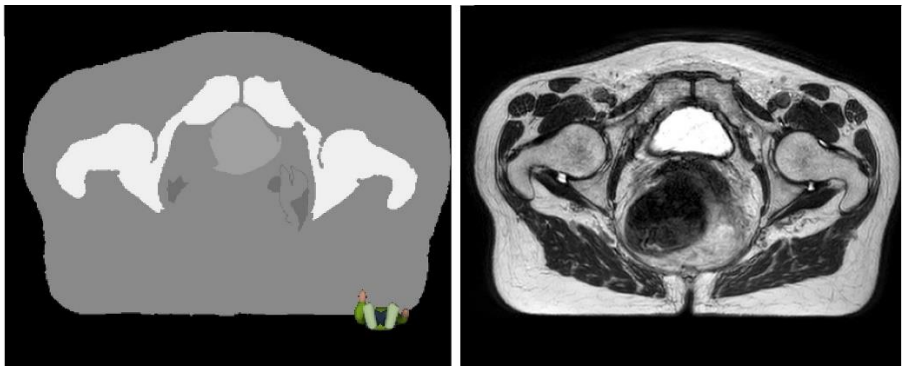


Figure 17 : Illustration of the 1.5T MR image quality and bulk density assignment for a pelvic case. Left: pseudo-CT created by bulk density assignment on MR image; Right: 1.5T MR image. Adapted from [169].

Ongoing research and development efforts focus on enhancing the efficiency of the adaptive process to minimize overall treatment durations. Current adaptive sessions typically range from 30 to 60 minutes, encompassing patient set-up, imaging, re-planning, and quality assurance procedures. It is however strongly dependent on the kind of re-optimization performed and the anatomical location. For example, for prostate bed a research group reported a mean duration of 20.6 minutes for patients where the adaptation consisted only in a virtual couch shift without contour adaptation, while it was of 31.3 minutes for a full re-optimization adaptation [170]. For prostate, another group reported a mean duration of 38 minutes but with noticeable timing differences from months to months of clinical usage [171]. For liver, a mean duration of 68.2 minutes was reported for full plans re-optimization [40] and a custom imaging workflow.

While enabling treatment with high image accuracy as well as precise gating with cine imaging, the Unity showed its ability to ensure target coverage and reduce dose to OAR through its complete re-optimization workflow [28, 172-174].

9.3. The MRIdian

The MRIdian (ViewRay Inc, Denver) is a radiotherapy treatment system incorporating a 0.35T MRI scanner (whose image quality is illustrated in Figure 18). It was introduced first in 2014 with a ^{60}Co source as photon emitter upgraded then with a 6MV linac in 2018 [29].

The workflow shares several similarities with the Unity. Once the volumetric MR image has been acquired, the contours are deformably registered from the initial CT to the MR image of the day. Review and corrections are then manually operated. Then the plan can be adapted following different workflows that either re-optimize the segment-weight, the fluence or re-launch a full new re-optimization with initial or updated objectives. Only conformal and IMRT plans are proposed by the optimizer. Before starting the treatment delivery, research groups proposed a QA step with a second independent MC dose calculation and a gamma analysis tool to ensure the validity of the computed dose [175, 176]. To prevent potential shifts between the start of the adaptation process and the moment of treatment delivery, a second image can be acquired to ensure correctness of isocenter position. However, for the MRIdian only 2D cine imaging seems to be implemented for such purpose [177].

The system also disposes of the capability of using such 2D cine imaging while treating to shut-off the beam when necessary if anatomical movement are observed, i.e. to perform gating during treatment [29].

While the magnetic field of such MR-imager being much lower than the one from the Unity, the dose computation is also done through a MC engine taking its influence into account. Also, as for the unity, it is possible to compute the dose on the registered CT or on the MRI with bulk electron density assignment.

As the Unity, the MRIdian showed its ability to preserve target coverage through the different treatment fractions while reducing dose to OARs [178-180].

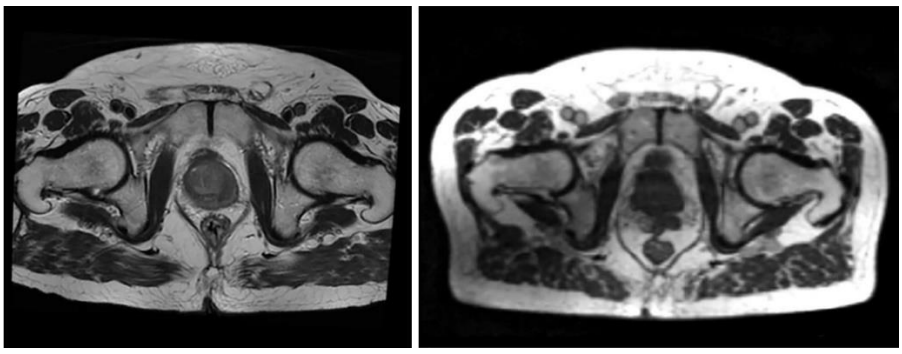


Figure 18: Illustration for a pelvic case of MR image quality as a function of the intensity of the magnetic field. Left: Unity 1.5T imager; Right: Mridian 0.35T imager. Adapted from [181].

As for the Unity, the adaptation workflow can thus last quite a time due to time-consuming contours review, lowly automatized optimization of planning and potential supplementary imaging protocols. The reported adaptation time is also quite dependent on the anatomical treatment sites, but most studies report a median time above 45 minutes [176, 182, 183].

Nonetheless, as for the Unity it is important to note that such timing values result from specific imaging workflows and specific treatment monitoring that are enabled by the MR image quality and would not be possible with CBCT. Consequently, one must be careful when making timing comparisons with the timing ranges achieved by the Ethos since the treatment indications are not necessarily the same.

9.4. The Aurora-RT

The Aurora-RT (MagnetTx Oncology Solutions, Edmonton, Alberta, Canada) is another linac incorporating an MR imager which has been presented in 2024 [184]. Aurora MR images are acquired with a 0.5T magnetic field. It differentiates from the other MR-linac by providing biplanar irradiation field, either parallel or perpendicular to the static magnetic field, which can lead to some differences in electrons transportation and thus in the spatial shifting of the dose. It also enables treatment couch motion to ease the IGRT with regards to previously presented MR-linac.

While the Aurora-RT has received both FDA and CE approval, it is important to note that the system is still in the early stages of clinical implementation. Future research and clinical studies will be crucial in elucidating the specific benefits and potential advantages of this device in various treatment scenarios.

It is particularly noteworthy that the Aurora-RT system appears to be placing significant emphasis on AI integration. The ongoing development of AI applications for this platform is expected to encompass several key areas: AI-based MR reconstruction; AI-based intra-fraction cine-imaging; AI-based tumor segmentation and target movement prediction [184].

9.5. The Evo

Recently, Elekta has also announced to provide another OART-compatible linac equipped this time with a CBCT imager: the Evo. Evo CBCT image quality is

announced to be enhanced by AI. It is also announced that such linac should be compatible with a new planning software, the Elekta One Planning, where segmentation and dose planning would be automatized by AI.

However, CE mark approval has only recently been received. It is therefore likely that further details of this treatment device and its clinical benefits will only become available in a few months' time.

10. State of the art of online adaptation for proton therapy

Currently, there exists no commercial implementation of OA for PT. PT presents unique challenges that have hindered the development of commercial implementations. Several factors contribute to this complexity:

1) Dose calculation precision: Monte Carlo (MC) simulations remain the clinical gold standard for PT dose calculations due to their accuracy. However, the computational time required for MC simulations has traditionally been a limiting factor for online adaptive planning.
2) Imaging requirements: Accurate proton range prediction necessitates high-quality volumetric imaging that provides precise stopping power information. This requirement often exceeds the capabilities of conventional in-room imaging systems.
3) Robustness considerations: PT is particularly sensitive to anatomical changes and treatment uncertainties. Ensuring plan robustness in an adaptive setting adds another layer of complexity to the optimization process. Robust optimization is currently considered as a gold standard in PT [85, 86, 185]. However, its computational intensity often leads to a prolonged optimization time.

These technical challenges are also combined with some pragmatic aspects:

1) The absolute demand for PT is smaller than for RT, i.e. there is much less PT treatment centers than RT ones.
2) There is no company providing both the necessary hardware and software together.
3) As previously introduced, adaptation time is even more constraint than in RT due to the economic factors [33-35].

These combined challenges make OA for PT significantly more demanding than for conventional RT, necessitating innovative approaches and technological advancements.

Over the years, research groups started to use different tools and ideas that were proposed in the community [186]. Several groups started to leverage the capability of Graphical processing units (GPUs) for speeding up the MC dose calculation and goals optimization. Simple tricks such as re-using the optimization constraints from initial planning and making use of registered anatomical volume started also to be commonly used to speed-up the adaptation process. Moreover, to palliate the cost of having an in-room CT to provide reliable information for the dose computation, CBCT started to be considered with scattering correction either from analytically- or DL-based or as for the Ethos by generating a sCT from the registration of the initial CT on the CBCT of the day [187].

Research groups have developed various adaptive processes for OA PT using these tools and simplifications. However, these approaches commonly face two limitations: they most often trade dose distribution optimality for shorter computation times; and like conventional OART, they do not yet have an adequate solution for cases that require extensive auto-segmentation review and correction. The following subsections will delve into the distinct methodologies proposed by different research groups, highlighting their unique features and potential advantages.

As the field of OA PT continues to evolve, it is anticipated that these research efforts will pave the way for future clinical implementations. The ongoing advancements in computational efficiency, imaging technologies, and optimization algorithms are gradually bridging the gap between the complex requirements of adaptive proton therapy and clinical practicality.

10.1. Fully new optimization with use of analytical algorithms

A first proposed approach for adapting PT plans involves generating a completely new plan based on the daily anatomy. However, as mentioned above, such an approach would typically require clinically prohibitive computation times. To address this limitation, one research group proposed using GPU-accelerated analytical algorithms for rapid proton dose calculation, enabling treatment plan generation within seconds [188]. While PT analytical algorithms still suffer of a

limited accuracy, it has been shown that impact of the dose distribution uncertainty using such algorithms can be lower than the negative impact of non-adapting the dose distribution to the anatomy of the day [189]. Using this approach, Matter et al. achieved in-silico plan adaptation within 5 to 25 seconds for paraspinal, cranio-spinal axis, brain, and paranasal cases [188].

This approach has led to the first clinical implementation of an OA workflow for daily PT plan adaptation [30]. The implementation utilizes an in-room CT-on-rails for daily imaging, with anatomical structures rigidly registered from the initial CT to the daily CT. Patient selection was limited to tumors in rigid anatomical regions to justify this rigid registration approach and to bypass the time-consuming contours manual corrections. As for previously presented photon adaptive workflows, an independent dose computation is performed as QA before treating the patient based on the generated control file from the daily plan. If the adapted plan fails to meet QA requirements or clinical quality standards, a "fallback" non-adapted plan, independently generated with potentially different beam angles or optimization parameters, can be used as an alternative.

The use of such adaptive workflow in this first investigation allowed to maintain the dose in target while ensuring reasonable dose to OARs, which was not the case for all non-adapted falloff plans. For the five patients clinically treated with this adaptive workflow, the reported necessary time for the adapted sessions ranged from 15 minutes to 30 minutes.

Future investigations should explore extending this work to cases with non-rigid anatomical changes and develop methods to replace target margins with robust optimization within clinical time constraints. This is particularly important as target margins in proton therapy have been shown inadequate for ensuring accurate target coverage under various conditions.

10.2. Dose restoration

A second proposed method consists in restoring a dose distribution similar to the initial one but on the image of the day by restoring Bragg peak position and re-optimizing the beamlet weights [190]. This process generally eases the validation of the plan since it aims at achieving a dose close to the reference. The conventional implementation begins with rigid registration of the initial dose onto

the daily image. This enables straightforward determination of dose constraints through several possible approaches [191-193]:

- A DVH-based approach, where the optimization aims to fulfill the same dose-volume constraints as the initial plan.
- A voxel-based approach, where objective functions are replaced by voxel-wise optimization functions minimizing dose differences voxel-by-voxel.
- An isodose-based approach, where isodose volumes are created from the initial dose distribution (typically in 2Gy or 5Gy steps), with minimum and maximum dose objectives set according to each isodose volume's level.

Since the initial dose cannot ensure to be an optimal reference on the new anatomy of the day, researchers have also explored methods to provide better dose reference to restore. These seek to generate reference doses that better approximate optimality for the daily anatomy, either through deep learning-based dose prediction or by applying non-rigid registration of the initial dose onto the daily image [192, 194].

10.3. Re-optimization with initial parameters

A third approach aims at re-using most of the initial plan parameters, e.g. beam configuration, spots position and weights, while fine-tuning only a subset of these parameters for the daily anatomy [195, 196]. Unlike dose restoration methods, this approach does not strictly aim to reproduce a reference dose distribution. They rather aim at quickly producing a clinically acceptable plan through the use of initialized parameters that do not necessarily need to be updated. Such approaches are thus much less computationally intensive than a complete re-planning from the ground up.

Botas et al. demonstrated the efficiency of this approach for head and neck cases through a combination of initial spot transfer to the daily anatomy, selective spot re-optimization, and GPU-accelerated Monte Carlo dose calculations, achieving average adaptation times of just over 5 minutes [195].

As for the dose restoration, initial plan optimization constraints cannot ensure to be optimal on the new anatomy of the day. However, as previously explained tremendous progress has been made in approaches predicting relevant constraint

base on the new patient anatomy [36, 119] suggesting potential for further improvement of this adaptation strategy.

Chapter 2: Getting the most out of automatic planning

This chapter is published as:

Comparison of ethos template-based planning and ai-based dose prediction: General performance, patient optimality, and limitations.
Benjamin Roberfroid, Ana M Barragán-Montero, David Dechambre, Edmond Sterpin, John A Lee, and Xavier Geets. *Physica Medica*, 116:103178, 2023.

1. Introduction

In the field of radiation therapy (RT), adaptive RT (ART) has been a major research topic for many years [197-199]. Unlike conventional radiation therapy that relies on single computed tomography scans (CT) acquired prior to treatment, ART aims at accounting for anatomical changes that can occur in the tumor (shrinkage, progression) and organs at risk (OAR) (weight loss, air cavity filling, organs deformation, …), over the course of a treatment. In practice, ART requires repeated image acquisitions to capture the daily anatomy, to delineate the updated patient volumes, and to re-optimize the dose based on the current snapshot of the patient anatomy.

Varian®(Varian Medical system, Palo Alto, USA) has recently released a new generation of linear accelerator: Ethos, which proposes a commercial implementation of Online-ART (OART) guided by cone beam computed tomography (CBCT) [27]. With CBCT iterative reconstruction, daily automatic organ segmentation, and automatic dose optimization, Ethos allows the treatment plan to be adapted within a time frame of 20 minutes [37, 144-147]. Ethos OART approach has already proven its efficiency and utility in several contexts and setups [37, 144-146, 150, 151].

The automatic plan optimization enabled in Ethos uses an "*Intelligent Optimization Engine*" (IOE), an algorithm that optimizes objectives directly coming from a list of clinical goals, ordered by importance and created by the user prior to the treatment. Any list of clinical goals can be saved as a template of objectives and constraints to be used for other patients requiring the same treatment intent. Preliminary studies have shown that Ethos generated plans quality was globally satisfying and, when compared to manually generated plans, dose metrics differences were quite small [37, 146, 147, 200, 201]. However, this general perspective does not point out specific cases where optimization difficulties can arise from more challenging anatomies: although the template-based optimization of Ethos can produce clinically acceptable plans at the patient population level, it might encounter difficulties to generate the optimal dose distribution for any given patient.

Concurrently to Ethos, various other automatic planning methods have arisen [202, 203], some of which might intrinsically better integrate patient features to

produce anatomy-specific optimal plans. Notably, the recently popular dose prediction (DP) approach, based on deep convolutional neural networks, has demonstrated the ability to predict optimal, patient-specific three-dimensional dose distribution [122, 123, 204, 205]. Following this dose prediction, a subsequent step of inverse planning, also called "dose mimicking" (DM), allows machine parameters to be determined such that the predicted dose can actually be delivered [123, 206]. Such a sequence of steps is meant to generate anatomy-specific plans, which might turn out to be more tailored to individual patients than template-based plans.

Therefore, the goal of this study is twofold. On the one hand, we aim at assessing the general performance of the Ethos solution (i.e., the template-based IOE), by analyzing the global quality of the Ethos generated plans (EG) with respect to our clinical standards. On the other hand, we aim at investigating the capability of Ethos solution to produce patient-optimal plans and compare it to the aforementioned sequence “dose prediction & mimicking” (DP+DM) workflow. These two state-of-the-art automatic treatment planning approaches are compared by looking at their performance on selected challenging plans and their capability of generating specific dose trade-off when necessary.

Notice that this study looks only at the quality of the initial treatment plans and does not analyze the quality of the per-fraction plans.

2. Materials and methods

To achieve the goals of this study, 2 planning studies using 4 different planning methods were designed. These will be presented in the following sections. Figure 19 illustrates the schematic study design and data distribution.

2.1. Data

A database of 45 prostate cancer patients, without pelvic nodal irradiation, was used in this study. All patients were previously treated on our Halcyon (Varian Medical system, Palo Alto, USA) and initially planned on the Eclipse treatment planning system (TPS). CT images were acquired on a Toshiba Acquilion CT scanner with a slice thickness of 2mm. The dose prescription was 60Gy in 20 fractions, with a CTV-to-PTV isotropic expansion of 7mm.

Local ethics committee approval was obtained for the use of all patient data under the agreement "Learning from the past: the MIRO treatment planning database" which received the ethical approval by the CEHF (Comité d'Ethique hospitalo-facultaire).

2.2. Treatment planning

For this study, new treatment plans were retrospectively generated on the planning CT using the different dose optimizations methods described in this section. All planning methods used the 9-fields IMRT configuration and isocenters from Ethos TPS. Concerning automatic methods, they were assessed without any further manual optimization: no other operations than loading the CT scans, the contours, and optimization goals were performed.

2.2.1. Ethos planning with initial template (EG_init plans)

As previously mentioned, the Ethos IOE solution generates a plan by using a set of optimization objectives that are directly coming from a template of clinical goals. Within a given template, the clinical goals have priorities ranking from 1 (most important) to 4 (less important), and internal priorities within their own group between each other. The IOE converts this list into objective functions for the photon optimizer and then monitors the optimization process: it initializes weights for all objective functions and regularly adapts them with respect to the established priorities. It also resolves the possible conflicts between the target volumes and nearby or overlapping organs at risk, and it generates optimization structures to decrease dose to normal tissues [27].

All 45 patients were retrospectively planned using Ethos IOE based on our 60Gy prostate template (Table 2). The CT scans and the planning contours were exported from Eclipse TPS V16.01.10 to a remote Ethos emulator TPS (V02.01.00).

The Ethos TPS typically generates several plans with different IMRT and VMAT configurations. Nine-fields IMRT configuration was selected for this study, because VMAT has been reported to achieve lower plan quality in the early version of the Ethos TPS [37, 200, 201].

Table 2: IOE initial clinical goals template. Each line refers to an optimization goal for a defined volume with a defined optimization priority. Femur_L = left femoral head, Femur_R = right femoral head.

Priority	Volume	Goal
1: Most Important	PTV	D95% >= 95%
1: Most Important	PTV	D2% < 104%
1: Most Important	CTV	D98% >= 98%
2: Very Important	Rectum	V60Gy < 1%
2: Very Important	Bowel	D0.1cm^3 < 60Gy
2: Very Important	Anal canal	V60Gy < 1%
2: Very Important	Bladder	V60Gy < 3%
2: Very Important	Rectum	V52.8Gy < 15%
2: Very Important	Rectum	V30Gy < 35%
2: Very Important	Bladder	V50Gy < 20%
2: Very Important	Bladder	V40.8Gy < 20%
3: Important	PTV	D50% >= 99%
3: Important	PTV	D50% <= 101%
3: Important	Anal canal	V35Gy <25%
3: Important	Anal canal	V20Gy < 50%
3: Important	Bowel	V55Gy <3%
3: Important	Bowel	V20Gy <300cm^3
4: Less Important	Femur_L	V40Gy <20%
4: Less Important	Femur_L	V35Gy < 5%
4: Less Important	Femur_R	V40Gy <20%
4: Less Important	Femur_R	V35Gy < 5%
4: Less Important	Penile bulb	V42.5Gy < 50%
4: Less Important	Penile bulb	V54.1Gy < 10%

2.2.2. Manually-generated planning (MG plans)

A single planner manually generated plans for the 45 patients, with the goal of achieving the optimal clinical trade-offs. The beam configuration is the same as Ethos 9 fields IMRT. The plans were specifically re-optimized with this configuration for this study. For this reason, MG plans are considered as the gold standard, that is, benchmark plans for the dose comparisons that are reported in Section 2.3. The Eclipse Photon Optimizer algorithm was used for plan optimization and the Acuros External Beam 16.1.0 algorithm for the final volumetric dose calculation with a 2.5mm calculation resolution.

2.2.3. Ethos planning with updated template

Based on the acquired experience from manual planning and Ethos TPS, we updated the initial template from Table 2. The V20Gy constraint for the anal canal was decreased to 40% and optimization goals "V20Gy<40%" were added for rectum, bladder, and penile bulb with priority 4: "*less important*". Later in our experimentations (see section 2.3), 10 challenging patients were selected, and the updated template was applied in order to investigate whether the plan clinical quality could be improved for these 10 specific patients. Those are the 10 testing patients mentioned in Figure 19.

2.2.4. Dose prediction and dose mimicking planning (DP+DM plans)

The dose prediction model used in this work is a hierarchically densely connected U-Net (HD-UNET). The in-house implementation of the architecture is publicly available in https://gitlab.com/ai4miro/ntcp_predicted_dose [207]. Dose prediction with U-net architectures has already been studied over many treatment locations [121, 208-211] and demonstrated good performances across several architecture variations including the HD-UNET version [103, 120, 212]. Our HD-UNET model contains 11 input channels: one channel for the CT-scan; one for the dose prescription mask, which consists of the prescription dose (60Gy) for the voxels inside the PTV and 0 for voxels outside; and 9 channels for the masks corresponding to the organs at risk, namely, the anal canal, bladder, rectum, left femoral head (Femur_L), right femoral head (Femur_R), colon, sigmoid, small bowel, and penile bulb.

To prevent exceeding the GPU (Nvidia A100 40GB) memory limit, the CT scan, masks, and dose were resampled at a resolution of [3x3x3 mm^3] and model training with image patches size of [176 x 176 x 96] was used. Data were augmented by flipping training data in the left-right direction to double the dataset size. The model was trained, validated, and tested using the 45 MG plans, with a 5 folds cross-validation. The mean squared error between the predicted and ground-truth doses (from MG plans) was used as a loss function. The training set included 30 patients, and the validation set 5 patients (those are the "training patients" in Figure 19). The test set contained the 10 selected patients. The model was trained for 300 epochs for each fold. After model training, the weights yielding the lowest mean squared error on the validation set were kept. This was done for all 5 folds, which resulted in 5 different models. For each test patient, the

dose was predicted with these 5 different models, producing 5 different dose maps. A final dose map was then computed from the mean of these 5 dose maps. This way of aggregating the different dose maps over several folds is somehow analogous to bootstrap aggregation (bagging), which has already been reported as increasing the accuracy of dose prediction [213]. The test patients were implicated in neither the model training nor the model selection process.

Concerning dose mimicking, the goal is to optimize machine parameters that best achieve the predicted dose. For that purpose, a template was devised in the Ethos TPS to reproduce closely the dose prediction into a deliverable plan (see Table 3). This mimicking template sets up an isodose-based optimization (IBO) [214]. This template consists in requiring a classical target coverage (PTV:D95%>95% ; CTV:D98%>98% ; PTV:D2%<=104%), to then apply maximum dose constraints on the predicted isodose levels. Isodose volumes were computed from 57-60Gy (95% and 100% of the prescription dose) to 0Gy by steps of 10 Gy (60, 57, 47, 37, 27, 17, 7, 0) with an in-house Python script. The isodose template was refined by including maximum dose constraints on the intersection of some close OARs and isodose levels, to enable a more accurate mimicking on these specific anatomical regions. This mimicking template was set-up and tested on prostate plans that were not part of this study.

Table 3: Dose mimicking template for Ethos IOE. "AnoRectum" refers to the merging volume of anal canal and rectum contours. "Dose [D1-D2 Gy]" refers to the predicted dose volume ranging from dose D1 to dose D2. AnoRectum, bladder, or bowel followed by values in square brackets [D1, D2 Gy] refers to the intersection volume between the anatomical volume and the predicted dose volume ranging from dose D1 to dose D2.

PRIORITY	VOLUME	GOAL
1: Most important	PTV	D95%>= 95%
1: Most important	CTV	D98%>=98%
1: Most important	AnoRectum[60-57Gy]	D0.1cm^3<60Gy
1: Most important	Bladder[60-57Gy]	D0.1cm^3<60Gy
1: Most important	PTV	D2%<=104%
2: Very Important	AnoRectum[47-57Gy]	D0.1cm^3<57Gy
2: Very Important	Bladder[47-57Gy]	D0.1cm^3<57Gy
2: Very Important	Bowel[+57Gy]	D0.1cm^3<60Gy
2: Very Important	AnoRectum[37-47Gy]	D0.1cm^3<47Gy
2: Very Important	Bladder[37-47Gy]	D0.1cm^3<47Gy
2: Very Important	AnoRectum[27-37Gy]	D0.1cm^3<37Gy
2: Very Important	Bladder[27-37Gy]	D0.1cm^3<37Gy
2: Very Important	AnoRectum[17-27Gy]	D0.1cm^3<27Gy
2: Very Important	Bladder[17-27Gy]	D0.1cm^3<27Gy
2: Very Important	AnoRectum[7-17Gy]	D0.1cm^3<17Gy
2: Very Important	Bladder[7-17Gy]	D0.1cm^3<17Gy
2: Very Important	AnoRectum[0-7Gy]	D0.1cm^3<7Gy
2: Very Important	Bladder[0-7Gy]	D0.1cm^3<7Gy
3: Important	Dose[+57Gy]	D0.1cm^3<60Gy
3: Important	Dose[47-57Gy]	D0.1cm^3<57Gy
3: Important	Dose[37-47Gy]	D0.1cm^3<47Gy
3: Important	Dose[27-37Gy]	D0.1cm^3<37Gy
3: Important	Dose[17-27Gy]	D0.1cm^3<27Gy
3: Important	Dose[7-17Gy]	D0.1cm^3<17Gy
3: Important	Dose[0-7Gy]	D0.1cm^3<7Gy

2.3. Planning studies

Two planning studies were performed, in order to investigate the general performance of the Ethos template-based and its patient-optimality (i.e., the capability to produce plans that achieve the best possible dose distribution for a patient, with the highest number of met clinical goals) against DP+DM plans, with respect to reference MG plans. All plans were scaled either to D95=57Gy on the

PTV or D98=58.8Gy on the CTV so that both PTV and CTV coverage constraints could be met.

2.2.5. Study S1 -Performance of ETHOS using initial template

This first study evaluates the global performance of Ethos planning with our initial clinical template (EG_init) against conventional manual planning (MG) over the 45 patients.

The 45 EG_init plans and MG plans were compared through the OAR clinical metrics used in our Ethos template goals and on the mean doses to the OARs. Metrics were extracted from Ethos TPS for EG_init plans and from Eclipse TPS for MG plans. Homogeneity index (HI) and conformity number (CN) were also computed to provide supplementary information about quality of the dose distribution. HI and CN were computed with the following formulas:

$$HI = \frac{D_{2\%} - D_{98\%}}{D_{50\%}} \quad (6)$$

$$CN = \frac{V_{95\%,PTV}^2}{V_{PTV} \times V_{95\%,Body}} \quad (7)$$

where $V_{95\%,PTV}$ and $V_{95\%,Body}$ are the volume receiving 95% or more of the prescribed dose for the PTV and the body, respectively.

Statistical significance between MG and EG_init plans is reported for the clinical goals and mean dose to OARs with a Wilcoxon signed rank test. All the statistical tests are operated with the “stats.Wilcoxon” function from the scipy Python package.

From the results of this comparison, we noticed 10 EG_init plans showing specific difficulties compared to MG plans, either they did not meet some clinical goals that were met with manual planning or the dose to OARs could be overall further improved.

2.2.6. Study S2 - Performance of ETHOS using an updated template for a selected set of patients and comparison with deep learning-based auto-planning

The second study highlights the 10 patients from previous study where Ethos IOE shows optimization difficulties with generic clinical constraints typically used in

template-based optimization. Consequently, a second template which has been updated on basis of results of the first study is used to evaluate if clinical plan quality might be substantially increased. The 10 EG_init plans were kept (EG_init_selected) and corresponding patients were re-planned with the updated template (EG_upd_selected). Moreover, these 10 test patients were also planned with the alternative DP+DM approach. The 10 EG_init_selected, EG_upd_selected and DP+DM plans were compared using homologous MG plans as benchmark. Clinical metric and mean dose to OAR for EG_init_selected plans, EG_upd_selected plans and DP+DM plans were compared. In addition, the numbers of clinical goals that are met were also provided for each approach.

HI and CN are also provided following the same formulas than for Study S1. Modulation complexity score (MCS) was computed for each Ethos multileaf collimator (MLC) with an in-house python script. Details about MCS computation can be found in supplementary materials S1 and in the original MCS article [215]. Deliverability was reported with Mobius gamma index. Recent studies have demonstrated and used the capability of Mobius to serve as a reliable indicator of plan failure [36, 142].

Statistical significance between EG_init_selected and DP+DM as well as between EG_upd_selected and DP+DM is reported for the clinical goals, mean doses to OARs, monitor unit (MU), MCS, and gamma index with a Wilcoxon signed rank test. Statistical tests are operated with the “stats.Wilcoxon” function from the scipy Python package.

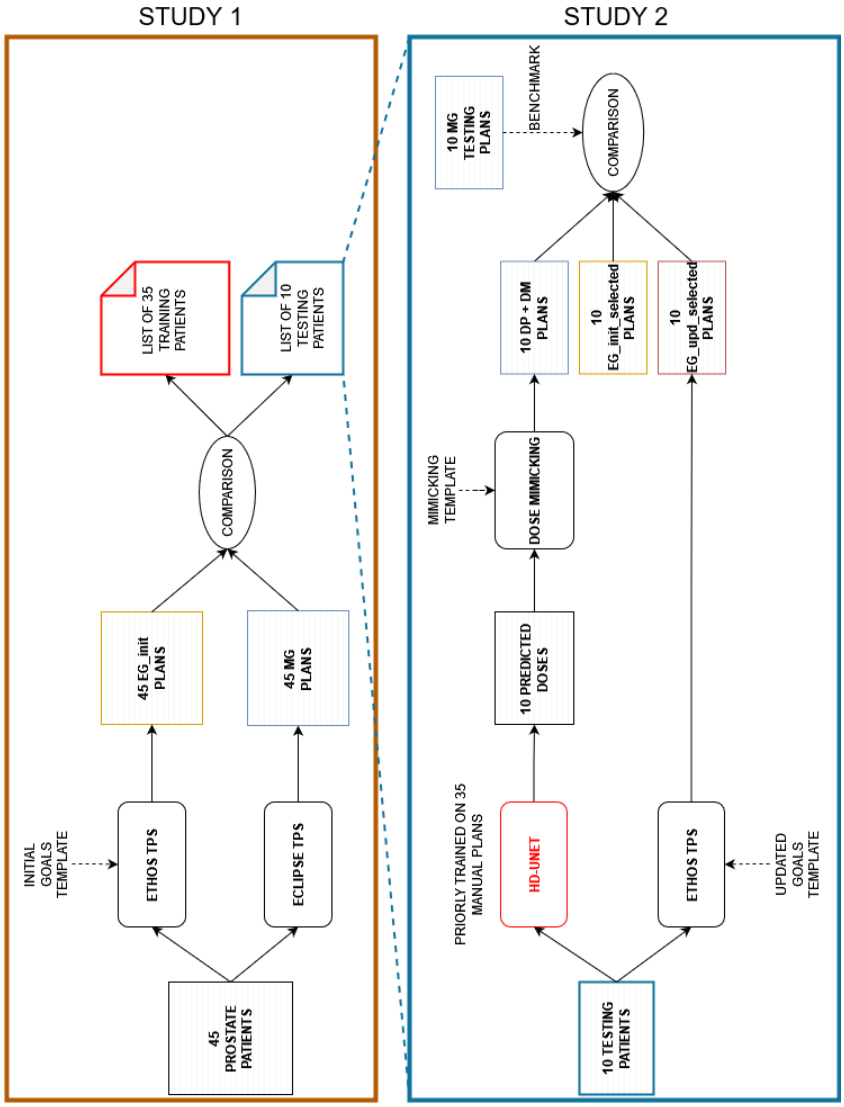


Figure 19

Figure 19: Workflow of the planning studies and data distribution. The patients used in the study 2 come from a manual selection of 10 patients from study 1. The 10 EG_init_selected plans did not undergo any modifications between study 1 and study 2 as for the 10 MG plans used as benchmark in the specific comparison. The 10 EG_upd_selected plans were generated the same way as the 10 EG_init_selected, excepted that the updated goals template was used.

3. Results

3.1. Study S1 - Performance of ETHOS using initial template

Comparison of the metrics between MG plans and EG_init plans are reported in Table 4. For metrics close to prescription dose (50Gy and higher), EG_init plans achieved similar mean values than MG plans, absolute differences (|EG_init minus MG|) never exceeding 0.6%. Some of them are established as statistically significant, however, we consider these differences of less than 1% to be clinically insignificant.

For Lower dose levels, EG_init plans tend to show larger dose differences with MG plans: 13%, 6.6%, and 5.3% more, for V20Gy the anal canal, V30Gy to rectum and V42.5Gy to penile bulb, respectively. For the rest of the considered metrics, results achieved by EG_init plans were deemed satisfactory.

The previously observed trend about lower dose levels seems to be confirmed when looking at the mean dose distribution for each OAR in Figure 20. Again, OARs that show noticeable differences are the anal canal, bladder, rectum, and penile bulb with median mean dose difference of 5.0Gy, 3.7Gy, 4.9Gy, and 5.8Gy, respectively.

Table 5, reports the HI and CN for both planning approaches. We can notice that the mean HI for MG plans is slightly better than for EG_init plans. Mean CN looks comparable between the two approaches.

The V20Gy<40% constraints for the updated template were chosen because precise mean V20Gy for the anal canal, bladder, rectum, and penile bulb are, respectively, 21.3%, 28.1%, 35.6%, and 23.2% over the 45 MG plans. These values over the 10 MG plans for the selected patients rise to 30.0%, 32.9%, 42.5%, and 23.4% for the anal canal, bladder, rectum, and penile bulb, respectively. Such additional V20Gy constraints aim at reasonably reducing the most extreme differences for lower dose levels of EG_init plans. Results of this template update can be found in the following section for the 10 selected patients.

Table 4: Mean metrics achieved for EG_init plans and MG plans over the 45 patients. Differences (EG_init minus MG) are also provided. The p-values are extracted from the comparison of EG_init plans and MG plans. Those presented in bold font are deemed statistically significant ($p < 0.05$). EG_init = Ethos generated plan (using initial template), MG = Manually generated plan.

Volume	Constraint	EG_init plans	MG plans	EG_init – MG	P-value
Anal canal	V60Gy (%)	0.3 ± 0.7	0.5 ± 0.6	-0.2	**0.006**
	V35Gy (%)	17.4 ± 9.3	14.3 ± 10.2	3.1	**$<10^{-3}$**
	V20Gy (%)	34.3 ± 14.5	21.3 ± 13.2	13.0	**$<10^{-3}$**
Bladder	V60Gy (%)	0.8 ± 1.7	1.0 ± 0.8	-0.2	**$<10^{-3}$**
	V50Gy (%)	10.8 ± 4.0	10.3 ± 4.6	0.5	**0.002**
	V40.8Gy (%)	15.8 ± 5.3	14.1 ± 6.0	1.7	**$<10^{-3}$**
Bowel	D0.1cm³ (Gy)	23.0 ± 24.1	21.4 ± 23.8	1.6	**0.020**
	V55Gy (%)	0.2 ± 0.7	0.2 ± 0.6	0.0	**0.020**
	V20Gy (cm³)	5.0 ± 9.7	3.7 ± 7.8	1.3	**0.002**
Rectum	V60Gy (%)	0.7 ± 1.6	0.7 ± 0.7	0.0	**0.040**
	V52.8Gy (%)	14.1 ± 2.8	14.3 ± 3.9	-0.2	0.780
	V30Gy (%)	34.1 ± 2.9	27.5 ± 6.5	6.6	**$<10^{-3}$**
Femur_L	V40Gy (%)	0.0 ± 0.1	0.1 ± 0.1	-0.1	**0.009**
	V35Gy (%)	0.0 ± 0.2	0.5 ± 1.0	-0.5	**$<10^{-3}$**
Femur_R	V40Gy (%)	0.0 ± 0.1	0.1 ± 0.2	-0.1	**0.004**
	V35Gy (%)	0.0 ± 0.3	0.4 ± 1.0	-0.4	**$<10^{-3}$**
Penile Bulb	V54.1Gy (%)	3.8 ± 5.5	4.4 ± 7.7	-0.6	**$<10^{-3}$**
	V42.5Gy (%)	15.5 ± 12.8	10.2 ± 11.4	5.3	0.120

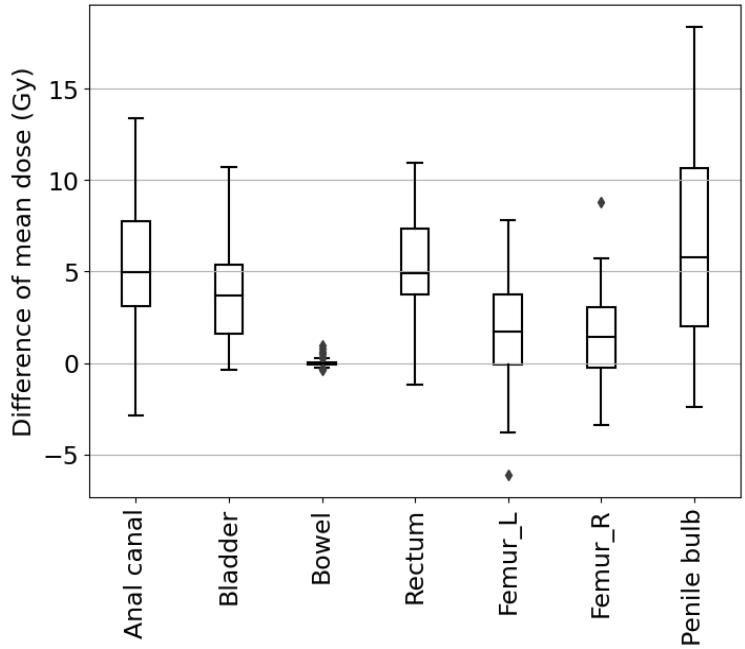


Figure 20: Difference of mean dose to OAR between EG_init and MG plans over the 45 patients. The inner line represents the median of the dataset, the box are the quartiles, the whiskers show the range of the distribution without outliers which are denoted by a diamond. Outliers are determined by the boxplot function of the seaborn Python package. All mean dose differences are statistically significant.

Table 5: Mean homogeneity index (HI) and conformity number (CN) for Ethos generated (EG_init) and manually generated (MG) plans

	EG_init	MG
HI	0.12 ± 0.02	0.10 ± 0.02
CN	0.89 ± 0.01	0.89 ± 0.03

3.2. Study S2 - Performance of ETHOS using an updated template for a selected set of patients and comparison with deep learning-based auto-planning

Metric differences for DP+DM, EG_init_selected and EG_upd_selected are reported in Table 6. Concerning dose levels above 50Gy, several of these metrics are now above 1% difference for EG_init_selected plans: V50Gy to bladder, V60Gy and V52.8Gy to rectum and V54.1Gy to penile bulb. As for S1, large mean differences can be observed for lower dose levels such as 16.9%, 8.5%, and 11.2%, for V20Gy to the anal canal, V30Gy to the rectum, and V42.5Gy to the penile bulb, respectively. This repeats the same trends as for S1, although the differences with MG plans get amplified due to patient selection.

Concerning EG_upd_selected plans, we noticed a slight increase of the V60Gy to rectum compared to EG_init_selected, while lower dose levels differences with MG dropped: from 16.9% to 11.5% for the anal canal V20Gy, from 8.5% to 5.7% for the rectum V30Gy, and from 11.2% to 3.4% for the penile bulb V42.5Gy.

The DP+DM approach overall achieved metrics closer to those from MG, in comparison with EG_init_selected and EG_upd_selected. Noticeable exceptions are the penile bulb V54.1Gy of 4.6% instead of 1.9% and 0.5% for EG_init_selected and EG_upd_selected, respectively. However, these differences did not reach statistical significance between DP+DM and EG_init_selected and between DP+DM and EG_upd_selected.

Figure 21 illustrates the difference of mean dose to OARs with regards to MG plans for the three automatic approaches. As previously observed, the updated template reduced the dose difference for the low dose levels, and consequently for the mean dose to OAR. However, DP+DM outperformed the other automatic methods, and achieved the closest dose metrics to the MG method. Statistical difference between EG_upd_selected and DP+DM plans is moreover established for the anal canal and bladder.

Table 7 depicts when clinical goals were met or not for each approach. While the 10 MG plans could not meet all clinical goals, they were deemed clinically acceptable by our physicians. As to the automatic approaches, DP+DM succeeds in meeting the largest number of “most important” and “important” goals, followed by EG_init_selected and EG_upd_selected. The 10 DP+DM plans were

also deemed clinically acceptable, but 1 EG_init_selected plan and 3 EG_upd_selected plans were not, due to critical unmet goals. Among these unmet goals, there is notably the "V60Gy<1%" to rectum that could not be met for 2 EG_init_selected and 4 EG_upd_selected plans. Figure 22 illustrates the distribution of V60Gy to rectum for each approach. We can notice that V60Gy of 2 EG_init_selected plans were significantly above 1% constraint and that template update resulted in a noticeable increase of V60Gy for 2 other patients, compromising the clinical quality of these EG_upd_selected plans.

Table 8 reports the HI and CN for the different planning methods, it can be noticed that MG and DP+DM achieved a better homogeneity while their conformity number is slightly lower than for EG_init_selected and EG_upd_selected plans.

Table 9 reports the mean total MU, MCS and Mobius gamma index for each approach. There are statistically significant differences between MU for MG plans and EG as well as for DP+DM and EG plans. MCS for both Ethos MLCs confirmed the higher complexity of DP+DM plans. This is expected since the dose metrics are probably more demanding for some constraints than the general Ethos template. However, no prostate plans were flagged as problematic while the mean gamma index is slightly lower for the MG plans than for other approaches.

Figure 23 illustrates the dose distribution for the different optimization methods for a specific patient where the V60Gy constraint to rectum was difficult to meet. Some facts already mentioned above can be observed: the DP+DM and MG methods reduce at best the V60Gy to rectum while ensuring PTV coverage. Such rectum sparing seems to have a cost in terms of dose conformity when compared to EG_init_selected and EG_upd_selected.

OAR	Constraint	DP+DM - MG	EG_init_selected - MG	P-value DP+DM;EG_init_selected	EG_upd_selected - MG	P-value DP+DM;EG_upd_selected
Anal canal	V60Gy (%)	0.3 ± 1.0	0.1 ± 0.7	0.079	0.4 ± 1.3	0.500
	V35Gy (%)	1.2 ± 3.1	6.0 ± 5.9	**0.006**	4.8 ± 4.8	**0.004**
	V20Gy (%)	4.6 ± 4.7	16.9 ± 11.4	**0.004**	11.5 ± 7.9	**0.006**
Bladder	V60Gy (%)	-0.3 ± 0.5	0.1 ± 0.6	0.062	0.4 ± 1.0	**0.028**
	V50Gy (%)	0.6 ± 1.2	1.4 ± 0.9	0.130	1.3 ± 0.8	**0.014**
	V40.8Gy (%)	0.6 ± 1.8	2.9 ± 2.9	**0.010**	2.2 ± 1.3	**0.004**
Rectum	V60Gy (%)	-0.2± 0.3	1.4 ± 3.1	0.058	1.9 ± 3.1	**0.034**
	V52.8Gy (%)	0.0 ± 1.2	1.0 ± 2.6	0.193	0.7 ± 2.5	0.275
	V30Gy (%)	3.4 ± 2.8	8.5 ± 5.1	**0.027**	5.7 ± 5.0	0.375
Bowel	V55Gy (%)	-0.1± 0.8	-0.2 ± 0.8	0.625	-0.2 ± 0.8	0.769
	V20Gy (cm^3)	-1.3± 5.6	0.8 ± 5.8	0.180	-0.6 ± 5.9	0.102
	D0.1cm^3(Gy)	0.5 ± 2.8	1.0 ± 2.5	0.109	0.1 ± 1.8	0.108
Femur_L	V40Gy (%)	0.0 ± 0.2	-0.1 ± 0.2	0.180	-0.1 ± 0.2	0.180
	V35Gy (%)	0.0 ± 1.1	-0.6 ± 1.1	**0.028**	-0.4 ± 0.9	**0.043**
Femur_R	V40Gy (%)	0.0 ± 0.1	0.0 ± 0.1	0.157	0.0 ± 0.1	0.157
	V35Gy (%)	0.0 ± 0.6	-0.3 ± 0.5	**0.043**	-0.2 ± 0.5	**0.043**
Penile Bulb	V54.1Gy (%)	4.6 ± 7.7	1.9 ± 4.4	0.080	0.5 ± 5.5	0.080
	V42.5Gy (%)	6.6 ± 9.0	11.2 ± 11.8	**0.028**	3.4 ± 9.4	0.068

Table 6 : Mean difference of dose metric between automatic plans (EG_init_selected, EG_upd_selected; and DP+DM) and MG plans for the 10 selected patients. P-values between DP+DM and EG_init_selected metrics distribution as well as between DP+DM and EG_upd_selected are provided. Those presented in bold font are deemed statistically significant ($p < 0.05$).

Table 7: The total number of met OAR goals for the 10 patients are reported for each method. For each of the 10 testing plans there are 8 "very important" goals, 4 "Important" goals and 6 "less important" OAR goals to satisfy. Highest number of met goals per importance for automatic planning methods are bolded. EG_init_selected = Ethos generated (using initial template) for patient manually selected among 45 initial patients, EG_upd_selected = Ethos generated (using updated template) for patient manually selected among 45 initial patients, DP+DM = dose prediction followed by dose mimicking, MG = manually generated.

Priority	2: Very important	3: Important	4: Less Important
EG_Init_selected	66/80	37/40	**59/60**
EG_upd_selected	62/80	38/40	**59/60**
DP+DM	**70/80**	**39/40**	58/60
MG	74/80	40/40	60/60

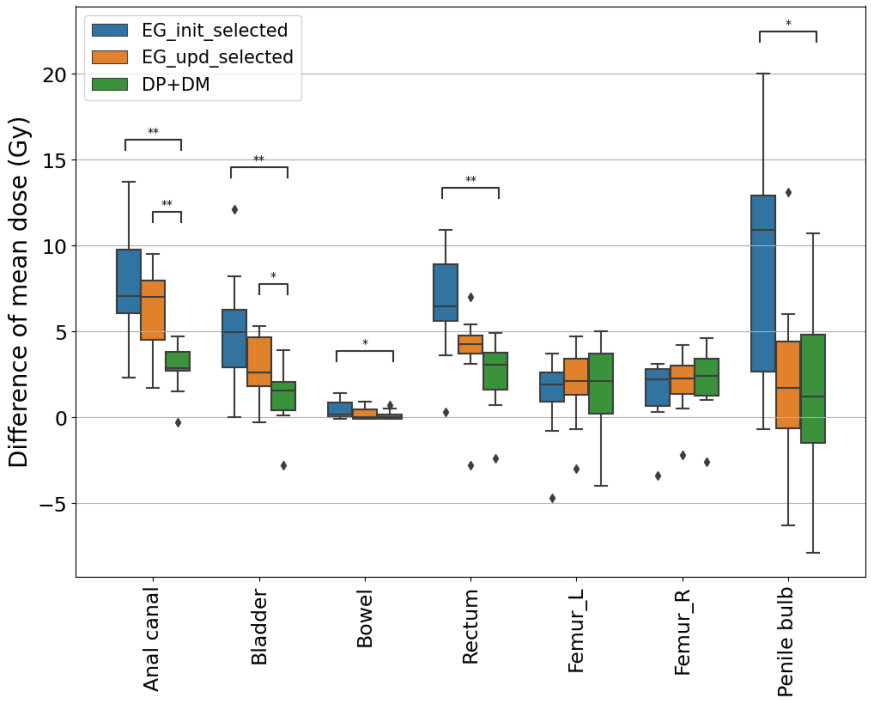


*Figure 21: Difference of mean dose to OAR between automatic planning methods and manual ground truth planning. The horizontal links between boxes denote a statistically significant mean dose difference between the 2 linked boxes (p-values lower than 5% are denoted by *; p-values lower than 1% are denoted by **). The inner line represents the median of the dataset, the box are the quartiles, the whiskers show the range of the distribution without outliers which are denoted by a diamond. Outliers are determined by the boxplot function of the seaborn Python package. EG_init_selected = Ethos generated (using initial template) for patient manually selected among 45 initial patients, EG_upd_selected = Ethos generated (using updated template) for patient manually selected among 45 initial patients, DP+DM = dose prediction followed by dose mimicking.*

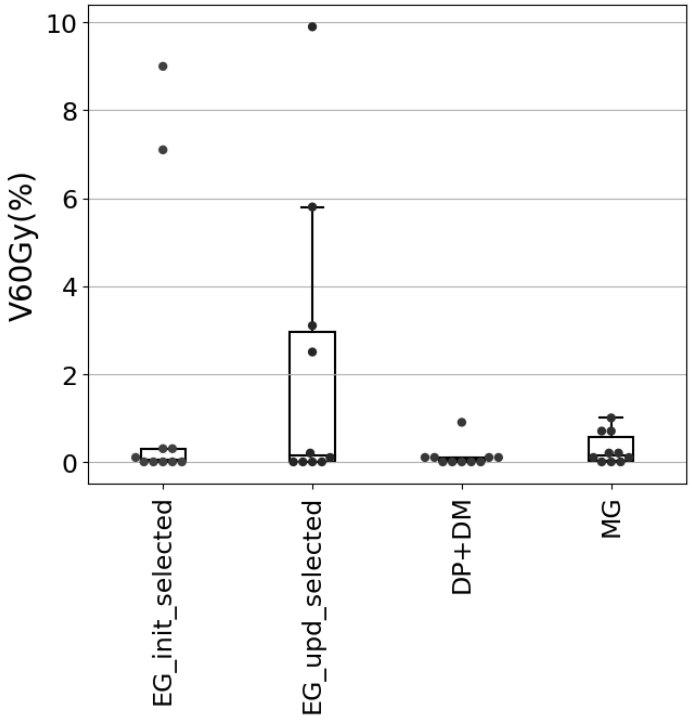


Figure 22: Rectum V60Gy for each optimization method over the 10 selected patients, individual values are denoted by a dot. The inner line represents the median of the dataset, the box are the quartiles, the whiskers show the range of the distribution without outliers. We can notice two patients having their V60Gy noticeably increasing above 1% when using the updated template instead of the initial one. EG_init_selected = Ethos generated (using initial template) for patient manually selected among 45 initial patients, EG_upd_selected = Ethos generated (using updated template) for patient manually selected among 45 initial patients, DP+DM = dose prediction followed by dose mimicking, MG = manually generated.

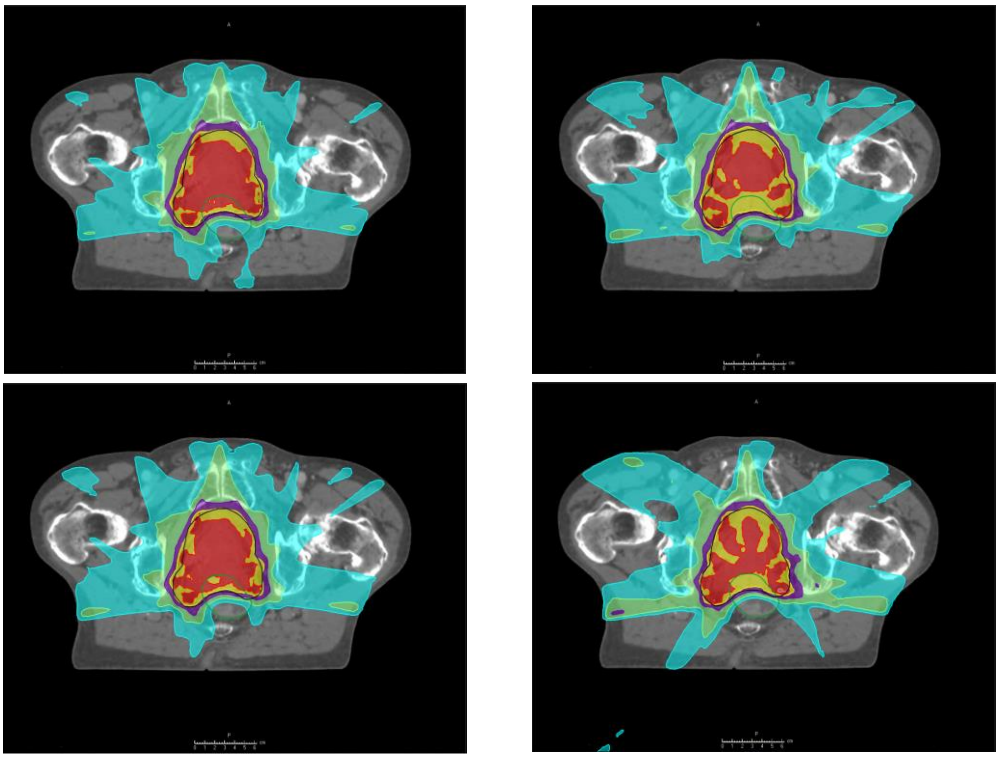


Figure 23: Dose distribution of each optimization method for a patient where the V60Gy constraint to the rectum was difficult to meet. Black contour: PTV; green contour: rectum. Red isodose: dose ⩾ 60Gy; yellow isodose: 57Gy⩽dose<60Gy; purple isodose: 48Gy⩽dose<57Gy; light green isodose: 36Gy⩽ dose<48Gy; light blue isodose: 24Gy⩽dose<36Gy.Top left: EG_init_selected ; bottom left: EG_upd_selected ; top right: DP+DM ; bottom right : MG. EG_init_selected = Ethos generated (using initial template) for a patient manually selected among 45 initial patients, EG_upd_selected = Ethos generated (using updated template) for a patient manually selected among 45 initial patients, DP+DM = dose prediction followed by dose mimicking, MG = manually generated.

Table 8: Mean homogeneity index (HI) and conformity number (CN) for the different planning methods over the 10 selected patients. EG_init_selected = Ethos generated (using initial template) for a patient manually selected among 45 initial patients, EG_upd_selected = Ethos generated (using updated template) for patient manually selected among 45 initial patients, DP+DM = dose prediction followed by dose mimicking, MG = manually generated.

	EG_init_selected	**EG_upd_selected**	**DP+DM**	**MG**
HI	0.12 ± 0.02	0.12 ± 0.02	0.08 ± 0.01	0.09 ± 0.02
CN	0.89 ± 0.01	0.89 ± 0.01	0.88 ± 0.02	0.87 ± 0.02

Table 9: Mean total monitor unit (MU), modulation complexity score (MCS) for the two Ethos multileaf collimators and Gamma index (Mobius γ). EG_init_selected = Ethos generated (using initial template) for a patient manually selected among 45 initial patients, EG_upd_selected = Ethos generated (using updated template) for a patient manually selected among 45 initial patients, DP+DM = dose prediction followed by dose mimicking, MG = manually generated.

	EG_init_selected	P-value EG_init_selected-DP_DM	EG_upd_selected	P-value EG_upd_selected-DP_DM	DP+DM	P-value DP+DM-MG	MG
Total MU	1947.8 ± 434.1	0.010	1941.0 ± 301.9	0.002	2397.5 ± 332.6	0.002	2920.03 ± 564.4
MCS 1	0.26 ± 0.04	0.049	0.26 ± 0.04	0.002	0.23 ± 0.03	0.014	0.20 ± 0.03
MCS 2	0.27 ± 0.05	0.020	0.28 ± 0.04	0.002	0.24 ± 0.02	0.014	0.21 ± 0.03
Mobius γ (%)	99.9 ± 0.00	0.020	99.9 ± 0.00	0.020	99.8 ± 0.06	0.008	99.4 ± 0.43

4. Discussion

This work consisted in 2 planning studies aiming at evaluating the general performance of Ethos template-based automatic plan optimization and its patient-optimality against another popular approach: deep-learning dose prediction followed by dose mimicking. For study S1, we generated plans for 45 patients using our initial Ethos clinical goals template (EG_init), and compared them to manually generated plans (MG). Study S2 used 10 specific patients from S1 where IOE with initial goals template faced optimization difficulties when compared to MG plans. For these 10 selected patients, plans were generated with 3 different automatic approaches: Ethos IOE with initial template (EG_init_selected), Ethos IOE with updated template (EG_upd_selected) and DP+DM. The plans from these 3 automatic approaches were then compared to their homologous manual plan.

We first showed that the performance of the Ethos template-based approach over our database of 45 patients, using our initial clinical template (EG_init), was globally satisfactory for dose levels close to the prescription dose, although lower levels looked loosely optimized, especially for the anal canal, rectum, bladder, and penile bulb. We then compared the Ethos template-based approach to the DP+DM approach on 10 selected challenging patients and showed that while most EG_init_selected plans were still clinically acceptable, the DP+DM plans outperformed them. Number of clinical goals that were met, doses to OARs, homogeneity index and clinical acceptability were favorable to our DP+DM method while the deliverability did not seem to be impacted according to our results. MG plans performance suggested that quality of EG_init plans might be enhanced by updating our Ethos template. We investigated that hypothesis by adding low priorities constraints to the clinical template. While this enhanced some clinical metrics, this also deteriorated some other high-priority metrics, such as the V60Gy to the rectum. EG_upd_selected plans unfortunately resulted in fewer clinically acceptable plans and highlighted the limitation of using generic constraints from a template when optimizing plans with IOE.

The V60Gy to rectum that were noticeably above 1% for 2 EG_init_selected and 4 EG_upd_selected plans might stem from the limited number of iterations allotted for the dose optimization process for such challenging plans. During dose optimization, even if the IOE sequentially updates the weights of goals from higher to lower priority [27], low-priority goals still affect the global solution convergence, just due to their simple presence in the objective function from its initialization. Considering that the more complex the goals are to achieve, the more iterations are necessary, lower-priority goals might make the objective function very complex and prevent its convergence from occurring within the iteration budget allotted by the manufacturer [216]. This would explain why 2 plans that were initially clinically deliverable were no longer acceptable after the update: adding an extra dose objective likely required more iterations, which in turn exceeded the preset default limit for some plans. Such assumption about an insufficient number of iterations might also lead to additional conclusions related to our DP+DM approach. DM consists in optimizing constraints derived from predicted dose volumes through IOE. It might thus be supposed that with unrealistic predicted doses and optimization objectives, the IOE would naturally have needed an extremely high number of iterations to try fulfilling conflicting

goals. However, it seems that DP+DM plans were not much affected by this iteration limit, which tends to confirm the reliability of the predictions made by our HD-UNET models. Moreover, this would also highlight the capability of the IOE to produce qualitative plans for challenging cases within the iteration budget when patient-specific dose constraints are used instead of generic ones.

From these results, DP+DM seems to better approximate optimality for a wider range of patients than the Ethos template. The overall quality of Ethos plans is not questioned, but this study illustrates the difficulty of having an Ethos template that can both deliver clinically acceptable plans for a vast population of patients and spare OARs as much as possible for every individual case. Nonetheless, it is important to analyze from a broad perspective the potential and limitations of both automatic planning approaches (Ethos template and DP+DM):

First, DP+DM requires some time to set up. It is necessary to have access to a database of clinical plans large enough to be able to train the model, this can take several weeks of data collection or generation, as well as model testing. Dose mimicking can also consume time to best reproduce the predictions. From our experience, mimicking set-up and testing can take from days to weeks. Moreover, it is noteworthy that our model was exclusively trained for Ethos 9 fields IMRT configuration, while if 12 fields IMRT or VMAT plans were to be also considered, then new dose prediction models should be trained for each new beam configuration. In contrast, a single IOE template can be used for all beam configurations. Even using transfer learning from 9 fields IMRT model to a new configuration model, this would require additional plans with the new beam configuration [209]. These processes lead to an even longer implementation and contrast with the Ethos template solution designed to be fast and easy to implement.

Second, the DP+DM workflow is inherently longer than Ethos initial planning workflow. Both methods lead to a deliverable plan through the Ethos IOE, however, DP+DM necessitates firstly to load CT and contours in a trained HD-UNET to then predict the dose. This prediction step in our case takes approximately 50 to 60 seconds per plan (this accounts for the 5 predictions from the cross-validation models on a single GPU used to

produce the final prediction of the dose map). Then, the generated isodose volumes are computed with our in-house script and uploaded to Ethos server in less than 10 seconds. Last, the dose mimicking step needs to be performed, which takes a time similar to the optimization with a template for EG plans. This results in a plan generation for our DP+DM method that is about 1 min longer than the EG approach in our case. However, such additional time might probably be optimized, for example by computing the 5 predictions in parallel.

Third, DP+DM was not evaluated over the whole patient database. In order to demonstrate the global superiority of DP+DM over Ethos template approach, it would require proving superiority over a larger dataset and not only on plans were Ethos IOE templates faced difficulties. Nevertheless, this does not detract from the conclusion that with an approach such as DP+DM, a more optimal automatic dose optimization seems possible at the cost of a longer and more demanding (in terms of training database generation) implementation of the workflow.

Concerning the adaptive workflow, clinical goals defined at initial planning are then further used for dose optimization at each adaptive session, consequently a question arises: "*Are challenging plans suited for the Ethos adaptive workflow?*". In the case where dose optimization would fail to converge to an acceptable solution for initial planning, we can suppose that results would not be substantially better during the adaptive session and that there is a need for specific tuning of clinical objectives. However, excessive specific tuning of clinical goals during initial planning step might be risky later in adaptive workflow since patient anatomy changes over time. Passed a tipping point of tuning, IOE could no longer converge to a satisfying dose distribution for each plan of the day before reaching the supposed-critical number of iterations.

To prevent scenarios where dose would not converge for adaptive sessions, a recommendation based on our results would be to stick to general constraints when setting-up Ethos clinical goal templates. If even with a generic template, dose optimization during initial planning proves to be challenging, clinicians should consider excluding these patients from Ethos adaptive workflow. If the IOE cannot converge towards a dose optimum for most fractions, it is likely that the

adaptive workflow will increase treatment time without clear benefits from plan adaptation. Keeping general constraints would, however, mean that some plans will inevitably be sub-optimal in comparison with what is manually achievable.

Note that the adaptive workflow in Ethos needs to be selected before the first fraction and is then maintained throughout the entire treatment. Considering the current limitations of the template-based approach, an interesting option could be to allow for a switch-up or switch-down of the adaptive workflow during the treatment course, based on the observed evolution of patient anatomy. This flexibility could address the cases where Ethos plan adaptation may not necessarily provide a clear added value for a specific patient. Alternatively, it could also be beneficial to directly visualize and select the initial plan recomputed on the daily image without having to wait for the re-optimized plan.

In the end, this is likely to be one of the keys to exploit the full potential of OART on Ethos: getting as patient-optimal as possible in a robust way. Although there seems to be room for improvement, and even if added value of treatment adaption cannot be ensured for every patient, recent studies have shown promising results in demonstrating the added value of Ethos plan adaptation for the global patients' cohort [37, 144-146, 150, 151]. To achieve further value for adaptive treatments, future steps could involve utilizing DP as a reference of specific achievable metrics similarly to the DVH prediction of RapidPlan (Varian Medical system, Palo Alto, USA) but with extra-information such as the predicted 3D dose map. Also, having an indicator of how close to the iteration limit the optimizer gets during the initial planning could convey an interesting and complementary piece of information when trying to fine-tune dose constraints or when considering whether a patient should enter in the adaptive workflow or not.

5. Conclusion

This study has analyzed the general performances and capability to produce patient-optimal plans for Ethos template-based optimization in comparison with deep-learning dose prediction followed by dose mimicking (DP+DM). Our results suggest the existence of a limitation of the Ethos template-based optimization with increasing complexity of the template (i.e. higher number and more aggressive clinical goals). This limitation might be related to a limit in the number of iterations. Consequently, the clinical goals in the Ethos template must be

defined very carefully: too restrictive or too stringent goals might lead to a significant proportion of plans that will not have time to converge to an optimal dose distribution; conversely, too lenient goals will lead to an important proportion of loosely optimized plans. Concerning DP+DM, this approach outperformed the classical template-based approach of Ethos for initial and updated template of clinical goals, showing that a more patient-optimal automatic dose optimization seems possible at cost of a longer and more demanding implementation workflow.

To some extent, these investigations reflect the antagonism between explicit formulation of objectives using human expertise and brute-force connectionism relying on big data to blindly reach similar objectives. Although the first approach might be more interpretable and thus reassuring for users, the second approach raises more and more interest due to its flexibility and capability to cover very specific or rare cases, although it might struggle to gain the experts' trust [217]. Considering these advantages and shortcomings, as well as the clinical convenience of a template-based approach, the choice of such approach in the Ethos automatic optimization appears to be rational and relevant, until AI systems further progress with, for instance, the capability to interact with natural language.

Chapter 3: A dosimetric approach to the automatic segmentation problem

This chapter is published as:

DIVE-ART: A tool to guide clinicians towards dosimetrically informed volume editions of automatically segmented volumes in adaptive radiation therapy.

Benjamin Roberfroid, John A Lee, Xavier Geets, Edmond Sterpin, and Ana M Barragán-Montero. *Radiotherapy and Oncology*, 192:110108, 2024.

1. Introduction

Automatic segmentation in radiation therapy has raised considerable interest for many years, since it allows the clinicians to automate the tedious task of manually contouring anatomical structures and tumor volumes. The accuracy of these algorithms is crucial, particularly in the context of online adaptive radiotherapy, where the time for review and correction should be as short as possible.

In recent years, the accuracy of auto-segmentation algorithms has reached a tipping point that enables a broad adoption in clinical environments, mainly thanks to the use of machine learning methods. In particular, deep convolutional neural networks (DCNNs) have demonstrated their ability to delineate various anatomical volumes quite accurately across multiple locations [37, 102, 117, 144, 147, 218-220]. However, in their current state, DCNNs can still not be used clinically without human supervision, since they cannot guarantee consistently good accuracy for every single volume. Few studies have analyzed the dosimetric quality of treatment plans using uncorrected auto-segmented contours [151, 221]. While these studies have shown overall good dosimetric quality, some cases might remain, for which treating without correction can lead to severe consequences for the patient. In order to spot these cases, clinicians typically perform slice-by-slice visual inspection and manual correction of every volume to ensure patient's safety [12, 28, 176, 222, 223]. This review process can take from few minutes up to half an hour for complex cases, such as for head and neck cancer [117, 224], which considerably slows down or even hamper online adaptive workflows [144-146, 176].

Although cumbersome and inefficient, this review process is currently the only solution to an existing problem: the lack of quality assurance (QA) tools or metrics for auto-segmentation algorithms. Recently, some groups have started to approach this problem from different perspectives. For instance, predicting geometric indicators like the Dice similarity index [225-227]; using a secondary and independent algorithm (i.e., deformable registration or yet another DCNN) to detect segmentation errors [228, 229]; or estimating the uncertainty of the DCNN prediction [230, 231], among other approaches. However, all these remain purely focused on global segmentation and geometric quality, irrespective of subsequent steps and objectives in the process of treatment planning. Indeed, when planning a radiotherapy treatment, organ and target volume segmentation is not a goal per

se and the actual endpoint remains the dosimetric quality of the plan. QA tools relying on geometric indicators are thus agnostic proxies that, in the end, may not correlate well with clinical quality [218, 232, 233]. From that perspective, it is expected that segmentation accuracy matters more in some regions of the anatomy and less in others. Within a restricted time budget for correction, like in online adaptive radiotherapy, it then makes sense to focus the contour editions on those regions that have higher dosimetric impact, whereas regions in the dose map that are hardly sensitive to segmentation quality become secondary.

In this work, we propose a QA tool for auto-segmentation algorithms designed from a dosimetric perspective instead of a geometric one: the DIVE (Dosimetrically Informed Volume Edition) map, which combines spatial deformations of the auto-segmented volumes and DCNN to predict the dosimetric relevance of different regions [121, 208-211]. Our solution takes the form of a graphical indicator, in the spirit of a heatmap, which can be overlaid on the planning image (e.g., CT or CBCT). We hypothesize that the DIVE-map could guide clinicians visually to disregard irrelevant corrections and focus on those with high dosimetric impact, significantly speeding up the contour review process. From a broad perspective, we believe that such a visual tool could mitigate the last remaining shortcomings of automatic segmentation, by further reducing the subsequent contour edition time in clinical conditions.

2. Materials and methods

This section describes the process to generate a DIVE-map and the data used for it. As a proof-of-concept , we analyze the added value of the DIVE-map, i.e. how much correction could the DIVE-map discard while preserving the dosimetric quality, for online adaptive treatments of prostate cancer patients using a commercially available adaptive solution: ETHOS, from Varian (Palo Alto, USA) [27]. The generation process walks through four main steps that are summarized in Figure 24. We focus on two main organs at risk (OAR): bladder and rectum, which were generated for each treatment fraction by the DCNN-based ETHOS auto-segmenter. Consequently, the reduction of manual correction was also evaluated per fraction.

2.1. Data

Our database consisted of 12 prostate cancer patients (9 for model development; 3 for evaluation), without pelvic nodal irradiation. The dose prescription was 60 Gy to the prostate and 48 Gy to the seminal vesicles, delivered in 20 fractions. The CTV-to-PTV expansion consisted in an isotropic margin of 7 mm. CT images were acquired on a Toshiba Acquilion CT scanner with a slice thickness of 2mm. All plans were optimized with the ETHOS “intelligent optimization engine” (IOE), using the clinical goals template at Table 1 and a 9-fields IMRT beam configuration. For each treatment fraction, we have at our disposal both the raw auto-segmented volumes on the daily CBCT and the corresponding manually corrected contours that were clinically used. These corrections were done by the physician in charge, following the conventional process of slice-by-slice visual inspection and review. The conventionally corrected contours are referred hereafter to as “ground truth”.

Local ethics committee approval by the CEHF (Comité d'Ethique hospitalo-facultaire) was obtained for the use of all patient data under the agreement "Learning from the past: the MIRO treatment planning database".

Table 10: IOE clinical goals and "Impact of correction" (IOC) metrics. Each line refers to an optimization metric for a specific volume with a user-defined optimization priority, if used for IOE. Femur_L = left femoral head, Femur_R = right femoral head, PTV_prostate= PTV volume for the prostate, PTV_vs = PTV volume for the seminal vesicles. IOC, metrics will be used later for the DIVE-map generation (see section 2.2).

Volume	Metric + goal if used for IOE	Priority if used for IOE	Usage for IOC (Y/N)
PTV_prostate	D95% >= 98%	1: Most Important	Y
PTV_prostate	D2% <= 104%	1: Most Important	N
PTV_vs	D95% >= 98%	1: Most Important	Y
PTV_vs	D2% <= 104%	1: Most Important	N
Rectum	V60Gy < 1%	2: Very Important	N
Anal canal	V60Gy<1%	2: Very Important	N
Bladder	V60Gy < 3%	2: Very Important	N
Rectum	V52.8Gy < 15%	2: Very Important	Y
Rectum	V30Gy < 35%	2: Very Important	Y
Bladder	V50Gy < 20%	2: Very Important	Y
Bladder	V40.8Gy < 20%	2: Very Important	Y
Bowel	D0.1cm^3<60Gy	2: Very Important	N
PTV_prostate	D50% >= 99%	3: Important	N
PTV_prostate	D50% <= 101%	3: Important	N
PTV_vs	D50% >= 99%	3: Important	N
PTV_vs	D50% <= 101%	3: Important	N
Bowel	V55Gy <3%	3: Important	N
Bowel	V20Gy <300cm^3	3: Important	N
Anal canal	V35Gy<25%	3: Important	N
Anal canal	V20Gy<50%	3: Important	N
Femur_L	V40Gy <20%	4: Less Important	N
Femur_L	V35Gy < 5%	4: Less Important	N
Femur_R	V40Gy <20%	4: Less Important	N
Femur_R	V35Gy < 5%	4: Less Important	N
Penile bulb	V42.5Gy < 50%	4: Less Important	N
Penile bulb	V54.1Gy < 10%	4: Less Important	N
Skin	D1cm^3<50Gy	4: Less Important	N
Rectum	D1%	Not used	Y
Bladder	D1%	Not used	Y
Anal canal	D1%	Not used	Y
CTV_prostate	D98%	Not used	Y
CTV_vs	D98%	Not used	Y

2.2. DIVE-map generation

2.2.1. General workflow

For a given auto-segmented volume, the DIVE-map is generated in four steps (Figure 24). First, in order to simulate possible contouring errors of the given volume, N spatial deformations are applied to its contour (the planning image remains untouched), following different amplitudes and directions (section 2.2.2). The deformed volume is added to the rest of the planning structures, namely, the non-deformed OARs and targets, resulting in a total of N deformation scenarios ($S_{d,d=1,..N,}$). Second, we used a DCNN to predict a fast and precise approximation of the 3D dose distribution for the initial (S_0) and deformed (S_d) scenarios (section 2.2.3). This aims at evaluating the dosimetric impact of the simulated contouring errors in the different S_d scenarios. The third step consists in computing a surrogate of the global dosimetric impact of generating a plan with scenario S_0 rather than scenario S_d (assuming S_d would be a correction of S_0). This surrogate of the dosimetric impact, referred hereafter to as "impact of correction" (IOC), is computed as the weighted sum of relevant dose-volume metrics differences between S_d and S_0 (section 2.2.4). The chosen metrics are typically used in our clinic to evaluate the quality of plans for prostate cancer patients. The IOC associated with each S_d is thus a scalar value reflecting the importance of considering this scenario as a relevant correction or not. In the fourth step, each IOC is spread uniformly back over the corresponding deformed volume of S_d, leading to a 3D map where each voxel in the deformed volume is equal to the IOC. Then, all IOCs associated to the differently deformed volumes are averaged locally in each voxel to get the complete DIVE-map (section 2.2.4), highlighting the buildup of strong IOCs, or alternatively the lack thereof.

DIVE-maps for bladder and rectum are independently generated, i.e., there is no combination of deformation between them. In total, 60 DIVE-maps (30x2 OAR) were generated for the 10 first fractions of 3 randomly chosen patients in our database. An example of generated DIVE maps is depicted in Figure 25.

2.2.2. Deformation scenarios

Our volume deformation process uses the RBF mesh deformation of the PyGeM Python toolbox [234]. By varying the direction and the amplitude of deformation, we generated a total of N = 162 deformation scenarios S_d per organ and per fraction. See Supplementary Materials B. for further details.

2.2.3. DCNN dose prediction model

Our DCNN was a hierarchically densely connected U-Net (HD-UNET) [103, 120, 212], with 9 input channels containing the CBCT scan, the PTV volumes, and seven OARs (the anal canal, bladder, rectum, left femoral head, right femoral head, bowel, and penile bulb). Notice that when predicting the dose for the initial S_0 and deformation scenarios S_d of bladder or rectum, the corresponding input channel contained the initially auto-segmented or deformed volume of the organ, respectively. Notice also that PTV of prostate and seminal vesicles are rigidly registered from the initial CT to the CBCT since, in the Ethos workflow, target volumes for daily CBCT are not yet available when correcting Ethos auto-segmented OARs.

The model was trained with data from 180 adaptive fractions of 9 patients (9x20) randomly chosen among the 12 initial ones. See Supplementary Materials B. for more details.

2.2.4. Dosimetric Impact Of Correction value (IOC) computation and aggregation

Each of the *N* deformation scenarios of a given volume is characterized by an IOC scalar value. The IOC formula consists in a weighted sum of *M* dose-volume metrics differences between S_0 and S_d (M=11, see Table 10). Notice that metrics for deformed scenarios are extracted from the predicted dose with our DCNN (section 2.2.3). The weights for each metric difference, α_i, are the inverse of the mean standard deviation of the evaluated metrics on the 20 fractions *Fx* of each 9 training patients p:

$$IOC = \sum_{i=1}^{M} \alpha_i \left| Metric_{i,S0} - Metric_{i,Sd} \right| \quad (8)$$

$$\alpha_i = \frac{1}{\bar{\sigma}_i} \quad (9)$$

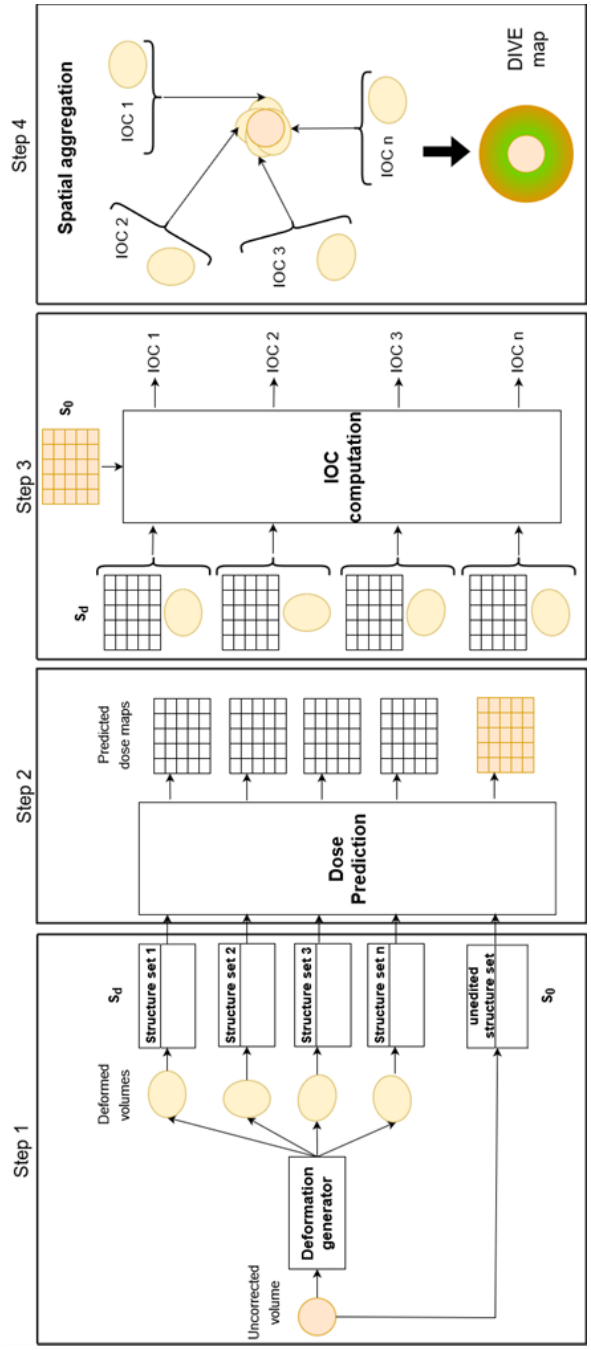
with

$$\bar{\sigma}_i = \frac{\sum_{j=1}^{P} \sigma_{j,i}}{P} \quad (10)$$

$$\sigma_{j,i} = \sqrt{\frac{1}{Fx} \sum_{k=1}^{Fx} \left(Metric_{i,j,k} - \overline{Metric}_{i,j} \right)^2} \quad (11)$$

The idea behind taking the inverse of the standard deviation of each metric is to normalize each metric difference according to their sensitivity to volume changes. Observing a change in an unsteady metric (e.g. the V30Gy to rectum which, in our case, has a mean standard deviation of 2.7%) could be considered as quite common in the database and should, therefore, less influence the weighted sum. Conversely, observing a change in a stable metric would be less common (e.g. the V60Gy to rectum which, in our case, has a mean standard deviation of 0.6%) and should consequently have a more substantial impact on the sum. The idea is not necessarily to penalize high variance metrics but rather to ease the detection of dose changes in low variance metrics that are usually more clinically important. In addition, Ethos optimizer further accentuates the variability of less important metrics. Indeed, objectives of lower optimization priority in Ethos IOE are only optimized once the higher priority objectives are met. Besides, IOE will mostly try to meet these goals and will not optimize much further than this [235]. Consequently, weighted metric differences might better reflect their clinical importance, rather than the sensitivity to volume changes.

The computed IOC value for each S_d is then assigned uniformly to all voxels inside the mask of the deformed volume of interest to form a 3D map. The final DIVE-map results from averaging out these 3D IOC maps over all S_d starting from the smallest deformation amplitude *A*, then locking the concerned voxels, and continuing iteratively to the next deformation amplitudes, like a tree grows bigger wood rings every year (step 4 in Figure 24). For more information, see Supplementary Materials B.



DIVE-map generation process for a given auto-segmented volume. Uncorrected = raw auto-segmented volume, without any modifications, S_0. IOC = "Impact of correction". Step 1: deformation of the considered uncorrected volume, incorporation of the deformed volume to the structure set (other volumes are not deformed), resulting in S_d scenarios. Step 2: Dose is predicted for each deformation scenario as well as the S_0 scenario. Step 3: "Impact of correction" IOC scalars are computed as a weighted sum of dose-volume metric differences between S_0 and each Sd. Step 4: IOC and deformed volumes are spatially aggregated around the uncorrected volume to generate the DIVE-map.

Figure 24

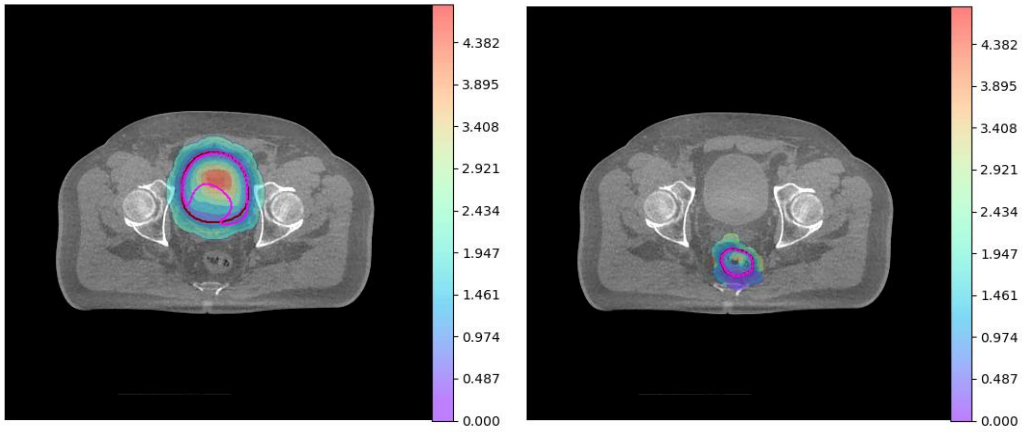


Figure 25: Example of generated DIVE-map for bladder(left) and rectum(right), for which the values are denoted by a colormap ranging from purple to red. maroon = auto-segmented contour; fuchsia= ground truth contour.

2.3. Contours corrections guided by the DIVE-map and Ethos plan generation

The capability of our DIVE-map to indicate dosimetrically relevant regions for correction was assessed with an automatic slice-by-slice correction strategy based on the values of the DIVE-map and the ground truth contours (considered as our reference). We define a "misclassified" voxel when the auto-segmenter did not assign the correct label to this specific voxel: either auto-segmentation considered this voxel as belonging to the nearby OAR while it should not or, conversely, auto-segmentation did not include this voxel into the OAR while it actually belongs to it. In our DIVE-correction strategy, if all misclassified voxels of a given slice had corresponding DIVE-map voxels values (DMi) below a threshold (*t*), no correction was performed on the slice. In contrast, when there was at least one misclassified voxel with DMi above *t*, the entire slice was corrected in the same way as in the ground truth contour. Three thresholds *t* were evaluated to trigger the corrections guided by the DIVE-map, i.e., *t*=1, *t*=1.5, and *t*=2.0. Consequently, it must be noted that no hand-correction are operated: the slices highlighted by the DIVE-map are simply replaced by their homologous ground truth slices.

To evaluate the dosimetric impact of using DIVE-corrected contours, we generated treatment plans for each of the 10 first fractions of the remaining 3 patients not used to train our DCNN dose prediction model. Plans were automatically optimized by the Ethos IOE with the clinical goals in Table 10, the synthetic CT (I.e., the planning CT registered on the daily CBCT) and the custom structure set composed of the DIVE-corrected bladder and rectum, as well as the other OAR and ground truth PTVs. Notice that, in our institution, other volumes than bladder, rectum, and targets are hardly corrected, consequently, those were kept uncorrected.

2.4. Results analysis

To evaluate the added-value of our DIVE-map we present two types of results: dosimetric values and relative number of corrected voxels. On the one hand, we analyze the clinical goals (i.e., dose-volume metrics, Table 10) for the treatment plans generated with DIVE-corrected and ground truth volumes. We specifically focus on high dose metrics, i.e., V60Gy for OAR and D95% to PTV_prostate, since they are of the highest relevance for the plan clinical quality. The goal here is to demonstrate that the dosimetric quality of the plan for DIVE-corrected volumes is the same as the one for ground truth volumes or that deviations are small and clinically not relevant. Consequently, concerning V60Gy metrics, we will consider as a reasonable deviation, a difference of less than +2%. On the other hand, we report the relative number of corrected voxels for auto-segmented rectum and bladder according to the value of *t*. In this context, a relative correction of 100% indicates that all voxels that were corrected to obtained the ground truth contour were deemed necessary when looking at the DIVE-map. In contrast, a value of 0% means that no corrected voxels are actually necessary, consequently, the contour remains fully uncorrected.

3. Results

Dosimetric differences of plans generated with ground truth versus DIVE-corrected and auto-segmented (without any correction) bladder and rectum are illustrated in Figure 26, for OAR V60Gy and PTV_prostate D95%. Using auto-segmented contours led to an OAR V60Gy difference above +2% for 13 fractions (12 fractions showed a difference of V60Gy to rectum greater than 2%, 7 fractions showed a difference of V60Gy to bladder greater than 2% and 4 fractions showed a difference greater than 2% for the V60Gy to anal canal), while this was reduced

to 6 fractions for the DIVE-corrected contours with threshold *t*=2.0. The best reduction of correction able to preserve the dosimetric quality was found with *t*=1.5, since this is the value allowing the largest contour correction reduction while keeping all V60Gy differences below 2% difference for the considered OAR. The lowest difference for D95% on PTV_prostate was -0.5%, which still ensures an acceptable target coverage of 97.7% of the prescription.

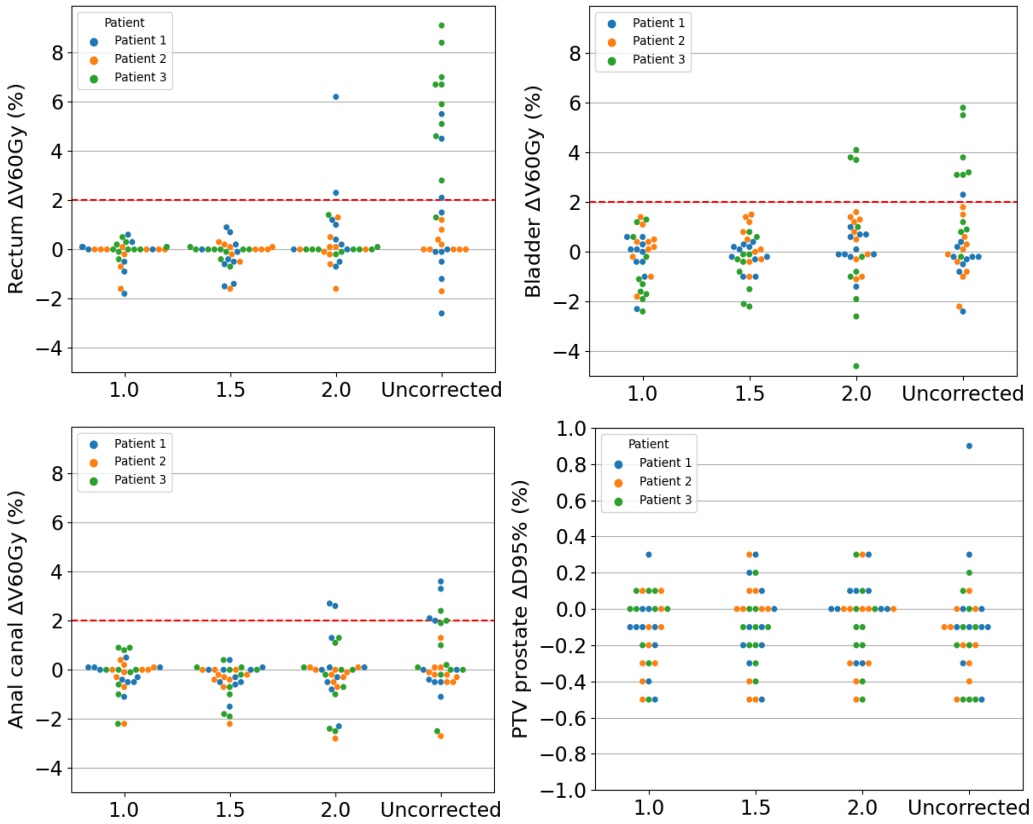


Figure 26: Dosimetric differences between plans generated with DIVE-corrected (t=1.0, 1.5, and 2.0)/auto-segmented volumes and plans generated with ground truth volumes. Individual fraction values are denoted by a dot. Horizontal dashed line represents the limit of +2% difference. Upper left: rectum V60Gy differences. Upper right: bladder V60Gy differences. Lower left: Anal canal V60Gy differences. Lower right: PTV prostate differences. "1.0", "1.5", and "2.0" denote the t threshold value of the DIVE-map below which no slice correction is triggered. "Uncorrected" corresponds to the auto-segmented case where no corrections are operated on the bladder and rectum contours.

Table 11 reports the OARs mean metric differences between the plans using auto-segmented/DIVE-corrected and ground truth contours. Note that, except for the V60Gy in auto-segmented cases, no mean metric difference exceeds +1%. Concerning the bowel D0.1cm³, although it remains below the 60Gy clinical goal, it presents a large difference range. Such dose metric is naturally unstable due to the tiny volume evaluated by the D0.1cm³ over the large bowel volume.

Table 11: Differences (auto-segmented/DIVE-corrected minus conventionally-corrected) on OAR metrics. Note that the bowel D0.1cm³, despite showing very large range of differences, was significantly below the 60Gy threshold for all plans

Volume	Dose-volume metric	Uncorrected minus conventionally-corrected		DIVE(t=1.5) minus conventionally-corrected	
		Mean	Range	Mean	Range
Rectum	V60Gy (%)	2.3	[-2.6; 9.1]	-0.2	[-1.6; 0.9]
Rectum	V52.8Gy (%)	-0.6	[-2.8; 0.8]	-0.5	[-1.9; 1.1]
Rectum	V30Gy (%)	0.4	[-8.8; 7.3]	0.9	[-7.1; 7.3]
Bladder	V60Gy (%)	0.8	[-2.4; 5.8]	-0.1	[-2.2; 1.5]
Bladder	V50Gy (%)	0.0	[-4.2; 1.8]	0.3	[-2.5; 2.2]
Bladder	V40.8Gy (%)	0.2	[-5.5; 3.1]	0.4	[-3.6; 2.8]
Bowel	D0.1cm³ (Gy)	1.0	[-4.5; 18.9]	-0.9	[-15.2; 16.2]
Bowel	V55Gy (%)	0.0	[0.0; 0.0]	0.0	[0.0; 0.0]
Bowel	V20Gy (cm³)	0.0	[-3.8; 7.3]	0.7	[-3.2; 6.7]
Anal canal	V60Gy (%)	0.3	[-2.7; 3.6]	-0.4	[-2.2; 0.4]
Anal canal	V35Gy (%)	-4.9	[-11.0; 0.2]	-3.5	[-8.5; 2.3]
Anal canal	V20Gy (%)	-7.1	[-17.9; 2.2]	-5.9	[-13.4; 3.0]
Femur_L	V40Gy (%)	0.0	[0.0; 0.0]	0.0	[0.0; 0.0]
Femur_L	V35Gy (%)	-0.1	[-0.8; 0.2]	-0.1	[-0.8; 0.0]
Femur_R	V40Gy (%)	0.0	[0.0; 0.0]	0.0	[0.0; 0.0]
Femur_R	V35Gy (%)	-0.2	[-1.8, 0.1]	-0.2	[-2.5; 0.1]
Penile bulb	V54.1Gy (%)	-1.1	[-5.5; 3.2]	-1.1	[-6.4; 3.6]
Penile bulb	V42.5Gy (%)	-0.7	[-5.2; 9.2]	-0.2	[-5.0; 9.4]

Figure 27 reports the relative number of corrected voxels according to the DIVE-map threshold value. As expected, the higher the threshold *t*, the smaller the number of corrections. We can notice global trends. For the rectum, fractions from patient_1 seem to require the highest amount of corrections, followed by patient_2 and patient_3. Concerning the bladder, fractions from patient_3 are those requiring more corrections, followed by patient_1 and patient_2. Notice that fractions from patient_2 do not require any bladder correction for *t*=1.5 and 2.0. Overall, the number of voxels requiring a correction could be reduced by

more than 50% in 16 and 18 fractions for rectum and bladder, respectively, when using the DIVE-map with t=1.5 (i.e. points below the value of 50% in Figure 27). However, for some fractions, plans generated with auto-segmented contours achieved similar dosimetric quality as the ground truth contours. This was the case for 6 fractions in patient_1, all fractions in patient_2 and one fraction in patient_3. For these fractions, the DIVE-corrected rectum with *t*=1.5 required more than 50% of corrections for 5 out of 6 fractions of patient_1, 4 out of 10 fractions of patient_2 and no fraction in patient_3. For the bladder, it resulted in one out of 6 fractions for patient_1, and no fractions for patient_2 and patient_3. Detailed correction values can be found in appendix E.

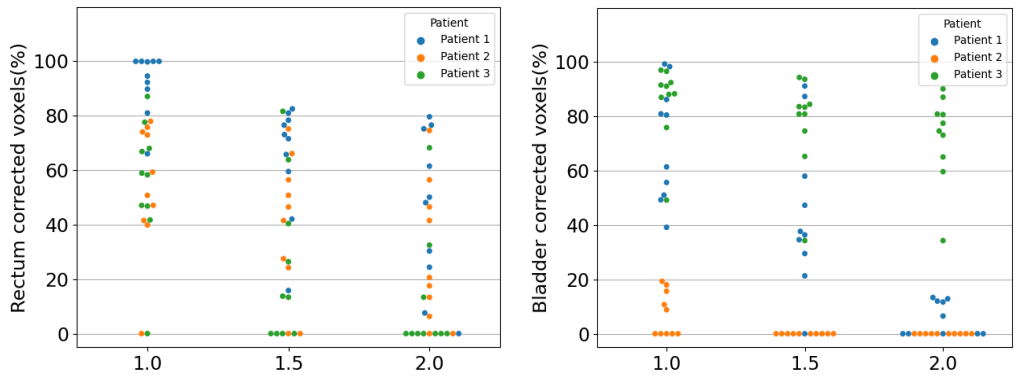


Figure 27: Relative number of corrected voxels for each fraction according to each DIVE-map threshold value. A value of 100% means that all the corrected voxels from ground truth contour are deemed necessary, 0% means that no correction is done, and the corresponding volume remains fully uncorrected. Left: Rectum relative corrected voxels. Right: Bladder relative corrected voxels. Notice the increasing number of fractions requiring no correction at all as the correction threshold increases.

4. Discussion

This work presents a "heatmap" tool to assist clinicians in correcting auto-segmented contours, by indicating dosimetrically relevant regions when it is overlaid on the (re-)planning image. We coined it the DIVE-map, standing for dosimetrically informed volume edition. We evaluated it on auto-segmented bladder and rectum from 3 prostate cancer patients (10 fractions each) receiving adaptive radiotherapy with ETHOS (Varian).

It was first shown that the use of auto-segmented rectum and bladder volumes (with ETHOS auto-segmenter) without any correction would result in 13 plans out

of 30 where OAR V60Gy values would strongly exceed their homologous values achieved with conventionally-corrected contours. This highlighted the dosimetric importance of correcting auto-segmented volumes. Then, using the DIVE-map with a reasonable correction threshold *t* of 1.5, we demonstrated the preservation of the dosimetric quality while reducing the amount of corrections. In particular, there were 6 and 11 fractions where, following the DIVE-map indications, no correction had to be applied to the rectum and bladder, respectively. Moreover, the amount of corrections was reduced by more than 50% in 16 and 18 fractions for rectum and bladder, respectively. However, for 9 rectums and 1 bladder, the DIVE-map marked as relevant more than 50% of conventional corrections, while auto-segmented contours presented the same dosimetric quality. This highlighted a good sensitivity of the necessary corrections when using *t*=1.5, i.e. the slices leading to important dosimetric changes were corrected, while the specificity was more limited, i.e. the corrections in slices leading to no dosimetric changes were not all discarded, especially for the rectum. This could hint that some of the used rectum deformations are not fully relevant. Investigation of different rectum deformations to enhance the correction specificity could be a next investigation step. This could also be a result of a limited precision of the dose prediction on some deformation scenarios (see appendix E).

Concerning the DIVE-threshold *t*, our results suggested to use a value between 1.5 and 2. This value might not generalize well to other locations and specific treatments, since it is influenced by the quality of the predicted doses, the quality of the auto-segmenter, or the dose-volume metrics used to compute IOC factors. Therefore, anyone wishing to use this tool in their own clinical conditions will have to go through a calibration step to find the optimal *t* value, which can take some time and requires representative data. This calibration requirement stems from the limited clinical interpretability of the dimensionless IOC values underlying the DIVE maps. Future investigations could focus on expressing IOC and *t* in more clinically meaningful units, potentially eliminating the need for site-specific calibration. A potential improvement to IOC to make it a more consistent value across different implementation conditions could be its normalization by the number of objectives.

This lack of interpretability highlights also a limitation of this study. In the absence of a clear interpretation for such IOC values and *t* parameter, we evaluated the tool over 30 fractions. However, it remains unclear what constitutes an adequate database to validate the DIVE-map. Is a total of 30 fractions sufficient to the validation of the threshold *t*? Should we use the entirety of the treatment fractions rather than half (i.e. 20 instead of 10 in this case)? Should we use the same number of fractions but from more diverse patients, i.e. more than 3 patients? All these questions remain to be answered in future investigations to ensure DIVE reliability and human trust.

A third hitch concerns the DIVE-map generation time. It took around one hour to generate a single DIVE-map, using a purely sequential process, where each deformation scenario (0.5 seconds) and predicted dose (10-20 seconds) is computed one after the other. Such generation timing is incompatible with online adaptive treatments. Therefore, parallelization of these processes and reduction of the number of deformation scenarios (e.g., by keeping only the most relevant ones) could be a next step towards the clinical use of such a tool. Alternatively, it could be feasible to build a database of DIVE-maps and then train a DCNN model to predict them, with a workflow and timing similar to the dose prediction process.

To clarify an important aspect of the methodology, it should be noted that the DIVE-map generation process intentionally employs non-realistic deformations. This is not a limitation but rather an inherent design choice aligned with the tool purpose. The primary objective is to evaluate the dosimetric impact of organ presence across the surrounding spatial regions. This does not necessitate to simulate physiologically plausible organ movements. What matters is the exhaustive exploration of spatial locations at increasing distances from the original contour, with sufficient directional sampling at each range increment. From this perspective, incorporating realistic deformation behavior is not relevant to the fundamental purpose of the tool. However, the use of realistic deformations could potentially enhance dose prediction accuracy in cases where the prediction model has been exclusively trained on realistic anatomical data, as it is the case in our implementation.

Another potential weakness of our approach is that the DIVE-maps of bladder and rectum are generated independently, whereas both volumes are likely to be correlated in terms of dose optimization. Ideally, this would require making deformation combinations with both volumes, which would, however, add a lot of scenarios and further slowdown the generation process. In our case, with only two volumes and globally good segmentation quality, DIVE-maps gave sufficient information to perform reliable corrections. However, for other cases with more volumes to correct or lower auto-contouring quality, performance remains to be evaluated.

This study did not evaluate the time that a clinician could spare by using the DIVE-map, but the tool is designed to be used for that purpose and that is the reason why its performance was evaluated in the same way as humans would probably use it, namely, one slice after the other. Future work could investigate the joint use of an uncertainty map [230], indicating the most probable regions of contouring error for DCNN auto-segmentation, with the DIVE-map. This would allow us to highlight the regions where contouring errors are more probable instead of the entire space where the deformations were evaluated.

On a side note, the time for manual corrections might not be the only concern when using automatic segmentation in adaptive conditions. Some studies have demonstrated the presence of automation bias, which results in noticeable differences between auto-segmented contours followed by manual corrections and fully manually segmented contours. While it may have a beneficial impact by decreasing the inter-observer contouring variability [117, 118], it might also result in over-reliance of clinicians on automated contouring [118, 218], jeopardizing the contouring and dosimetric quality. Such an excessive reliance on automation has already been observed in other medical fields [236-238]. Consequently, DIVE-maps might also help to prevent possible omission of dosimetrically important corrections and again ensuring plan clinical quality.

In conclusion, with the limited information from auto-segmented contours and approximative position of target, the DIVE-map was able to discard most of irrelevant corrections. Even if some enhancements of the approach in terms of correction specificity and generation time seem possible, such tool could have great future potential in guiding physicians during manual contour corrections.

Chapter 4: Online adaptation for Proton Arc Therapy

This chapter is published as:

Towards faster plan adaptation for proton arc therapy using initial treatment plan information

Benjamin Roberfroid, Margerie Huet Dastarac, Elena Borderías Villarroel, Rodin Koffeing, John A. Lee, Ana M. Barragán-Montero and Edmond Sterpin. Physics and Imaging in Radiation Oncology, 100705.

1. Introduction

Proton arc therapy (PAT) is an emerging cancer treatment modality that aims at delivering a proton beam in a rotational pattern around the patient, in contrast to the traditional intensity modulated proton therapy (IMPT), which delivers beams at fixed angles. Conceptually, PAT has more degrees of freedom than IMPT and has thus the potential to achieve better dose distributions [239-243]. However, several challenges remain to be addressed to harness its full potential.

Like IMPT, PAT remains sensitive to treatment delivery uncertainties and anatomical changes [244]. While robust optimization is commonly used during plan optimization [24, 245] to account for these, it entails larger volumes irradiated at higher doses and may not cover all possible anatomical changes. Given this limitation, proton online adaptation (OA) has until now promised to be the best option to reduce treatment uncertainties while improving dose conformity [191, 246, 247]. Enabling OA for proton therapy is thus often considered as a primordial objective to realize the full potential of this treatment modality.

Unfortunately, the challenge for PAT is to determine optimal energy layers per beam angle. This is needed to achieve optimal dosimetric quality while maintaining a reasonable treatment time [69, 70]. Excessive energy switches, particularly switch-ups (those are much slower due to magnetic hysteresis effect in the delivery system), might indeed result in significantly prolonged treatment delivery that would hardly be compatible with clinical conditions. Different research groups have tackled the problem from different perspectives, resulting in various optimization approaches [70, 248-253]. These methods exhibit differences in computation times, final energy layer patterns, and dose distribution. Although some methods are faster than others, total PAT optimization times (including the energy layer selection process) are generally slower than conventional IMPT, thus representing a significant obstacle for OA.

In this context, implementing a dedicated online adaptive workflow for PAT, by speeding up plan re-optimization, would ensure that PAT would not lag behind IMPT for situations requiring OA. Currently for IMPT, several solutions have already been proposed by different research groups. Some examples include making use of analytical algorithms to speed up optimization of the dose on the

computed tomography (CT) of the day [188, 189], bypassing the manual plan optimization loop by restoring the dose distribution of initial CT onto the CT of the day [190, 254, 255], or mimicking a dose predicted by artificial intelligence (AI) on the CT of the day [191]. Alternatively, Botas et al. proposed a workflow where they re-used the initial spot grid onto the image of the day through a deformation field between both images, followed by the tuning of a subset of these spot weights. Eventually, they managed to adapt IMPT plans in a time span compatible with clinical conditions [195]. Our proposed PAT adaption workflow re-uses the key elements of this workflow while considering the specific challenges of PAT.

This study aims to propose a time-efficient OA workflow for PAT that effectively addresses uncertainties and anatomical changes detrimental to proton therapy, enhancing the clinical quality of treatment plans and ensuring a promising future for the PAT treatment modality.

2. Materials and methods

2.1. Patient data for workflow evaluation

The study has been retrospectively conducted on five head and neck cancer patients with nodal involvement. The dose prescription was 70 Gy to the clinical target volume of primary tumor (CTVp) and 54.25 Gy to CTV of the nodes (CTVn), except for the 5th case where a part of the nodal target also receives 70 Gy (CTVn,high). CT images were acquired on an Acquilion CT scanner (Toshiba Medical Systems Corporation, Japan) with a slice thickness of 2mm. For each of these patients, an initial CT (CT1) was acquired before treatment and a second CT (CT2) was acquired several days after treatment started. In each case, CT2 was acquired due to observation of important anatomical changes regarding CT1. Organ at risks (OARs) and target contours were segmented by the physician in charge for both CT1 and CT2. Local ethics committee approval by the CEHF (Comité d'Ethique hospitalo-facultaire) was obtained for the use of all patient data under the agreement "Learning from the past: the MIRO treatment planning database" (2019/16SEP/402).

2.2. Adaptive workflow

The PAT adaptive workflow comprises 4 steps. Steps 1-3 form the geometrical adaptation (left of Figure 28) while Step 4 handles spot weights adaptation (right

of Figure 28). Implementation was done in OpenTPS, a Python-based open-source treatment planning system (TPS) [256].

Step 1 involves registering the planning CT and spot grid to daily anatomy. After deformably registering CT_1 to CT_2, the resulting displacement field transfers spots from CT_1 to CT_2. To do so, Bragg peak positions on CT_1 are retrieved and used as surrogate for the spatial position of the *i* spots ($x_{i,1}$, $y_{i,1}$, $z_{i,1}$) in DICOM (patient-system) coordinates. The displacement field then maps these positions onto CT_2, i.e., ($x_{i,2}$, $y_{i,2}$, $z_{i,2}$). We used a hexagonal spot grid, and the deformable intensity-based deformable registration tool of the Raystation TPS (Raysearch laboratories, Sweden). The deformation field was then exported to OpenTPS to operate the displacements.

Step 2 converts updated spot positions ($x_{i,2}$, $y_{i,2}$, $z_{i,2}$) to beam-eye-view coordinates ($x^*_{i,2}$, $y^*_{i,2}$, $E_{i,2}$), where x* and y* represent horizontal and vertical positions, and E is the necessary energy to reach the desired depth in patient. The determination of the energies can be achieved by a raytracing process on the new anatomy. However, PAT treatments require careful energy assignment to maintain treatment efficiency and quality. Unfortunately, spot registration between CT_1 and CT_2 typically disrupts initial spot clusters (spots sharing the same energy layer and beam angle on CT_1), as this clustering is not considered during the registration process. Consequently, our method proposes to preserve the clusters by updating the energy on CT_2 while considering the energy layer pattern of CT_1, i.e., spot clusters are preserved and angles where energy switch-ups occur on CT_1 remain the same on CT_2. To do so, for each initial cluster *c* the median energy of updated spots $\widetilde{E_c}$ is computed and forced to be assigned to the spots of the considered cluster. If for a given angle, $\widetilde{E_c}$ value leads to a non-authorized energy switch-up (i.e., an increase of energy between two consecutive angles that is not occurring on the initial energy pattern), then the energy for this specific cluster *c* is instead set at the same energy level than the previous cluster, i.e., $\widetilde{E_c} = \widetilde{E_{c-1}}$. We hypothesized this energy layer preservation would maintain the plan quality while keeping beam parameters similar to the initial plan, facilitating PAT plan adaptation.

Step 3 verifies the dose distribution. The dose is computed in CT_2 with the updated position and energy of the spots as well as their weights from CT_1 plan. At

this step, if the dose is deemed clinically acceptable, the adaptive process can be stopped. Otherwise, the dose contribution of each spot *i* to the total dose (i.e. the so-called beamlets or dose-influences D_i) are then saved for the following step. These D_i are each generated using proton numbers proportional to their CT_1 weights. The total dose was computed using MCsquare, an open-source Monte-Carlo engine [257] integrated in OpenTPS, with $3*10^7$ protons.

Step 4 implements spot weights re-optimization through a two-criteria selection process. First, spots with largest weight (on CT_1) are selected, i.e., the smallest set of spots carrying at least a percentage *p* of the sum of the spots weight is selected. Spots with small initial weights result in beamlets with high dose uncertainty (i.e. they were generated at step 3 with a low number of protons), which are therefore not suitable for weight re-optimization. The second criterion selects spots with substantial impact on the objective function among those meeting the first criterion, using a gradient constraint.

$$\frac{\partial \mathrm{F}}{\partial w_i} > \mathrm{t} \tag{12}$$

where F is the objective function of the optimization problem, w_i is the *i* spot weight and *t* is an arbitrary chosen threshold. Note that discarded spots are not deleted, they still contribute to the total dose distribution similarly to a background dose:

$$TD = d_d + d_s = w_{d,1} D_d + w_{s,updated} D_s \tag{13}$$

with *TD* being the treatment dose, d_d the dose contribution from the discarded spots, d_s the dose contribution from selected spots, $w_{d,1}$ the spots weights of discarded spots (thus equal to those on CT_1 plan), D_d the dose influences matrix of the discarded spots, $w_{s,updated}$ the spots weight of selected spots that were re-optimized, and D_s the dose influences matrix of the selected spots. For our five patients, we selected spots carrying at least 65% of total weight (first criterion) and used t=3 for the gradient constraint (second criterion). These values were determined empirically. While Botas' paper used a 50% weight threshold, we increased it to 65% and added the gradient constraint for dosimetric reasons (see appendix S2 for more details).

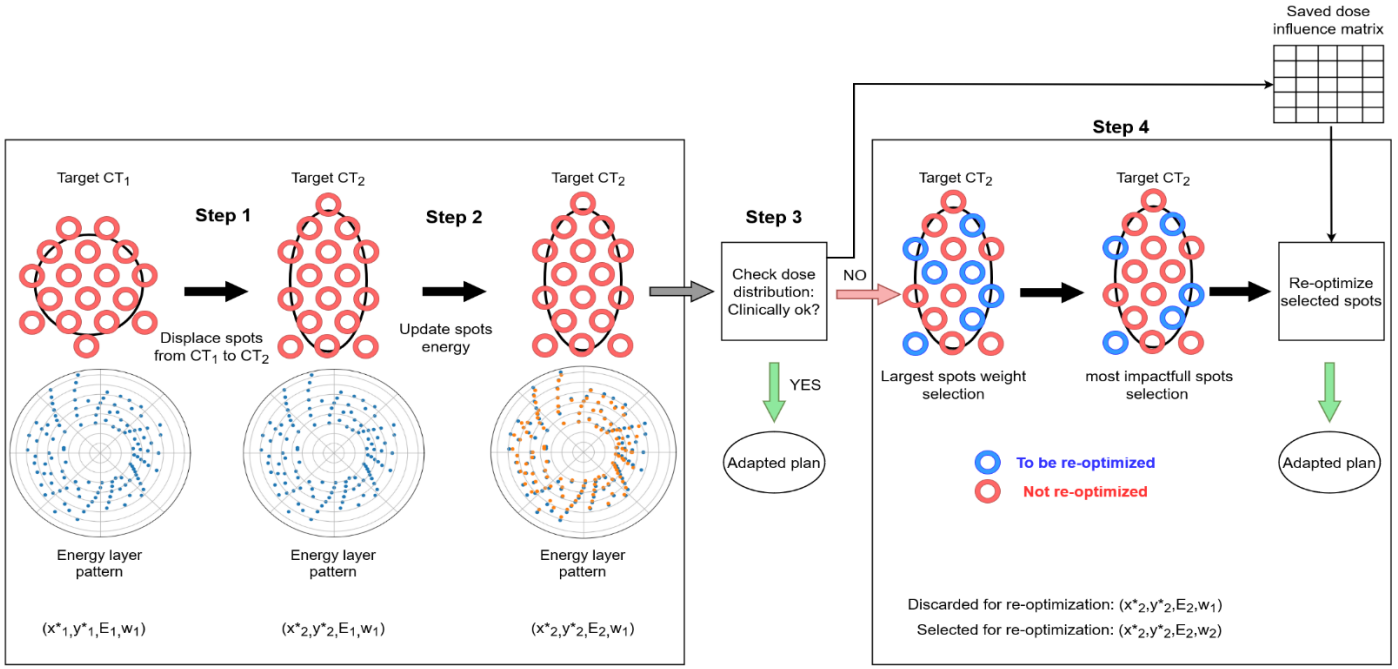


Figure 28: The proposed PAT adaptation workflow. It consists in two different steps: geometric and spot weights. The first step of the geometric adaptation is to displace the initial spots onto the CT of the day; the second is to update the spot energy while ensuring PAT energy layer pattern. At step three, the dose distribution is verified, if its quality is deemed clinical then the adaptation workflow can already stop there. Otherwise, the dose influence matrix resulting from the dose computation is saved for the step 4. The first step of the spot weights adaptation is to select the largest weights for the re-optimization; the second is to refine this first selection by selecting the spots with the largest derivative of the objective function according to spot weights. Eventually, the weights of the selected spots are re-optimized, using the saved dose influence matrix. Black circular shape on the top denotes a schematic representation of the target volume in 2D, the rings denote a schematic representation of the treatment spots for such target. In the middle, dots disposed on a polar plot illustrate a representation of the energy layers for each beam angle, with the radial distance being the energy and the angular position being the beam angle. The blue dots are the initial energy layers while the orange dots are the updated ones. In the bottom, the quadruplet (A,B,E,w) summarizes the state of the spots at each steps. A and B denote the spots position in Beam-eye-view, E denotes the energy of the spots, w denotes the spots weight. Index 1 indicates that the variable preserves its CT_1 value, index 2 indicates that the variable is updated for CT_2.

2.3. Plan parameters

Each plan featured a single 360-degree arc with control points at each degree. Energy layers were defined with the ELSA algorithm [4] integrated in the RayStation TPS version 11B. Initial optimization constraints were applied to CT_2 plans and adjusted when needed. Optimization parameters varied per patient, with constraints and priority weights determined case-by-case. Each adapted plan allowed 500 maximum iterations. All plans incorporated worst-case robust optimization using 2.6% range error and [4.0,4.0,4.0] mm setup error, based on previous studies [192, 207]. The robust constraints concerned the maximal and minimal dose to the CTV_p, CTV_n, $CTV_{n,high}$ as well as the max dose to the spinal cord (SC). Physicians in charge provided contours for both CT_1 and CT_2.

2.4. Plan comparison

PAT plans were generated for CT_1 and CT_2. Plans on CT_2 were generated with two approaches: "*reference*", optimized from the ground up; and "*smart-adapted*" generated following the proposed adaptive workflow. The reference plans used the same plan parameters as the smart-adapted plans, except that the spots and energy layers were initialized from the ground up and the dose influences matrices were generated with 10^4 protons for each spot.

To evaluate the proposed adaptive workflow, the plan quality for the geometric adaptation, i.e., adaption of the spot position and energy only (left box in Figure 28), on critical constraints was first investigated: D_{98} CTV_p, D_{98} CTV_n, and D_{max} on SC on nominal case, for the reference, the smart-adapted and the no-adapted plan computed on CT_2 were reported. For these critical clinical goals, we aimed to achieve D_{98}>95% of the prescribed dose (66.5 Gy and 51.54 Gy for CTV_p and CTV_n respectively) and D_{max}<44 Gy to SC. These metrics were evaluated following complete plan adaptation (comprising both geometric adaptation and spot weight refinement) (both boxes in Figure 28), on the nominal case and the worst-case scenario. Mean dose to OARs was also reported on the nominal case. For other objectives, the reference plans aimed at achieving the lowest mean dose to OARs and the lowest D_2 for CTV_p.

A timing comparison between smart-adaptation and reference methods is also provided. Spots and energy initialization, computation time to determine the dose influence matrix, and dose optimization time were reported.

2.5. Hardware

The registration process used a computer with an Intel Xeon Gold 6244, 96 GB RAM, and Nvidia RTX A600 GPU. Dose influence matrix generation, adaptive workflow, and optimization used a system with Tesla A100 80GB GPU, 128 Intel Xeon Gold CPUs, and 125 GB RAM.

3. Results

3.1. Geometrical adaption

Figure 29 shows an example of energy layer adaptation after workflow steps 1 and 2, demonstrating preserved angular positions of energy switch-ups. Table 12 indicates target coverage and SC maximum dose for all patients when using plan computed for CT_1 on CT_2 (No adaption) and when using the geometric plan adaption. For all cases the $D_{98\%}$ is noticeably below 66.5 Gy, i.e., 95% of the prescription dose. Geometric adaption slightly improves the target coverage and SC maximum dose for some patients. However, it was not sufficient to recover clinical quality.

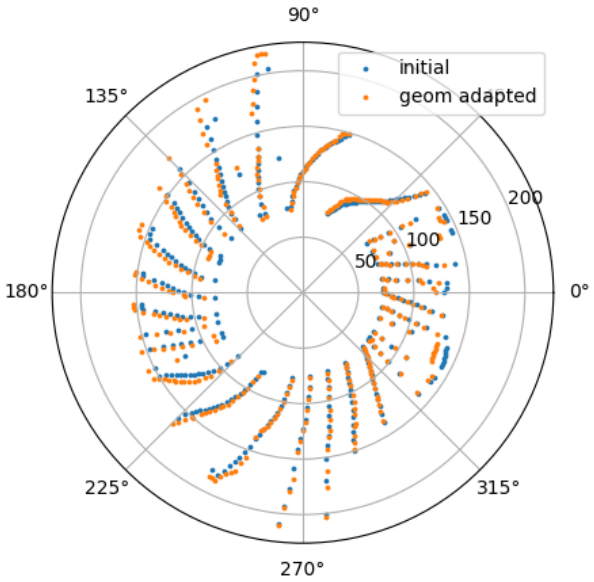


Figure 29: Example of energy layer pattern update. Radial position denotes the energy in MeV, angular position denotes the beam angle in degree. "initial" = initial arc energy layer pattern on CT1; "geom adapted" = geometrically adapted energy layer pattern on CT2.

Table 12: Comparison between no adaptation, geometric adaptation (without spot weights adaptation) and spot weights adaptation (with geometric adaptation first) on CT2 for nominal case for D98% of CTVp, D98% of CTVn and the maximal dose received (Dmax) by spinal cord (SC) the 5 patients.

		P1	P2	P3	P4	P5
CTVp D_{98} (Gy)	Clinical goal			>66.5Gy		
	No adaptation	29.1	53.0	62.7	56.4	65.2
	Geometric	30.4	56.1	64.4	57.5	62.4
	Spot weights	68.2	67.9	68.0	68.4	68.3
CTVn D_{98} (Gy)	Clinical goal			>51.54Gy		
	No adaptation	24.9	39.4	40.9	45.8	47.7
	Geometric	30.4	41.7	41.5	44.0	42.1
	Spot weights	52.8	53.0	53.4	53.0	53.2
SC D_{max} (Gy)	Clinical goal			<44Gy		
	No adaptation	66.6	41.9	30.3	22.6	24.7
	Geometric	62.1	36.4	29.3	24.9	24.1
	Spot weights	32.8	27.3	21.9	26.5	22.6

3.2. Spot weights adaptation

Table 12 also includes details for target coverage and SC maximum dose of the spot weights adaptation. The constraints on the targets and on the SC for the nominal case were largely met. Target coverage objectives and SC Dmax objectives were also satisfied for the reference plans. Table 13 displays metric differences between plans that were adapted and their reference. We can notice a slight median decrease for $D_{98\%}$ of 0.21 Gy for CTV_p and 0.44 Gy for CTV_n, while still being largely above the clinical constraint (i.e., 64.4 Gy and 51.54 Gy, respectively) for all plans. D_{max} difference for SC was globally a bit lower for the smart-adapted plans even if there is one smart-adapted case exhibiting an excess of 0.49 Gy compared to the reference plan. Mean dose differences to OARs were globally low: the highest median difference over all patients was found for the mandible and was of 1.86 Gy.

Table 13: Dose volume-metrics difference between smart-adapted plans and reference plans on CT2 for nominal case. SC: spinal cord

Volume	**Metric**	Smart **adaptation minus Reference**	
		Mean (Gy)	**Range (Gy)**
CTV_p	D_{98}	-0.2	[-0.6, 0.1]
CTV_p	D_2	0.2	[-0.1, 0.7]
CTV_n	D_{98}	-0.4	[-0.5, -0.1]
$CTV_{n,high}$	D_{98}	-0.4	/
$CTV_{n,high}$	D_2	0.0	/
SC	D_{max}	-0.5	[-5.0, 1.2]
SC	D_{mean}	0.1	[-1.6, 0.8]
Mandible	D_{mean}	1.9	[-8.7, 4.4]
Parotid_L	D_{mean}	0.9	[-3.6, 1.9]
Parotid_R	D_{mean}	0.1	[-0.8, 2.7]
Thyroid	D_{mean}	0.6	[-2.7, 1.6]
Esophagus	D_{mean}	-2.2	[-3.9, 2.1]
Larynx	D_{mean}	-0.1	[-1.3, 2.5]
Oral Cavity	D_{mean}	0.7	[-4.1, 1.2]

Concerning dose-volume differences for the worst-case scenario, results are displayed in Table 14. The trend is similar to the nominal case: a slight decrease in CTVp and CTVn coverage of 0.44 and 0.56 Gy, respectively; while CTVp D2% was 0.24 Gy higher, thus resulting in a lower dose homogeneity. As for the nominal case, Dmax to SC was about 1.47 Gy lower for smart-adapted plans than for reference plans.

Table 14: Critical dose volume-metrics difference between smart-adapted plans and reference plans on CT2 for worst case. SC : spinal cord

Volume	Metric	Smart **adaptation minus Reference**	
		Mean (Gy)	**Range [min, max](Gy)**
CTV_p	D_{98}	-0.4	[-0.8, -0.2]
CTV_p	D_2	0.2	[-0.1, 1.0]
CTV_n	D_{98}	-0.6	[-0.7, 0.0]
$CTV_{n,high}$	D_{98}	-0.3	/
CTV_{n_high}	D_2	0.4	/
SC	D_{max}	-1.5	[-3.1, 1.3]
Body	D_{1cc}	0.7	[-0.4, 1.9]

3.3. Timing analysis

Section “Spots initialization” in Table 15 displays the timings of the different steps. During spot initialization (steps 1 and 2), our adaptive workflow was 7.7 seconds slower than the reference method of generating a new spot grid and optimizing energy layers with ELSA, resulting in a relative speed factor of 0.76. Section “Dose influence computation” in Table 15 reports computation times to generate the dose influence matrix (step 3). The proposed adaptative workflow was on average 3.55 faster than the reference one. Concerning the optimization part in the section “spots weight optimization” (step 4 of the adaptive workflow), the smart-adaptation constantly re-optimized less than 50% of the total spots of CT_1 onto CT_2. This step only considers the required time to optimize the spot weights. The adaptation workflow was on average 2.32 times faster than the reference one. Overall, the adaptive workflow was 3.36 times faster than the reference method.

Table 15: Timing comparison between reference (Ref) and adaptation workflow (Smart) for the 5 patients. The "average acceleration of the step" rows are computed as average of reference divided by average of adaptation workflow. The "Average acceleration of plan adaption" row on the bottom is computed by dividing the sum of the time components of the reference workflow by the sum of the different time components of the smart-adaptive workflow. Percentage of re-optimized spots are also provided.

Step	Method	Details	**Patient 1**	**Patient 2**	**Patient 3**	**Patient 4**	**Patient 5**
Spots initialization	**Ref (s)**	**Spot + energy layer determination**	47	36	37	38	34
	SMART (s)	**Registration determination Time**	26	47	16	23	12
		Spots displacement and energy update	26	24	27	24	26
	Average acceleration of the step				0.76		
Dose influence computation	**Ref (s)**		15523	28538	15371	6253	20938
	SMART (s)		4975	6470	5083	2097	5001
	Average acceleration of the step				3.55		
Spots weight optimization	**Ref (s)**		2132	1308	1496	2817	1350
	SMART (s)		860	928	752	628	1080
	Average acceleration of the step				2.32		
	Re-optimized spots (%)		46.6	46.8	47.4	48.0	49.5
Overall average acceleration of the whole adaptation process					3.36		

4. Discussion

We implemented a "smart" plan adaptation workflow for PAT and evaluated it on 5 head and neck cancer patients. The workflow features a geometric and a spot weight step. It is inspired by the work of Botas et al. for IMPT [195] while making some adaptation to accommodate it to PAT modality, as well as adding a novel criteria to select the most relevant spots for re-optimization.

Geometric adaptation showed modest improvements in target coverage for some patients and failed to achieve clinically acceptable quality. Target coverage, and sometimes the maximum dose to the SC, did not meet the clinical constraints. In some cases (patients 4 and 5) geometric adaptation seemed to be worse than no adaptation. This followed a similar trend observed in Botas' paper and highlighted the need for re-optimization of spot weights for complex cases. The limited success of geometric adaptation might stem from our patient selection, as these cases initially required replanning due to substantial anatomical changes. Such changes might have exceeded the capabilities of the geometric adaptation.

The spot weights adaptation step, which re-optimized less than 50% of the total spots, produced treatment plans comparable to their reference counterparts created from the ground up. While the smart-adapted plans exhibited a minor reduction in dose homogeneity, this did not meaningfully impact their clinical acceptability.

Although marginally slower in initial spot and energy determination, the smart adaptation method achieved significant speedups in subsequent steps i.e., dose influence matrix computation and dose optimization. Smart adaptation method proved to be 3.55 times faster than classical optimization from the ground up (reference) for the computation of the dose influence matrix and 2.32 times faster for the dose optimization using the optimization constraint from CT_1. Such acceleration for beamlets generation stems from simulating fewer protons by leveraging prior weight information. The dose optimization acceleration results from both the initialization of the spot weights that lead the optimization process to start from a better initial solution and the reduction of total number of spot weights to optimize thus resulting in a less computationally intensive process.

However, among various approaches for energy layer determination, some could be significantly slower than the ELSA method used here [70, 249]. These findings suggest the proposed method may have even greater impact when applied to these slower techniques by allowing quick updates to initial angle/energy patterns rather than re-launching a new determination process. Future tests should evaluate the time savings for these other energy layer algorithms.

An interesting perspective would be evaluating this workflow on commercial TPS where some operations are generally faster than in OpenTPS. Currently, the time

span to produce a smart-adapted plan remains hardly compatible with clinical conditions. Notably, only a few matrix operations are GPU-implemented in this TPS, while in the original Botas' paper, most workflow operations were GPU-implemented, resulting in a time-scale compatible with OA. Therefore, evaluating the proposed workflow on different systems, would be instructive. However, in contrary to OpenTPS, certain parameters tuned in this workflow may not be readily accessible within these commercial systems. For example, saving spot dose contributions and reusing them for subsequent calculations, as well as freezing optimization of discarded spot weights, could be challenging due to their lack of accessibility to users.

Regarding optimization constraints, weight updates to target coverage and maximum dose to the SC were necessary. This adjustment loop is not perfectly compatible with OA. In addition, reusing constraints from CT_1 may theoretically be suboptimal for CT_2. However, in recent years, several techniques of AI constraint prediction and automatic planning have proven their efficiency [119, 202, 235]. Combining such predictions with this adaptive workflow would likely be valuable. Future work should investigate compatibility with updated constraints that are more specific to the anatomy of the day, i.e., reduce dose to OARs on the CT of the day when possible by the use of stricter constraints.

In conclusion, while the proposed workflow achieved comparable dose-volume metrics to full re-optimization with substantial time savings, further development is needed to achieve adaptation times compatible with online adaptation.

Chapter 5: Online adaptation for conformal FLASH

This chapter is submitted as:

Anticipating Potential Bottlenecks in Adaptive Proton FLASH Therapy: A Ridge Filter Reuse Strategy

Benjamin Roberfroid, Macarena Chocan Vera, Camille Draguet, John A. Lee, Ana M. Barragán-Montero and Edmond Sterpin. Currently, in revision in *Physics in Medicine and Biology*.

1. Introduction

Since its rediscovery by Favaudon et al. [32], the FLASH effect, referring to a sparing effect of the healthy tissues without loss of tumour control at high dose-rate, has provoked a vast new wave of research. Lots of these studies aim at understanding the effect itself: why it occurs, how to quantify its strength or what are the ideal conditions to trigger it [258-263]. Complimentary to radiobiological investigations, feasibility in clinical conditions with the current hardware is also explored [264-269]. Specifically, clinical feasibility for proton-triggered FLASH effect is mostly studied since current proton therapy installations may already be able to reach the necessary dose rate to trigger the FLASH effect [270].

In proton therapy, along the years, pencil beam scanning (PBS) method imposed itself as current state of the art due to excellent conformity achieved with it. However, using PBS with energy intensity modulation through beam energy layer switching prevents high dose-rates due to the high energy layer switching time. Consequently, PBS is envisaged for FLASH with a monoenergetic beam and two delivery modes: transmission and conformal PBS [269]. Transmission treatments consist in using high energy beam leading to Bragg peaks outside the patient but allow easy optimization of high dose rate treatment plans [271]. This modality avoids the need for patient specific device and creates plans that are naturally robust to treatment uncertainties, since only the proximal part of the Bragg peak curve is used. However, without the contribution of the Bragg Peak distal dose fall-off, the potential for protons improved target dose conformity is reduced. On the other side, conformal treatments offer an excellent conformity by placing the treatment spots within target range with the help of a patient-specific ridge filter (RF). Such device allows to obtain a spread-out Bragg peak within the target area even with a monoenergetic beam [272-274]. Unfortunately, this RF needs to be 3D printed to fit patient anatomy which is time consuming, i.e. one to several hours [275]. Printing a new RF would therefore not be feasible every day in clinical conditions, hampering daily delivery of conformal FLASH fraction. On the other side, delivering the initial conformal plan as it on the anatomy of the day could lead to a significant loss of dosimetric quality due to anatomical changes. Indeed, such conformal method is highly sensitive to uncertainties and anatomical changes. Consequently, it appears that finding a way to enable online adaptation of conformal FLASH plan would be of highest interest for the proton FLASH

therapy community and could definitively make conformal beam superior to transmission beam.

This paper addresses this problem by re-using the initial RF for each treatment fraction, while updating the treatment spot weights to the anatomy of the day. This approach drastically reduces the need for daily reprinting of the RF and allows partial online adaptation of FLASH conformal treatments. The other steps required for an online adaptive workflow, such as OAR contouring or treatment plan generation, are already automated in the current state of the art and could be directly combined with the proposed approach in clinical routine to enable an efficient workflow for FLASH conformal therapy. [191, 235, 276-278].

2. Materials and methods

This section describes the detailed workflow used to adapt FLASH plans and the data used for its evaluation.

2.1. Data

This study used an existing retrospective database of 6 head and neck cancer patients. CT images were acquired on a Toshiba Acquilion CT scanner with a slice thickness of 2mm. For each of these patients, an initial CT (CT_1) was acquired before treatment start and a second CT (CT_2) was acquired several days after treatment start. In each case, CT_2 was acquired due to observation of important anatomical changes with respect to CT_1. Organ at risks (OARs) and target contours were segmented by the physician in charge for both CT_1 and CT_2.

Local ethics committee approval by the CEHF (Comité d'Ethique hospitalo-facultaire) was obtained for the use of all patient data under the agreement "Learning from the past: the MIRO treatment planning database" (2019/16SEP/402).

2.2. Workflow for conformal FLASH planning

The conformal FLASH plans were all generated on an experimental version of the RayStation treatment planning system (TPS) (RaySearch, Stockholm, Sweden). The planning workflow consists first in choosing a beam angle per treatment fraction. For this considered angle, a single field uniform dose (SFUD) beam is generated. At first, it consists in a poly-energetic PBS plan where the spot weights are optimized a first time. Then, this poly-energetic plan is converted to a

monoenergetic RF plan through an algorithm incorporated in the TPS. This RF is thus defined to best suits the anatomy of the considered CT image and to reproduce at best the initial poly-energetic plan. The treatment spot weights are then optimized a second time with this RF.

2.3. Plan parameters

All the conformal FLASH plans generated for this study used the same parameters, described hereafter. The chosen prescription dose per fraction is 9Gy, as current literature indicates a minimal dose per fraction threshold varying from 4 to 10Gy to observe the FLASH effect [260]. The treatment plans are designed for the Proteus Plus system (IBA sa, Louvain-la-Neuve, Belgium). The clinical goals per fraction aim at reaching D_{98} >= 95% of 9Gy, D_2 as low as possible, D_{max} to spinal cord (SC) lower than 5 Gy. This constraint on spinal cord is derived from the recommendation of *Diez et al.* [279] when considering a fractionation scheme of 5 fractions. The mean dose to OARs was decreased as much as possible while trying to meet first the previously enounced objectives. The optimization constraints used in this study are the same for all plans while their optimization weight might vary from one case to the other.

Treatment plans were made only on the primary CTV target for each case. Fractions were separately planned for 4 different angles: -90°, -50°,50°,90°, excepted for patient 1 where angle of -90° and 90° were not possible due to shoulders relative position with CTV, i.e. the shoulders were in the beam path, and for patient 4 where -50° and 50° were not possible due to the presence of metallic implant, resulting in a total of 20 plans. The beam energy was set up at the maximum of the Proteus Plus system, i.e., 228 MeV.

All plans were robustly optimized following a worst-case scenario approach. The D_{98} and the D_2 to CTV as well as the D_{max} to SC were set as robust. The range error was defined as 2.6% and the isotropic setup error as 4 mm according to our past studies [192, 207].

2.4. Planning on CT_1

For all patients, plans were generated for the beam angles described above (1 SFUD beam per fraction). The generation process followed the workflow described above for the CT_1 image. Once the conformal plans were generated, their RFs were saved for reuse in the spot-on adjustment described below.

2.5. Full adaptation on CT_2

As a planning reference on CT_2, we generated plans in exact same way as in the previous section "*planning on CT_1*" but using the CT_2 image instead. This is thus a conventional optimization on the CT_2 image with an RF that is fully designed for this anatomy. Such a reference method will be referred to as "full adaptation" (FA).

2.6. Spot adaptation only on CT_2

The proposed adaptation workflow on CT_2 starts by conventionally generating a conformal FLASH plan on CT_2. Once created, the RF matrix from the homologous initial plan on CT_1 is loaded to replace the newly-generated RF matrix for CT_2. The treatment spot weights are then classically optimized on CT_2 patient anatomy with the initial RF. Such method will be referred as "Spot adaptation only" (SAO).

2.7. No adaptation on CT_2

With this approach, plans generated on CT_1 were recomputed on CT_2. To do so, CT_1 and CT_2 images were rigidly-registered. The reference plan from CT_1 is then computed on the rigidly-registered CT_2. No spot weight nor RF matrix is re-optimized. Such approach will be referred as "no adaptation" (NoA).

2.8. Plan comparison

This section describes how the different planning approaches on CT_2 were compared. Since only one repeated CT per patient was available, for each beam angle a conformal plan was generated on the CT_2. Additionaly, the performance of each treatment plan was evaluated independently of the other plans.

The different approaches were dosimetrically evaluated on the nominal case and the worst-case scenario. D_{98} and D_2 for the CTV, D_{max} for the spinal cord (SC) and D_{mean} for the other OARs when their contours were available, are provided. The conformity number was also computed to account for unnecessary dose in healthy tissue surrounding the target volume as followed:

$$CN = \frac{V_{95,target}^2}{V_{target} . V_{95,body}} \quad (1)$$

with $V_{95,target}$ the volume inside the CTV where the dose is equal or larger than 95% of the prescription dose, V_{target} is the volume of the CTV, $V_{95,body}$ is the volume

inside the body contour where the dose is equal or larger than 95% of the prescription dose.

Dose-rate was also evaluated between the different approaches. While in literature, the Dose-averaged dose rate (DADR) is often reported, we chose to not use it since it might strongly overestimate the real dose-rate as indicated by *Deffet et al.* [280]. Instead, we chose to use the percentile dose rate ($\dot{D}^{PERC}$) [281, 282] which is an extended definition of the PBS dose rate defined by *Folkerts et al.* [283]:

$$\dot{D}_i^{PERC} = \frac{(1-2p)D_i}{t_{1,i} - t_{0,i}} \quad (2)$$

with D_i being the total accumulated dose received in voxel i, p is an arbitrary percentage, equal in this case to 5% of the prescription dose, i.e. 5% of 9Gy, $t_{1,i}$ is the time elapsed to obtain (1-p).D_i at voxel i and $t_{0,i}$ the time elapsed to obtain the dose p.D_i at voxel i.

Following such dose-rate definition, the V40Gy/s, i.e, the percentage of the considered volume where the dose-rate is at least of 40Gy/s, was computed for the CTV and OARs; such dose-rate value (40Gy/s) being usually considered as the threshold to trigger the FLASH effect.

3. Results

Left panel in Figure 30 displays the distribution of the D_{98} to CTV for each approach on CT_2. The highest median is achieved by FA method with a value of 8.57Gy, followed by the SAO with a value of 8.52Gy and then the NoA with a value of 8.50Gy. It is noteworthy that even for FA, some plans presented a D98 lower than 8.55Gy due to the difficulty of achieving such value while minimizing D_2 and D_{max} to SC.

Distribution of D_2 to CTV is displayed on the middle panel of Figure 30. This time, the highest median is the SAO with a value of 9.49Gy, then the FA with a value of 9.45Gy and then the NoA approach with a value of 9.4Gy. Considering these results on CTV coverage, it appears that dose homogeneity is lower for SAO approach.

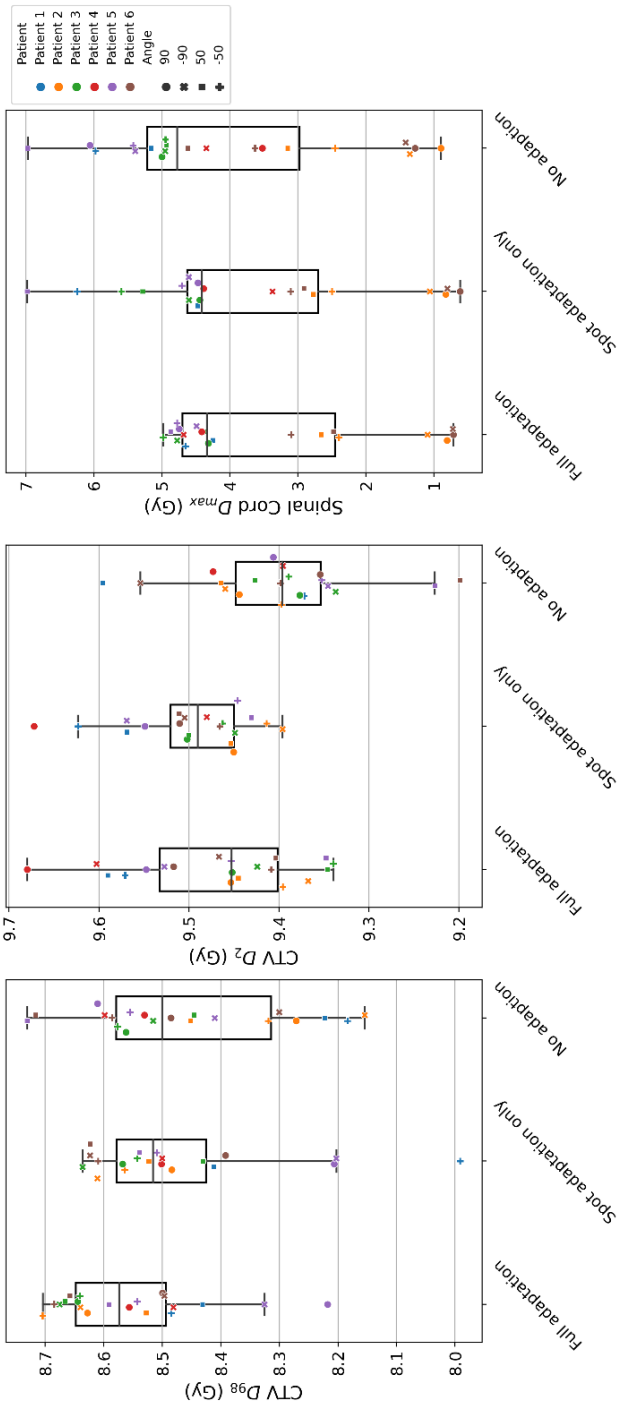
Distribution of D_{max} to SC is displayed on the right panel of Figure 30. The median value is 4.34Gy for the FA approach, 4.41Gy for the SAO and 4.77Gy for the NoA. For SAO and NoA methods some plans depicted noticeably high value. 4 and 6 plans show D_{max} higher than 5Gy for SAO and NoA, respectively.

Figure 31 illustrates the difference in target coverage that can happen between SAO and NoA approach. An anterior part of the CTV is not properly covered by the D95% isodose. Figure 32 illustrates the difference in target coverage that can happen between SAO and FA approach. A posterior part of the CTV is not properly covered, this time by the SAO approach.

Table 16 shows the difference in dose volume metrics for the nominal scenario. All SAO absolute average differences with FA are less than 0.1 Gy except for the D_{max} to SC, for which the average difference is of 0.22 Gy. For NoA, four metrics have an average absolute difference with FA greater than 0.1 Gy: the D_{max} to SC, which is of 0.61 Gy; the D_{mean} to oral cavity, which is of 0.20 Gy; the D_{mean} to mandible, which is 0.13 Gy and the D_{mean} to larynx, which is 0.14 Gy.

Table 17 shows the difference in dose volume metrics for the worst-case scenario. Trends are very similar to those of nominal case, except the CTV coverage with NoA that shows a mean decrease of about 0.17Gy compared to the nominal case while for SAO the average difference is of 0.07Gy.

Table 18 displays the difference of volume receiving at least 40Gy/s. For the different anatomical volumes, the $V_{40Gy/s}$ mean value is relatively close to 0 while some larger differences in maximal and minimal differences exist for both SAO and NoA approaches. We can notice that the volume of SC receiving 40Gy/s is often 0 or close to, suggesting that such volume would not beneficiate of the FLASH effect.



Distribution of dosimetric values on the nominal case. Left: boxplot of the CTV D98 for each treatment plan on CT2 for the different adaptation workflow. Middle: Boxplot of the CTV D2 for each treatment plan on CT2 for the different adaptation workflow. Right: Boxplot of the maximum dose to the spinal cord for each treatment plan on CT2 for the different adaptation workflow. Individual values are denoted by different marker and color according the patient and the beam angle. Full adaptation: re-optimization of the ridge filter and the treatment spots on CT2; Spot-only adaptation: re-optimization of the treatment spots while the ridge filter from CT1 is re-used; No adaptation: the treatment plan defined on CT1 is recomputed on CT2.

Figure 30

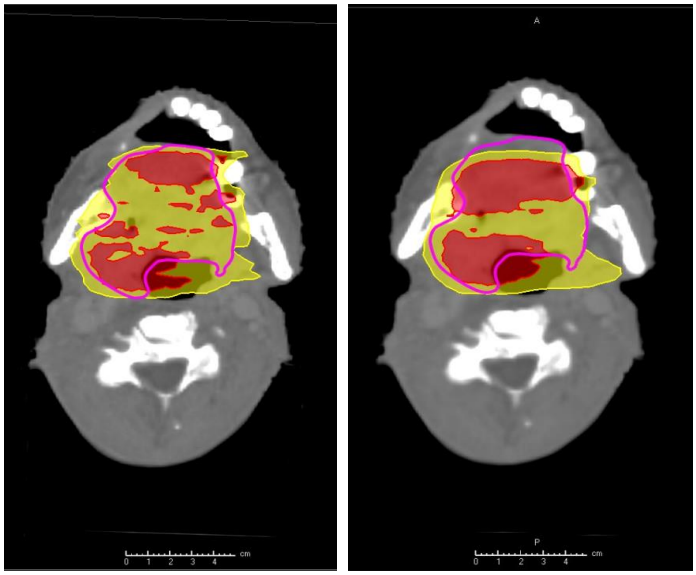


Figure 31 Dose distribution for Patient 2 with a beam angle of -90°. Left: SAO plan, Right: NoA plan. Pink contour is the CTV; yellow isodose depicts the area where dose is equal or higher than 8.55Gy, red isodose depicts area where dose is equal or higher than 9Gy. Notice the anterior part of CTV not cover by the yellow isodose for NoA plan.

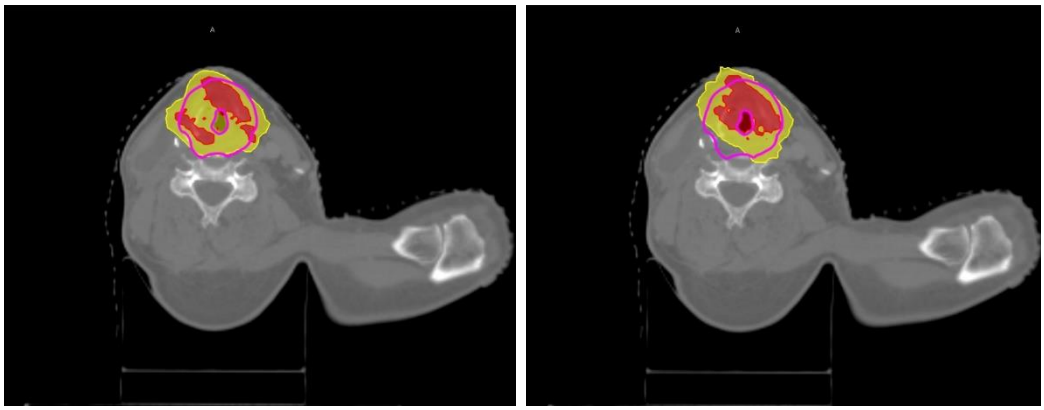


Figure 32 Dose distribution for Patient 1 with a beam angle of -50°. Left: FA plan, Right: SAO plan. Pink contour is the CTV; yellow isodose depicts the area where dose is equal or higher than 8.55Gy, red isodose depicts area where dose is equal or higher than 9Gy. Notice the posterior part of CTV not cover by the yellow isodose for SAO plan.

Table 16:Difference of dose metrics for NoA and SAO approaches compared to FA in nominal case. NoA: no adaptation (CT1 plan re-computed on CT2); SAO: spot adaptation only; FA: Full adaptation.

Volume	Metric	NoA minus FA		SAO minus FA	
		Mean	Range	Mean	Range
CTV_p	D_{98} **(Gy)**	-0.09	[-0.49; 0.39]	-0.08	[-0.49; 0.13]
CTV_p	D_2 **(Gy)**	-0.07	[-0.21; 0.09]	0.03	[-0.12; 0.15]
Body	CN	-0.03	[-0.15; 0.07]	-0.02	[-0.06; 0.05]
Spinal cord	D_{max} **(Gy)**	0.61	[-0.89; 2.15]	0.22	[-1.31; 2.10]
Parotid L	D_{mean} **(Gy)**	0.03	[-0.98; 1.08]	-0.01	[-0.61; 0.19]
Parotid R	D_{mean} **(Gy)**	0.04	[-0.82; 1.03]	0.08	[-0.98; 0.72]
Oral cavity	D_{mean} **(Gy)**	0.20	[-0.35; 1.15]	-0.04	[-0.42; 0.07]
Mandible	D_{mean} **(Gy)**	-0.13	[-1.35; 1.04]	0.05	[-0.44; 0.72]
Thyroid	D_{mean} **(Gy)**	-0.08	[-0.59; 0.03]	-0.03	[-0.25; 0.03]
Larynx	D_{mean} **(Gy)**	0.14	[-1.02; 0.90]	0.01	[-0.08; 0.24]

Table 17 : Difference of dose metrics for NoA and SAO approaches compared to FA in worst-case scenario. NoA: no adaptation (CT1 plan re-computed on CT2); SAO: spot adaptation only; FA: Full adaptation.

Volume	Metric	NoA minus FA		SAO minus FA	
		Mean	Range	Mean	Range
CTV_p	D_{98} **(Gy)**	-0.17	[-1.00; 0.57]	-0.07	[-0.57; 0.37]
CTV_p	D_2 **(Gy)**	-0.07	[-0.24; 0.13]	0.04	[-0.05; 0.20]
Body	CN	-0.03	[-0.11; 0.05]	-0.02	[-0.08; 0.05]
Spinal cord	D_{max} **(Gy)**	0.63	[-0.83; 2.16]	0.25	[-0.34; 1.53]
Parotid L	D_{mean} **(Gy)**	0.04	[-0.81; 1.16]	0.01	[-0.72; 0.21]
Parotid R	D_{mean} **(Gy)**	0.04	[-0.97; 0.96]	0.09	[-1.23; 0.95]
Oral cavity	D_{mean} **(Gy)**	0.24	[-0.36; 1.14]	0.00	[-0.47; 0.08]
Mandible	D_{mean} **(Gy)**	-0.21	[-1.71; 0.99]	0.06	[-0.36; 0.69]
Thyroid	D_{mean} **(Gy)**	-0.06	[-0.42; 0.04]	-0.03	[-0.24; 0.04]
Larynx	D_{mean} **(Gy)**	0.14	[-1.05; 1.04]	0.01	[-0.09; 0.25]

Table 19 shows the number of plans meeting the chosen tolerance parameters on dosimetric differences for nominal and worst-case scenario. If one of the chosen criteria is not met, either on the nominal or the worst-case scenario, the plan is deemed outside the tolerance range. Note that the tolerance criterion on D_{max} to SC only applies when this value on the nominal case of FA plan is larger than 4Gy, i.e. when the value is getting close to chosen limit of 5 Gy. We can notice that SAO

is much closer to the FA reference, for example the number of plans meeting a maximum difference of 0.3Gy on the CTV D_{98}, CTV D_2 and SC D_{max} being of 16 for SAO while being 5 for the NoA plans. Previous tables and figures showed reasonable dosimetric gap between NoA and SAO approach from a global patient cohort perspective. Nonetheless, when evaluating the gap between FA and NoA/SAO plans-by-plans this table highlights a more important gap between NoA plans and SAO ones.

Table 18 : Difference in volume receiving 40Gy/s (V40Gy/s) for each anatomical volume for NoA and SAO approaches compared to FA. NoA: no adaptation (CT1 plan re-computed on CT2); SAO: spot adaptation only; FA: Full adaptation.

Volume	**Metric**	**NoA minus FA**		**SAO minus FA**	
		Mean	**Range**	**Mean**	**Range**
CTV_p	V40 Gy/s (%)	-4.49	[-68.13; 33.89]	-2.13	[-18.74; 10.03]
Body	V40 Gy/s (cm^3)	31.49	[-55.61; 219.16]	1.08	[-43.39; 88.46]
Spinal cord	V40 Gy/s (%)	0.01	[0.00; 0.01]	0.00	[0.0; 0.01]
Parotid L	V40 Gy/s (%)	0.91	[-8.79; 10.83]	-0.75	[-19.07; 10.34]
Parotid R	V40 Gy/s (%)	3.5	[-1.14; 39.60]	1.26	[-18.88; 14.28]
Oral cavity	V40 Gy/s (%)	0.26	[-5.78; 8.60]	-0.35	[-3.33; 4.81]
Mandible	V40 Gy/s (%)	-1.08	[-13.60; 9.73]	1.36	[-2.20; 9.25]
Thyroid	V40 Gy/s (%)	-1.30	[-7.43; 0.00]	-0.50	[-4.72; 0.08]
Larynx	V40 Gy/s (%)	3.07	[-8.62; 12.44]	1.20	[-13.66; 21.33]

Table 19 Number of plans (amongst the 20 ones) meeting chosen difference tolerance on the D98 and D2 to CTV as well as Dmax to spinal cord (SC) on nominal and worst-case scenario. The tolerance on Dmax to SC only applies if the Dmax to SC is at least 4Gy in the nominal case for FA plans. Values are provided for NoA and SAO approach. To be considered as meeting the difference tolerance, each plan has to satisfy the three considered tolerances on the nominal and worst-case scenario. The prescription dose for each plan is 9 Gy to the CTV. NoA: no adaptation; SAO: spot adaptation only.

Chosen tolerances with respect to Full Adaptation plan			**Number of plans meeting the chosen tolerances**	
Max decrease D_{98} CTV (Gy)	Max increase D_2 CTV (Gy)	Max increase D_{max} SC (Gy)	NoA	SAO
0.1	0.1	0.1	**2**	**7**
0.1	0.1	0.3	**2**	**8**
0.15	0.15	0.15	**3**	**12**
0.2	0.2	0.2	**5**	**14**
0.2	0.2	0.3	**5**	**15**
0.3	0.3	0.3	**5**	**16**
0.4	0.4	0.4	**6**	**16**
0.5	0.5	0.5	**8**	**16**

4. Discussion

In this study we presented an adaptive workflow for FLASH conformal plans. Such workflow relies on updating spot weights on the image of the day while re-using the initial ridge filter (RF) for all treatment fractions.

Over the evaluated treatment plans generated on CT_2, it was shown that adaptation was necessary for up to 18 fractions over 20 to maintain plan clinical quality, while the others remaining non-adapted (NoA) treatment plans were close to the FA ones (Table 19). This indicated that while some major anatomical changes can be observed from an initial CT to a CT of the day, it does not necessary mean that plan adaptation will be necessary as anatomical changes might not necessarily occur in the area of the planned beam angle of the day. However, it could be noted that from a global patient cohort perspective, the general distribution of the different dosimetric values were generally not strongly different with the FA approach. The gap in dosimetric quality between NoA and FA was only highlighted through the analysis of dosimetric tolerance plan-by-plan (Table 19).

Such low global differences in dosimetric values could be explained by the simplicity of SFUD planning. Indeed, while less robust than transmission treatment, conformal FLASH consists in single SFUD beam. This approach remains simpler in term of modulation and thus target conformation, compared to conventional multi-field optimization where dose gradients from different beams need to be adequately combined to deliver an acceptable dose distribution, even in the presence of uncertainties. Consequently, it seems logical that SFUD plans with lower complexity could be less impacted by anatomical uncertainties, i.e., their simplicity leads to a certain robustness.

Regarding the dosimetrical quality of the spot adaptation only workflow (SAO), we showed that up to 16 over the 20 were within 0.3 Gy for key dose metrics to their homologous FA plans (while only 5 for NoA plans for same dose-volume metric differences). This highlighted the added value of this adaptation workflow. Since the SAO approach should not lead to a significantly prolonged planning, i.e., simply optimizing the spot weights of a single SFUD conformal beam with the initial RF, such workflow could be easily implemented in a clinical routine.

Concerning the change in volume for OAR receiving dose-rate of at least 40 Gy/s, we showed that on average no large differences were noticed between NoA, SAO and FA. However, some cases could show larger deviation in V40Gy/s for some anatomical volume. At the moment it is difficult to quantify how this would impact the clinical quality of a plan since it would be necessary to precisely know the flash dose modifying factor [284, 285] of the different tissues for the achieved dose-rates. Nonetheless, such divergence was not a global trend and more a consequence of different spot placement and spot weights between the different planning approaches. Furthermore, it is important to note that while this study considered 40 Gy/s as the reference threshold for FLASH effect activation, determining the precise threshold value remains an active research topic.

A first limitation of this study lies in the size of the patient cohort, i.e. 6 head and neck patients with a total of 20 evaluated plans. Future studies will have to investigate over a larger cohort of patient as well as over different beam angles and treatment locations to precisely evaluate the limits of the proposed SAO approach.

Another limitation of this study concerns the optimization goals used. We struggled to fulfill CTV D_{98} > 8.55Gy and SC D_{max} < 5.0Gy for several fractions even for the FA approach. This is probably due to the limited degree of freedom, implied by the use of single SFUD beam combined with the strictness of the chosen dosimetric objectives. However, it is hard to evaluate the clinical quality of such plans as the literature on clinical objectives for conformal FLASH is sparse. Nonetheless it is important to remember that the primary goal was to generate plans as qualitative as possible with a chosen set of dosimetric constraint to then evaluate the proposed SAO workflow.

Although the SC received doses exceeding 5 Gy, the dose-rate remained predominantly below 40 Gy/s, indicating the FLASH effect would not be triggered. As mentioned above, this outcome probably stems from the challenge of meeting both target coverage and SC objectives using a simple SFUD beam. The situation highlights the importance of clarifying the dose and dose-rate threshold triggering the FLASH effect. Indeed, if triggered at lower doses like 4 Gy, maintaining FLASH dose-rates in such OAR could become crucial. Given such a concern, this also highlights the importance for commercial TPS to include dose-rate and dose-modifying factor objectives in their optimizer [286, 287]. These kind of optimization terms could theoretically allow to use the full potential of the FLASH radiotherapy, resulting in significantly improved treatment plans.

Another limitation is that at least 4 cases could not achieve results comparable to FA using SAO approach. It means that in these cases, unless another beam angle is found to be only slightly affected by anatomical changes, treatment would need to be interrupted to allow time for a new RF to be printed. This could thus hamper the viability of such treatment workflow. Therefore, investigating alternative approaches such as dynamic RF could also provide a valuable solution for conformal FLASH implementation [273].

5. Conclusion

The proposed spot-adaptation workflow showed dosimetric results close to the full adaptation for up to 16 over the 20 evaluated fractions (within 0.3 Gy difference for key dose metrics). The dose rate results were globally close to the FA plans. Given the fact that SAO workflow only requires re-optimization of the spot weight of a single conformal beam, which is generally considered to be a fast

process, this approach could be applied in a clinical routine for each treatment fraction. This would leave the re-printing of the ridge filter only for extreme cases where SAO is not sufficient.

Chapter 6: Discussion and conclusion

The finding of the previous chapter can be shortly summarized as follow:

Chapter 2 addresses the difficulty of producing high-quality plan in OART. In this investigation, we compared two popular approaches for automatic planning. The first automatic approach, the template-based, consists in simply using a list of clinical goals ordered by clinical importance. The second one uses AI-based dose prediction to generate isodose contours that are then used to mimick the dose on the TPS (DP+DM). The considered TPS was the Ethos one using the IOE automatic optimizer. This study highlighted a dosimetric gap between template of conventional clinical goals and predicted patient specific goals (i.e. predicted doses followed by a dose mimicking). The DP+DM approach outperformed the template-based one on a test set of 10 prostate cases. The template-based approach showed difficulty providing a quality close to the manual planning and even produced several non-clinically acceptable plans when trying to tighten up the dose constraints. The DP+DM approach achieved results much closer to manual planning on evaluated patients (while remaining inferior) and the optimized plans were all clinically acceptable.

Chapter 3 addresses the challenge of efficiently reviewing and correcting automatically segmented volumes in OA conditions. While automatic segmentation has shown great progress in recent years with the arrival of AI models, achieving consistently perfect automatic contours currently remains an elusive goal. Given that contour review and correction represent one of the most time-consuming aspects of adaptive workflows, we developed a quality assurance tool, the DIVE map, that prioritizes corrections based on their potential dosimetric impact rather than geometric accuracy alone. Indeed, conventional reliance on geometric indicators such as the Dice coefficient and Hausdorff distance for evaluating segmentation quality is a limitation for OA conditions. While these metrics serve a purpose in benchmarking segmentation algorithms, they provide an incomplete picture of clinical relevance, particularly in the context of adaptive radiotherapy. More problematically, such geometric indicators could potentially mislead the clinicians, as a segmentation algorithm with low Dice scores might still be clinically acceptable if its errors occur in dosimetrically insignificant regions.

The developed DIVE map tool addresses these limitations through the generation of deformation scenarios of the automatically-segmented volumes and

subsequent prediction of dose distributions for these scenarios. This approach enables the identification of dosimetrically relevant regions requiring correction.

Using the proposed DIVE map, we demonstrated that contour corrections could be substantially reduced without compromising the dosimetric quality of Ethos treatment plans. In more than half of the evaluated cases, less than 50% of the conventional corrections were necessary to maintain dosimetric quality, revealing that many routine corrections may have limited clinical impact. Nonetheless, several points remain to be addressed to ensure its reliability and time-efficiency.

Chapter 4 tackled the problem of enabling OA for an emerging treatment technique close to clinical introduction: the proton arc therapy. PAT specific problems, i.e. determining optimal energy layer per angle and having to optimize numerous plan parameters, adds another layer of complexity to the PT plan adaptation. Nonetheless, since PAT emerged as the logical evolution of IMPT as the VMAT did to IMRT, we showed that a previously proposed IMPT workflow could be made compatible to PAT problems. The proposed PAT adaptive workflow achieved dosimetric results close to the full re-optimization from the ground up, while being 3.4 times faster than conventional full re-optimization. Nonetheless, the achieved adaptation time was still too slow to be compatible with clinical conditions. However, we believe that such timings are due to the research-oriented TPS we used to generate the results and that substantial time improvement could be achieved on commercial TPS.

In Chapter 5, we investigated OA for another emerging treatment modality: the conformal FLASH PT. In FLASH therapy, the problem of achieving a high-quality plan pass by ensuring an extremely high dose-rate in order to trigger the so-called FLASH effect, a radiobiological effect reducing the toxicity to healthy tissues while preserving the tumor control. For the conformal FLASH, the current limit is related to the necessary 3D-printing of a patient-specific RF that is often considered to last for at least one hour leading to unpractical FLASH plan adaptation in clinical conditions. Nonetheless, we demonstrated that such 3D-printing was not necessary for each fraction even when large anatomical changes were observed. We highlighted that simple spots weights re-optimization with the initial RF could lead to dosimetric results close to the full plan and RF re-optimization for most of the treatment fractions.

1. Time is key

As previously highlighted, the combination of OA with emerging techniques could enable critically low toxicity and thus unpreceded treatment quality. Since PAT has already demonstrated significant benefits on different anatomical sites [239-241], its combination with OA could be game-changer from a purely dosimetric point of view. Similar enthusiasm applies also for combination of FLASH effect with OA (provided clinical value of FLASH treatment is demonstrated in a near future).

However, as introduced before, dosimetric benefits of PT-based techniques should always be considered alongside time considerations due to their inherent cost implications. For instance, it has been shown that for IMPT lung patients, adaptation time should be contained within a range of 6 to 20 minutes to still remain relevant from a global NTCP point of view. Beyond this window, some patients would need to be redirected to less optimal treatment modalities due to limited proton center availability, entailing the global NTCP of the patient cohort [35]. This clearly demonstrates the critical importance of optimizing adaptation time for these techniques to be truly competitive.

This fundamental balance has already influenced existing OA PT implementation as evidenced by the first OA-PT clinical workflow utilizing fast analytical algorithms and currently limited to rigid contour propagation cases requiring no corrections [30, 189]. This emphasis on practical efficiency led us to investigate how to optimize automatic segmentation and automatic planning in chapter 2 and 3 before focusing on adaptive workflow for emerging techniques in chapter 4 and 5 in order to ensure viable OA workflows for emerging techniques. The underlying goal was to improve the ratio of dosimetric benefits to adaptation time. Chapter 2 aimed at increasing such ratio by investigating how to automatically achieve plans of dosimetric quality close to manual planning (the current non-adaptive gold standard). Chapter 3 aimed at increasing this ratio by significantly reducing the adaptation time through optimization of the review and correction of the automatic segmentation, a step often considered to be a critical bottleneck of OA.

These considerations become even more critical for PAT which is usually assumed to require prolonged planning time when compared to IMPT. This suggests that OA-PAT would need to demonstrate substantial dosimetric advantages over

alternative treatment techniques to justify potentially extended adaptation times. This also emphasized on the importance of evaluating our proposed workflow on efficient TPS to determine how close we can bring OA-PAT to relevant timings. However, to push the reduction of adaptation time further, beamlet-free optimization could also be an interesting perspective for OA-PAT. Beamlet-free approach holds particular promise for accelerating planning optimization. Unlike conventional methods, it eliminates the need to generate the dose influence matrix prior to plan optimization, instead computing it simultaneously with spot weight optimization [288]. The resulting time savings become more significant as the number of spots increases, making it particularly valuable for PAT which may require more spots than IMPT.

For FLASH therapy, timing considerations will depend on its stage of clinical implementation. Interestingly, the inherent requirements for triggering the FLASH effect currently lead to simplified plan optimizations. This potentially makes OA less challenging for FLASH than for PAT or IMPT. Indeed, the necessity of maintaining FLASH dose-rates has driven the use of single SFUD beams with fewer spots. These simplifications result in lightweight dose computation due to fewer beamlets and limited beam parameters to optimize. Furthermore, the proposed workflow in chapter 5 proved the validity of re-using the same RF for most of treatment fractions. This tends to indicate that plan optimization could not be a problem for OA-FLASH. Concerning, the fractions for which re-using the RF would not be valid, evolution in RF manufacturing will probably still occur in the following years. Refinements of dynamic RF concept [273, 289] or arrival of faster 3D printing methods could address cases requiring RF adaptation, further enhancing the clinical viability of adaptive FLASH treatments. Moreover, the development of FLASH-dedicated optimization constraints in the future, e.g. constraints that maximize the dose-rate or the FMF in relevant anatomical regions, could also help maximizing the dosimetric benefits to time ratios.

In the end, the future may not lie in perfecting every adaptive element, but in orchestrating a symphony of 'smart' compromises that maximize therapeutic benefit within practical clinical constraints. This strategic evolution could finally bridge the gap between the theoretical promises and clinical reality of adaptive RT.

2. The challenges of Artificial Intelligence

The successful implementation of OA for emerging techniques, by maximizing the dosimetric benefits over adaptation time ratio, relies heavily on advanced automation technologies. Approaches based on AI have demonstrated considerable potential in the recent years. It is therefore natural to consider them. However, several challenges still exist and need to be addressed in order to exploit the full potential of AI.

A first and fundamental challenge lies in the data requirements: as with most AI applications, it requires a sufficiently large training database to ensure reliable and relevant model outputs. Despite the existence of architectures such as the U-net that perform reasonably well with limited training data [102], having access to a relevant database even a small one, e.g. 30 training patients as in chapter 2, can be challenging for small RT centers. This challenge is particularly pronounced when considering the requirement for high-quality, homogeneous data characterized by consistent prescription doses, beam arrangements, and clinical requirements to achieve reliable AI predictions. The complexity increases further in the context of automatic dose planning, where the inherent institution-specific nature of training data introduces additional challenges. Parameters such as prescription doses, target volume definitions, beam configurations, and clinical objectives often vary significantly between treatment centers. Although TPS vendors attempt to address this through diverse model collections, achieving universal applicability remains challenging. Research also continues to address these limitations through various approaches such as integrating the beam parameters in the model inputs [120] or using transfer learning from an already well performing model [209], and despite the promising arrival of foundation models trained on vast datasets [290-292], significant concerns persist.

Furthermore, the possibility that AI models could produce critically bad results potentially impacting medical outcome often creates hesitancy among clinicians to implement such systems in clinical practice [217]. Particularly for the planning steps where other non-AI approaches can easily bring clinically acceptable results. The template-based approach, for instance, offers both simplicity in implementation - requiring only a defined set of clinical goals - and transparency in operation, characteristics that often prove more reassuring to clinicians than

the perceived opacity of AI methodologies [293-295] while it can lead to suboptimal result as demonstrated in Chapter 2.

The evaluation of AI models and AI-based tools, including our DIVE-map, raises another fundamental methodological question. While the standard approach involves validation on independent test sets, ensuring these sets are sufficiently large and diverse to demonstrate reliable global performance presents a significant challenge. The inherent difficulty lies in capturing the full spectrum of possible clinical scenarios within a finite test dataset, limiting our ability to make comprehensive assessments of model reliability [296, 297]. Several approaches could address these validation challenges. One promising direction involves the establishment and utilization of standardized benchmark databases such as the OpenKBP challenge [298] or Cancer Imaging Archive [299]. Additionally, generation of synthetic phantoms to challenge AI models could also be an interesting alternative approach in assessing a broader range of scenarios [220, 300].

In response to these challenges, interest in quantifying the quality and reliability of predictions is also emerging [293] with development of tools such as uncertainty quantification [301-304]. Recent developments in this field have produced tools capable of assessing model uncertainty at multiple scales - from comprehensive prediction-level evaluations to granular voxel-wise uncertainty maps. These methodologies represent a promising direction for enhancing clinical safety by enabling the detection and exclusion of potentially unreliable predictions before clinical conditions.

These few elements and examples illustrate the current challenges in achieving reliable automation of RT workflows through AI. However, while concerns about AI reliability are valid, it is essential to recognize that conventional automation methods are not immune to critical failures. Our research demonstrated this vulnerability through the Ethos IOE system, which produced clinically unacceptable plans when presented with overly stringent template goals. This observation emphasizes the fundamental importance of maintaining fallback treatment plans in automatic planning systems, a practice already established in current commercial RT platforms and initial clinical deployments of OA PT [30].

This also emphasizes that AI should be viewed as any other tool with its own set of capabilities and limitations, like other automation approaches in RT.

3. Managing automation uncertainties through workflow design

While significant advances in automation and AI are anticipated in the coming years, developing strategies to prevent critical automation failures in clinical conditions remains a pressing priority. Although maintaining fallback plans provides a safeguard against automation failures in OA, the field has begun to investigate AI robustness in RT applications.

As previously mentioned, recent research has explored robustness evaluation through offline testing using synthetic phantoms, representing a controlled approach to assessing system reliability [220, 300]. However, beyond these validation methods, novel approaches are emerging that specifically address the challenge of ensuring robust automated processes within the constraints of OA conditions.

In automatic dose planning, the dose mimicking process represents a critical step where robustness could be introduced. Research has demonstrated that mimicking parameters can exert substantial influence on plan quality, sometimes exceeding the impact of the initial AI prediction [122, 192, 204]. While the focus is often put on AI performance, it has been demonstrated that a well-designed mimicking process can effectively compensate for suboptimal dose predictions [204]. Consequently, a careful design of dose mimicking through the selection of a relevant method and mimicking parameters could account for a certain range of AI prediction errors.

Our implementation of dose mimicking in chapter 2 provides an illustrative example of this concept. The relative success of our dose prediction and mimicking approach compared to the template-based method, despite our limited training database, may be attributed to our specific mimicking implementation. Notably, we employed isodose contours with larger intervals than those typically reported in recent literature. While this broader spacing might theoretically result in slightly less optimal results, it likely provides greater robustness against local inaccuracies in predicted dose gradients by providing greater flexibility for the optimizer.

This exemplifies how robustness to these uncertainties can be integrated by design in the following workflow steps. Similar to how robustness optimization in PT accepts higher OAR doses to ensure target coverage under uncertainties, this approach might result in marginally less optimal plans while ensuring more consistent and reliable performance across diverse cases.

One way to push further this concept would be with the use of uncertainty estimation. Uncertainty quantification could guide the selection of mimicking methods and its parameters. As prediction certainty decreases, the mimicking process could automatically adjust to become more robust to potential errors. In cases of very high uncertainty, the system could even switch to conventional template-based optimization approaches, providing a graduated response to varying levels of prediction confidence.

Such concept of robustness can also be extended to automatic segmentation. One way could be to incorporate AI uncertainty in the planning step with either planning organ at risk volume (PRV) [305, 306], robust and/or probabilistic optimization [307, 308] on OAR probable volumes location. Initial investigations using different non-rigid registration algorithms have shown promising results [309], demonstrating that while this approach may be less dosimetrically optimal than full contour correction, it still offers significant benefits over non-adaptive approaches.

From a practical perspective and given current limitations of the proposed DIVE map, a simple approach would be to evaluate the worthiness of correcting each volume based on dosimetric impact and time considerations, i.e. build a list of critical volumes that should be corrected for each patient. Manual corrections, being time-consuming, would then be reserved only for volumes providing clear overall NTCP benefits. Volumes with lower "dosimetric benefit to time" ratios would either remain uncorrected or be handled through intermediate approaches such as PRV, robust, or probabilistic optimization. This pragmatic approach is already partially implemented in commercial systems like Ethos, where manual corrections are limited to target volumes and "influencer" OARs [27].

Drawing a parallel with the use of uncertainty quantification in the robustness process, uncertainty estimation could also potentially automate the adjustment of these robustness margins or robustness parameters. In cases of high uncertainty,

the workflow could even switch from AI-based segmentation to conventional non-rigid registration methods. This approach would mirror the previously proposed uncertainty-based selection of dose mimicking parameters, creating a more comprehensive framework for managing AI-related uncertainties in OA conditions. First investigations of such an approach have recently been carried out with non-rigid registration of volumes and provided promising results [310].

4. Taking review and correction of automatic segmentation a step further

Although introducing robustness to automatic segmentation processes could significantly reduce review time, this approach presents a critical trade-off. The introduction of additional margins and/or robustness parameters, while providing greater reliability, may excessively compromise dosimetric advantages. Such compromises could ultimately undermine the crucial balance between dosimetric benefits and treatment time. It remains thus critical to investigate different approaches to optimize at best the required time of auto-segmentation review and correction.

Chapter 3 addresses this challenge through the introduction of the DIVE-map, a tool indicating the dosimetric relevance of correction regions around automatically segmented volumes. This work aimed at paving the path for QA tools of automatic-segmentation, where online solution to efficiently ensure quality of auto-segmented contours are few [311, 312] and where even commercial solutions remain mostly specific to retrospective or offline conditions such as the AIqualis software from PTW or Verify from Mvision [313]. While the DIVE-map offers a novel perspective on contour verification, several significant challenges persist.

Key methodological concerns potentially affecting physician confidence include the reliance on dose prediction for deformation scenarios and the independent generation of DIVE-maps for dosimetrically interdependent structures. Although these limitations proved acceptable in our test set, broader clinical implementation would require methodological refinements to ensure reliability across diverse cases. Additionally, the current DIVE-map generation process exceeds clinically acceptable timeframes, necessitating optimization to achieve

one to two-minute processing times for meaningful integration into adaptive workflows.

Although the current implementations remain at an early stage, the demonstrated potential of these adaptive solutions provides compelling motivation for continued development. Such future developments might also benefit from other emerging perspectives for auto-segmentation, particularly the uncertainty quantification methods previously cited. While global uncertainty quantification, which provides a single score for volume segmentation, aligns more with geometric quality indicators of limited utility in OA conditions [314], voxel-level uncertainty quantification offers promising opportunities for guiding physicians to potential segmentation errors [315]. The integration of such local uncertainty information with dosimetric guidance tools like the DIVE-map could enable a comprehensive optimization of the contour review process. Currently, the DIVE-map provides dosimetric guidance across broad regions without considering the probability of segmentation errors. Combining it with local uncertainty information would help direct attention to areas where corrections are both dosimetrically significant and likely necessary. However, both uncertainty prediction and the DIVE-map require substantial refinements before such integration becomes feasible. The effective visualization of such QA tools remains a critical concern, as it could significantly influence their clinical adoption. For example, recent research has demonstrated varying levels of clinician acceptance for different uncertainty visualization approaches with binary voxel-level uncertainty indication (certain/uncertain) receiving the highest preference among practitioners [316].

Beyond these technical improvements in contour evaluation and correction tools, it is also interesting to note that to optimize auto-segmentation step, research groups also started to investigate how to efficiently distribute the work load of online segmentation corrections such as by delegating the correction to Radiation Therapist (RTT) or non-expert radio-oncologist [317] or by parallelizing the corrections on different computers and different reviewers [167]. Other groups also proposed more efficient edition tools as other potential ways to improve auto-segmentation step. Such as interactive editing tools where the physicians would simply click on error regions in order to trigger an automatic contour correction provided by the AI model [318, 319].

This multi-faceted development pathway offers promising perspectives toward more efficient and reliable auto-segmentation in OA conditions.

5. Final considerations about the future of online adaptive radiotherapy

This thesis explored different facets of online adaptive radiotherapy (OART) with the aim of contributing to its evolution toward a more refined and efficient treatment approach. However, the future trajectory of OA will be significantly shaped by both industrial developments and clinical experience. The following discussion examines these evolving trends and their implications for the future of radiation therapy, considering both the technical advances presented in this work and the broader context of emerging industry solutions.

The durable presence of AI in RT appears increasingly certain, as evidenced by recent industry developments. Elekta announcement of an AI-driven linear accelerator and MagnetTx introduction of comprehensive AI solutions for their MR-linac system signal a fundamental shift in treatment delivery technology. This evolution extends beyond mere software additions to become integral components of treatment delivery systems.

The emerging competition between Varian HyperSight and Elekta Evo system, both featuring enhanced CBCT capabilities but with an opposed approach (detector development against AI development to enhance CBCT quality), may provide crucial insights into the relative value of hardware versus software innovations in imaging technology. The widespread adoption of these improved CBCT-guided adaptive systems could fundamentally reshape the landscape of OART. This technological advancement might lead to a broader use of OART, potentially relegating MR-linac systems to specific clinical scenarios requiring superior soft tissue contrast or real-time imaging capabilities.

The future of OA imaging appears to be moving toward a stratified approach. While MR-linac technology will likely maintain its relevance, particularly with recent developments like Unity and the announced Aurora-RT, its application may become increasingly specialized. The financial considerations and workflow efficiency challenges associated with MR imaging, including longer treatment times and higher operational costs [320, 321], contrast sharply with the streamlined efficiency of CBCT-guided adaptation on systems like Ethos. MR-linac

systems, with their superior soft tissue contrast and capability for real-time imaging without radiation exposure, may find their niche in complex cases where CBCT-guided adaptation proves insufficient. Such niche could perhaps in the future get enlarged with potential development of true real-time treatment adaptation, i.e. where adaptation process does really occur at sub-second timeframe, through their cine-imaging capability [322, 323].

This evolution in imaging technology could significantly impact the PT field, where the traditional reliance on in-room CT systems has influenced facility design and workflow. The emergence of high-quality CBCT solutions (such as Ethos HyperSight, Elekta Evo), combined with advanced artifact correction algorithms, might eliminate the need for space-intensive CT-on-rails systems, potentially reducing facility costs and improving treatment efficiency.

Concerning the other advantages that OA could bring, it has to be noted that it provides clear potential for reduction of treatment margin [100, 101, 324] and treatment setup uncertainty [325, 326]. However, as previously said, such potential represents both opportunities and challenges. While daily adaptation theoretically allows for smaller treatment margins, this benefit must be balanced against the uncertainties introduced by automated processes. If one wanted to address the AI-uncertainties alongside with common treatment uncertainties, the cumulative effect of AI uncertainties in segmentation and planning, combined with residual setup and delivery uncertainties, would require careful consideration in margin definition. Nonetheless, the possibility of reducing uncertainty treatment setup in robust optimization would be particularly relevant for OA-PT techniques as this could strongly decrease the NTCP for the same adaptation time.

It is also interesting to note that even if OART leads to an increase workload during treatment sessions, it also represents interesting opportunities to significantly reduce the workload of pre-treatment session. By making use of developed automation process for pre-treatment steps it can alleviate the workload of clinicians. The concept can even be pushed further by considering its capability to enable simulation-free treatments where all the pre-treatment simulations steps would not have to be carried on. Most of the workload would remain for the treatment session. However, in current conditions the disposal of a

falloff treatment remains critical for fraction where automation would strongly fail.

In conclusion, automation in RT environment presents strong opportunities for the whole field but it remains necessary to consider all the potential noxious effects that it could also bring with it. The future of automation in RT and PT presents a critical point: while recent research promises transformative capabilities, successful implementation requires more than technological sophistication alone. The most significant advances may emerge not from creating increasingly complex automated tools, but from designing workflows that strategically integrate these innovations. Success in maximizing the potential of OA and emerging treatment techniques will depend on our ability to balance technological advancement with practical clinical constraints.

A. Supplementary materials chapter 2

1. S1

The modulation complexity score was calculated following the same formula than the original article:

Pos is the leaf positions, N is the number of open leaves.

$$pos_{max} = \langle \max(\mathrm{pos}_{N \in n}) - \min(\mathrm{pos}_{N \in n}) \rangle_{\text{leaf bank}}$$

$$LSV_{segment} = \left\langle \frac{\sum_{n=1}^{N} \left(pos_{max} - (pos_n - pos_{n+1}) \right)}{N \times pos_{max}} \right\rangle_{left\ bank} \times \left\langle \frac{\sum_{n=1}^{N} \left(pos_{max} - (pos_n - pos_{n+1}) \right)}{N \times pos_{max}} \right\rangle_{right\ bank}$$

$$AAV_{segment} = \frac{\sum_{a=1}^{A} \langle pos_a \rangle_{left\ bank} - \langle pos_a \rangle_{right\ bank}}{\sum_{a=1}^{A} \langle max(pos_a) \rangle_{left\ bank \in beam} - \langle max(pos_a) \rangle_{right\ bank \in beam}}$$

$$MCS_{beam} = \sum_{i=1}^{I} AAV_{segment\ i} \times LSV_{segment\ i} \times \frac{MU_{segment\ i}}{MU_{beam}}$$

$$MCS_{plan} = \sum_{j=1}^{J} MCS_{beam\ j} \times \frac{MU_{beam\ j}}{MU_{plan}}$$

2. S2

Table A1: Study-S2, detailed met goals for each approach. EG_init_selected = Ethos generated (using initial template) for patient manually selected among 45 initial patients, EG_upd_selected = Ethos generated (using updated template) for patient manually selected among 45 initial patients, DP+DM = dose prediction followed by dose mimicking, MG = manually generated.

Volume	Constraint	EG_init plans /10	EG_upd_ selected/ 10	DP+DM /10	MG /10
Anal canal	V60Gy (%)	9	8	9	10
	V35Gy (%)	8	8	9	10
	V20Gy (%)	9	10	10	10
Bladder	V60Gy (%)	10	9	10	10
	V50Gy (%)	10	10	10	10
	V40.8Gy (%)	6	6	7	9
Bowel	D0.1cm³ (Gy)	9	9	10	10
	V55Gy (%)	10	10	10	10
	V20Gy (cm³)	10	10	10	10
Rectum	V60Gy (%)	8	6	10	10
	V52.8Gy (%)	7	7	7	7
	V30Gy (%)	7	7	7	8
Femur_L	V40Gy (%)	10	10	10	10
	V35Gy (%)	10	10	10	10
Femur_R	V40Gy (%)	10	10	10	10
	V35Gy (%)	10	10	10	10
Penile Bulb	V54.1Gy (%)	9	9	8	10
	V42.5Gy (%)	10	10	10	10

A. Supplementary materials Chapter 2

B. Supplementary materials chapter 3

1. S1. Deformation process.

The deformation process consists in defining an initial mesh represented by six control points, two for each spatial direction, placed on a sphere that circumscribes the volume of interest. We define a deformation scenario S_d by the displacement of a single control point while conserving the initial position of the other control points. Displacing a single control point results thus in deforming the initial mesh along a specific direction, which in turn deform the volume of interest.

Figure S1 gives an example of 2D deformation scenarios generation for 3 directions and 3 amplitudes. In this paper, 18 different directions of deformation and 9 amplitudes were used resulting in 162 deformation scenarios.

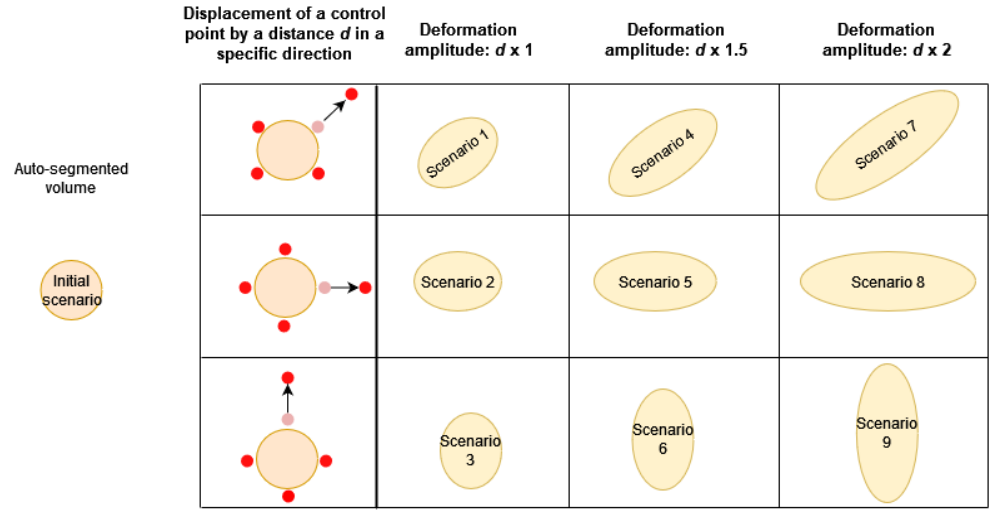
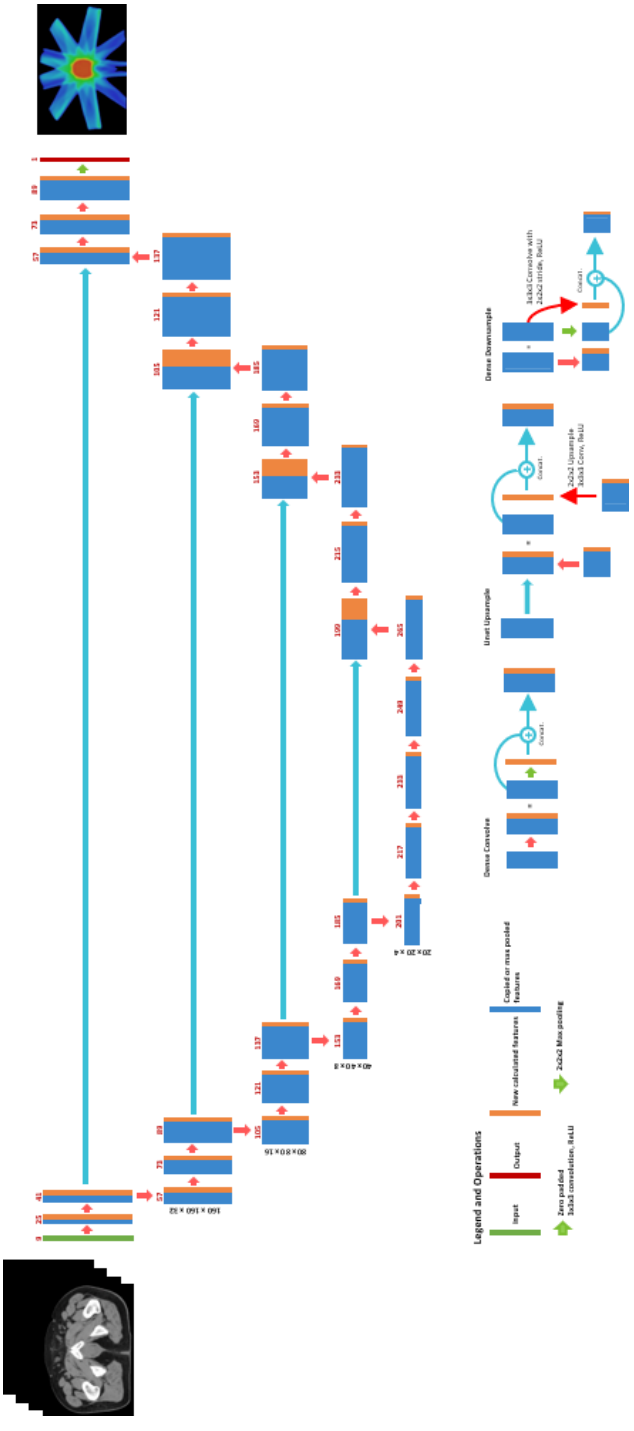


Figure S1: Illustration of deformation scenarios for a 2D case with 4 control points (red dots on the figure). Here, by making control point displacement in 3 directions and 3 amplitudes we obtained 9 deformation scenarios.

2. S2. Dose prediction model.

Figure 33 illustrates the structure of the HD-UNET used for the dose prediction. There are 9 input channels: 1 for dose prescription mask, which consists of the prescription dose for the voxels inside the PTV prostate (60Gy), PTV seminal vesicles (48Gy) and 0 for voxels outside; and 7 for the OARs (the anal canal, bladder, rectum, left femoral head, right femoral head, bowel and penile bulb). Convolutional operations were performed with 16 filters and max-pooling was used to lower the resolution by a factor of two (2x2x2).

The dose distribution matrix, and OAR masks were interpolated to the CBCT grid. Our HD-UNET was trained using image patches with a size of [320 x 320 x 64]. Data were augmented by flipping training data in the left-right direction to double the dataset size. We trained it for 200 epochs and kept the model weights of the last computed epoch.



Structure of HD-UNET used for dose prediction.

Figure 33

3. S3, IOC details.

The α values used as weight for IOC computations are depicted on Table S3. Those were computed on the 180 fractions also used as training data for the DCNN dose prediction. When the dose objectives were well met, α_i were halved.

Table S3: α values for IOC computations for each of the considered metrics.

Volume	Metric	α	Metric condition to use α/2
CTV prostate	D98%	1.78	>58.8Gy
CTV seminal vesicles	D98%	1.90	>47.04Gy
PTV prostate	D95%	1.31	>57Gy
PTV seminal vesicles	D95%	1.69	>45.6Gy
Rectum	D1%	1.65	<59Gy
	V52Gy	0.56	<8Gy
	V30Gy	0.37	<25Gy
Anal canal	D1%	0.49	<59Gy
Bladder	D1%	1.76	<59Gy
	V50	0.28	<12Gy
	V40	0.20	<15Gy

4. S4, Aggregation process of IOC values.

Figure S4 illustrates the aggregation process of the different IOC according to their deformation scenario. The aggregation process begins with the lowest deformation amplitude. The space is filled with the IOC value of each deformation scenario. When deformed masks overlap, the space is filled with averaged IOC. Once all the deformation scenario of the deformation amplitude has been used, the filled voxels are locked and can no longer be updated. The process continues with the deformation scenarios of the second lowest deformation amplitude and so on.

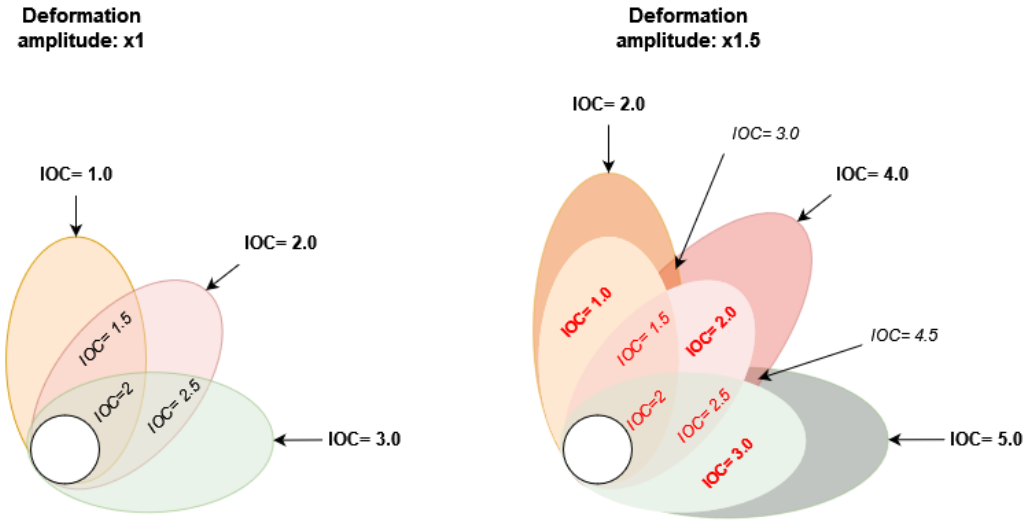


Figure S4: Illustration of DIVE-map aggregation for 3 directions of deformation and 2 amplitudes. Bolded IOC values are computed following the presented formula, italic ones are IOC average of overlapping scenarios. Once all the computations for an amplitude are done, all the updated voxels are locked. Those are represented by the red IOC values. Then, computations for the following deformation amplitude are performed and so on. Note that here in this example, contours color only helps to differentiate the scenarios and does not represent any dosimetric value.

5. S5, Amount of corrections per fraction and patient

This appendix reports the detailed results for the number of corrections indicated by the DIVE-map per fraction, patient, organs (bladder and rectum), and considered threshold (t). Table S5 and S6 present the corrections with respect to the number of voxels (I.e. a value equal to 100% means that the DIVE-map indicates to correct 100% of the voxels); while Tables S7 and S8 present the corrections with respect to the number of slices (I.e. a value equal to 100% means that the DIVE-map indicates to correct 100% of the slices).

		Threshold		
	Fraction	**1**	**1.5**	**2**
Patient 1	**1**	99.82	78.37	75.20
	2	100.00	73.09	61.50
	3	94.59	59.52	7.57
	4	100.00	82.54	79.64
	5	100.00	76.58	76.58
	6	92.28	71.57	30.33
	7	100.00	65.78	50.10
	8	66.06	15.82	0.00
	9	89.81	42.08	24.38
	10	81.02	81.00	48.11
Patient 2	**1**	75.78	75.17	74.61
	2	74.07	46.52	46.52
	3	0.00	0.00	0.00
	4	41.54	41.54	41.54
	5	47.10	0.00	0.00
	6	39.89	27.51	17.56
	7	77.97	66.12	13.33
	8	59.29	24.20	6.33
	9	72.96	56.47	56.47
	10	50.78	50.78	20.60
Patient 3	**1**	58.93	40.41	0.00
	2	58.31	13.73	0.00
	3	87.09	81.64	68.24
	4	0.00	0.00	0.00
	5	41.74	0.00	0.00
	6	68.02	26.40	0.00
	7	77.63	63.86	32.48
	8	47.09	0.00	0.00
	9	66.85	13.35	13.35
	10	46.83	0.00	0.00

Table S34: details of discarded corrections in term of relative rectum correction voxels. A value of 100% indicates that all corrections were operated while 0% indicates that all corrections were discarded.

		Threshold		
	Fraction	**1**	**1.5**	**2**
Patient 1	**1**	39.19	0.00	0.00
	2	55.71	37.70	12.87
	3	98.36	91.20	13.33
	4	80.54	36.40	0.00
	5	86.25	58.02	0.00
	6	51.06	29.52	0.00
	7	49.31	34.62	0.00
	8	80.96	21.31	12.02
	9	61.43	47.30	6.49
	10	99.31	87.38	11.67
Patient 2	**1**	0.00	0.00	0.00
	2	0.00	0.00	0.00
	3	0.00	0.00	0.00
	4	0.00	0.00	0.00
	5	8.81	0.00	0.00
	6	17.98	0.00	0.00
	7	15.66	0.00	0.00
	8	19.30	0.00	0.00
	9	10.72	0.00	0.00
	10	0.00	0.00	0.00
Patient 3	**1**	91.57	74.66	74.66
	2	96.63	93.70	90.16
	3	87.05	80.91	77.53
	4	97.07	94.37	87.10
	5	91.15	84.49	73.13
	6	49.20	34.29	34.29
	7	88.15	83.42	65.09
	8	75.92	65.29	59.66
	9	88.35	80.91	80.91
	10	92.49	83.63	80.70

Table S35: details of discarded corrections in terms of relative bladder correction voxels. A value of 100% indicates that all corrections were operated while 0% indicates that all corrections were discarded.

		Threshold		
	Fraction	**1**	**1.5**	**2**
Patient 1	**1**	78.38	51.35	43.24
	2	100.00	60.00	50.00
	3	72.22	38.89	11.11
	4	100.00	68.18	63.64
	5	100.00	63.64	63.64
	6	77.78	55.56	22.22
	7	100.00	50.00	20.00
	8	43.75	6.25	0.00
	9	66.67	37.04	33.33
	10	61.54	53.85	30.77
Patient 2	**1**	69.23	53.85	46.15
	2	56.25	37.50	37.50
	3	0.00	0.00	0.00
	4	40.00	40.00	40.00
	5	38.46	0.00	0.00
	6	52.94	35.29	29.41
	7	52.63	47.37	21.05
	8	41.46	14.63	7.32
	9	42.86	35.71	35.71
	10	47.06	47.06	11.76
Patient 3	**1**	42.86	28.57	0.00
	2	50.00	12.50	0.00
	3	53.12	46.88	37.50
	4	0.00	0.00	0.00
	5	28.57	0.00	0.00
	6	35.71	14.29	0.00
	7	71.43	57.14	28.57
	8	33.33	0.00	0.00
	9	50.00	12.50	12.50
	10	20.00	0.00	0.00

Table S36: details of discarded corrections in terms of relative rectum correction slices. A value of 100% indicates that all slices initially corrected were corrected by DIVE, while 0% indicates that no slice was corrected by DIVE.

		Threshold		
	Fraction	**1**	**1.5**	**2**
Patient 1	**1**	12.00	0.00	0.00
	2	48.15	14.81	3.70
	3	90.91	63.64	9.09
	4	47.37	15.79	0.00
	5	64.29	21.43	0.00
	6	37.14	14.29	0.00
	7	31.58	10.53	0.00
	8	51.85	11.11	7.41
	9	32.43	21.62	2.70
	10	88.89	61.11	5.56
Patient 2	**1**	0.00	0.00	0.00
	2	0.00	0.00	0.00
	3	0.00	0.00	0.00
	4	0.00	0.00	0.00
	5	3.45	0.00	0.00
	6	6.90	0.00	0.00
	7	7.14	0.00	0.00
	8	6.06	0.00	0.00
	9	3.23	0.00	0.00
	10	0.00	0.00	0.00
Patient 3	**1**	67.86	39.29	39.29
	2	68.97	58.62	51.72
	3	59.38	50.00	46.88
	4	84.00	76.00	64.00
	5	48.48	33.33	24.24
	6	21.05	10.53	10.53
	7	62.96	55.56	40.74
	8	46.43	35.71	32.14
	9	75.00	66.67	66.67
	10	60.00	43.33	40.00

Table S37: details of discarded corrections in term of relative bladder correction slices. A value of 100% indicates that all slice initially corrected were corrected by DIVE, while 0% indicates that no slice was corrected by DIVE.

6. S6. performances of the dose prediction model.

This appendix reports performance metrics comparing the dose distributions provided by our model and the ETHOS system (ground truth doses) for different cases: scenarios with raw (uncorrected) auto-segmented contours; scenarios with ground truth contours (auto-segmented by ETHOS and corrected by clinicians), and scenarios where selected (extreme) deformations were applied to contours, used later to generate the DIVE map.

1) Performance on raw (uncorrected) auto-segmented contours and ground truth contours (auto-segmented by ETHOS and corrected by clinicians)

Table S9 and S10 report the mean absolute error of our dose prediction model in metrics involved in the IOC computation for scenarios with raw auto-segmented and ground truth contours, respectively. Globally, the errors are quite low as well as the standard deviation: all metrics (except rectum V30Gy for ground truth contours, Table S10) show a mean absolute error below 2Gy. An example of DVH for predicted dose against ground truth dose for a raw auto-segmented case (no correction at all) is displayed in figure S11. On this example, the dose to the penile bulb presents some discrepancies due to the small volume of such organ.

Table S9: Mean absolute error of our dose prediction model for dose-volume metrics involved in the computation of the IOC over the 30 fractions with raw auto-segmented contours.

Volume	Metric	Mean Absolute Error (|Ground truth – dose prediction|)
CTV prostate	D98% (Gy)	0.74 ± 0.62
CTV seminal vesicles	D98% (Gy)	0.26 ± 0.41
PTV prostate	D95% (Gy)	1.26 ± 0.83
PTV seminal vesicles	D95% (Gy)	0.55 ± 0.74
Rectum	D1% (Gy)	0.21 ± 0.35
	V52Gy (%)	1.59 ± 1.02
	V30Gy (%)	0.77 ± 3.79
Anal canal	D1% (Gy)	0.14 ± 1.21
Bladder	D1% (Gy)	0.21 ± 0.30
	V50Gy (%)	0.65 ± 0.92
	V40Gy (%)	0.93 ± 1.39

Table S10: Mean absolute error of our dose prediction model for dose-volume metrics involved in the computation of the IOC over the 30 fractions with ground truth contours (auto-segmented + manual correction).

Volume	Metric	Mean Absolute Error (|Ground truth – dose prediction|) **(Gy)**
CTV prostate	D98% (Gy)	0.55 ± 0.55
CTV seminal vesicles	D98% (Gy)	0.14 ± 0.32
PTV prostate	D95% (Gy)	0.96 ± 0.84
PTV seminal vesicles	D95% (Gy)	0.36 ± 0.68
Rectum	D1% (Gy)	-0.06 ± 0.53
	V52Gy (%)	1.68 ± 1.81
	V30Gy (%)	-2.42 ± 6.56
Anal canal	D1% (Gy)	0.24 ± 1.05
Bladder	D1% (Gy)	0.18 ± 0.35
	V50Gy (%)	-0.17 ± 1.93
	V40Gy (%)	-0.98 ± 3.28

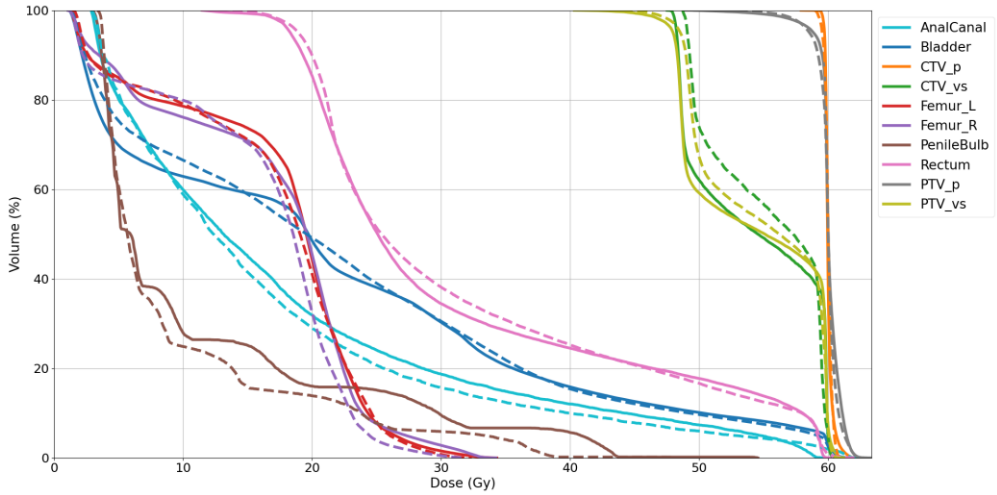


Figure S11: example of DVH for a predicted dose against ground truth dose for auto-segmented contours.

2) Performance on deformed contours

To illustrate the performance of our dose prediction model on scenarios with deformed contours, we start with an example of a deformed bladder and rectum and their corresponding DVHs in Figure S12 and S13, respectively. The applied deformation in both cases is the largest one used for these volumes in the posterior direction during the DIVE-map generation. We can notice a decrease in the quality of dose prediction for such a large deformation scenario. The main reason is probably that those deformed contours are very different to what can be found in the training database, where only non-deformed contours were included. However, what truly matters for this study is the difference in relevant dose-volume metrics, i.e. the quality of the IOCs associated to the predicted dose (see Figure S14). Indeed, extracting dose prediction performance metrics over all the deformation scenarios is too time consuming, since it requires to generate a ground truth dose with the ETHOS system for each scenario: 162 per DIVE-map, thus 324 per fraction. Instead, we propose to illustrate the types of problems that can arise when computing the IOC with predicted doses on deformed contours. We generated large deformations for the 6 spatial axis and analyzed the IOC corresponding to the predicted doses with respect to ground truth ones.

On the left side of figure S14 (bladder IOC), we can notice a large over-estimation of the IOC for predicted doses on contours with a large deformation in the caudal direction (#6). In such direction the bladder is expected to cover a significant part of the PTV, a large dosimetric change would then logically be observed to minimize as possible the dose to the bladder (as it was predicted by the dose prediction model). However, Ethos IOE ended up with a noticeably smaller dosimetric change. This could indicate that in such situation, there might not exist any good large dosimetric change that could strongly decrease the dose to bladder while not compromising other optimization goals. Such IOE behavior might thus not be correctly learned by the dose prediction model. On the contrary, on the right side of figure S14 (rectum IOC), we can notice the under-estimation of the IOC for deformations on the posterior and anterior directions (#1 and #2). Deformation in the posterior direction might cause to create a large proportion of rectum far from PTV, allowing the IOE to actually deliver more dose to the rectum to help satisfy other optimization goals. For the anterior direction, this deformation could cause to increase the volume of rectum in the high dose levels region. Given the rectum high priority of optimization, the optimizer might try to find another existing dose distribution in order to spare at most this deformed rectum. Again, these behaviors and/or the amplitudes of the dosimetric changes are not well learned by the dose prediction model for such large deformations.

Large IOC over-estimation might cause unnecessary corrections, while under-estimation might cause to discard relevant corrections. Although large errors in computing IOC from predicted doses do appear, our DIVE-map nevertheless showed quite good results. This might be explained first by the IOC aggregation method we used. Indeed, the aggregation consists in spatially averaging the different IOC where their corresponding deformation mask overlap. Such averaging could cause to mitigate the isolated IOC errors, thus making the DIVE-map more robust against dose prediction errors.

However, this averaging effect does not explain why our results did show some unnecessary corrections in specific slices while all important slices with dosimetrically relevant corrections were detected (considering threshold t=1.5). That could be explained by a second element: the corrections method used in this

study, i.e. full correction of the slice when any misclassified voxel[1] has a DIVE value greater than *t*. For a given slice, the region with misclassified voxels often spatially overlaps with several deformation scenarios. Consequently, since a single misclassified voxel with value higher than *t* can trigger the whole slice correction, IOC under-estimation for isolated deformed scenarios did not impact the whole correction process itself, as long as there exists other scenarios in the region where the IOC is not under-estimated. However, if one wishes to operate intra-slice corrections, the current quality of our dose prediction could prove to be not sufficient to detect dosimetrically relevant corrections.

To achieve higher robustness in the computation of the IOC from predicted doses, several options are possible. For instance, one could add some ground truth doses for scenarios with deformed contours in the training of the dose prediction model. Another possibility could be to replace the dose prediction model by a real auto-planner (like the Ethos IOE) in the DIVE generation process. The dose distributions for deformed scenarios would be generated much slower than with dose prediction, but there will be no associated uncertainty in dose distribution. In addition, after generating DIVE-maps with ground truth doses for several treatment fractions from several patients, one could build a large database and use it for training a DCNN model (e.g. an HD-UNET) to directly predict DIVE-map. This would generate a DIVE-map in a matter of seconds and thus solve the timing problem, although in this case there will be an uncertainty associated to the predicted DIVE-maps instead of the predicted doses.

[1] From section 2.3, we define a "misclassified" voxel when the auto-segmenter did not assign the correct label to this specific voxel: either auto-segmentation considered this voxel as belonging to the nearby OAR while it should not or, conversely, auto-segmentation did not include this voxel into the OAR while it actually belongs to it.

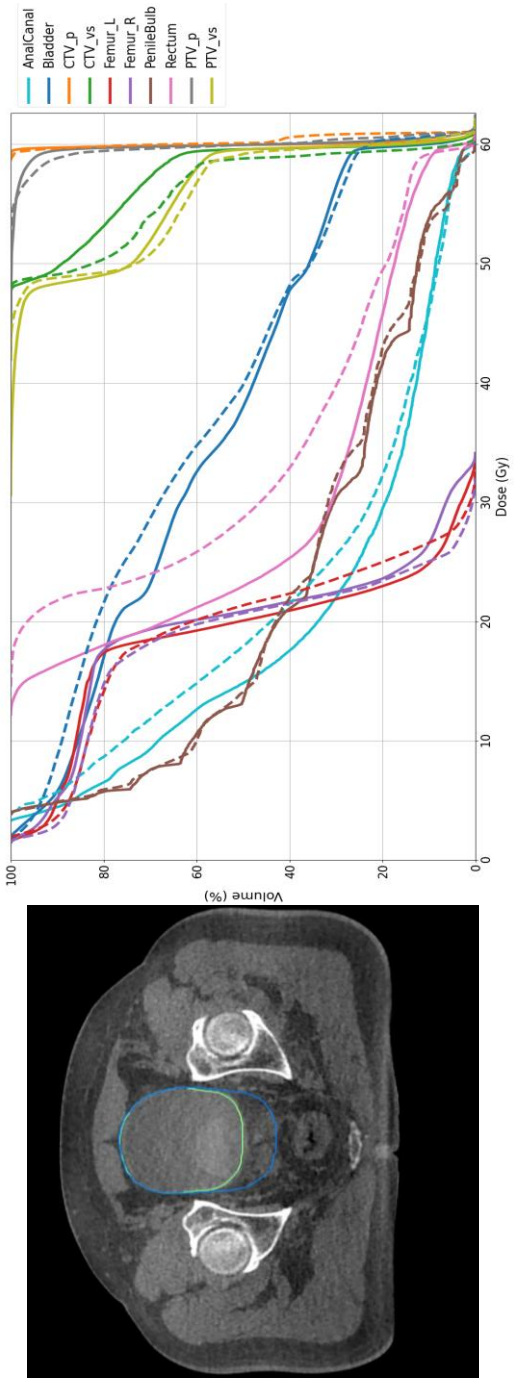


Figure S12. Left: Illustration of bladder deformation in the posterior axis with the largest used amplitude. Light green: auto-segmented contour; Blue: deformed contour. Right: DVH of dose prediction for such bladder deformation. Solid lines are for the ground truth dose; Dashed lines are for the predicted dose

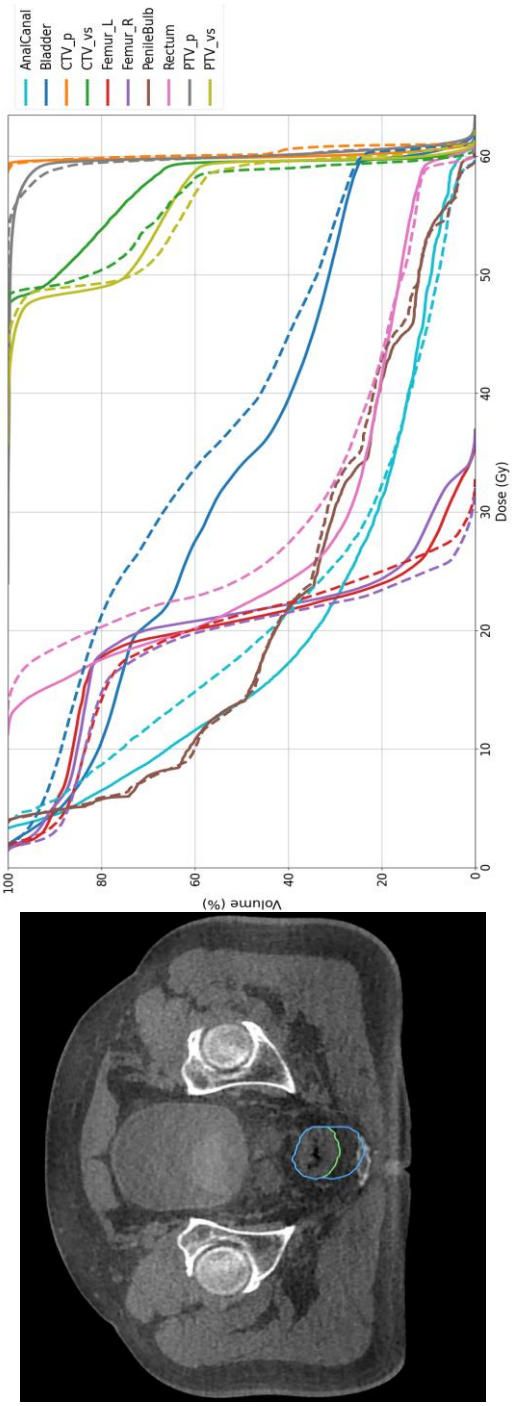


Figure S13. Left: Illustration of rectum deformation in the posterior axis with the largest used amplitude. Light green: auto-segmented contour; Blue: deformed contour. Right: DVH of dose prediction for such rectum deformation. Solid lines are for the ground truth dose; Dashed lines are for the predicted dose.

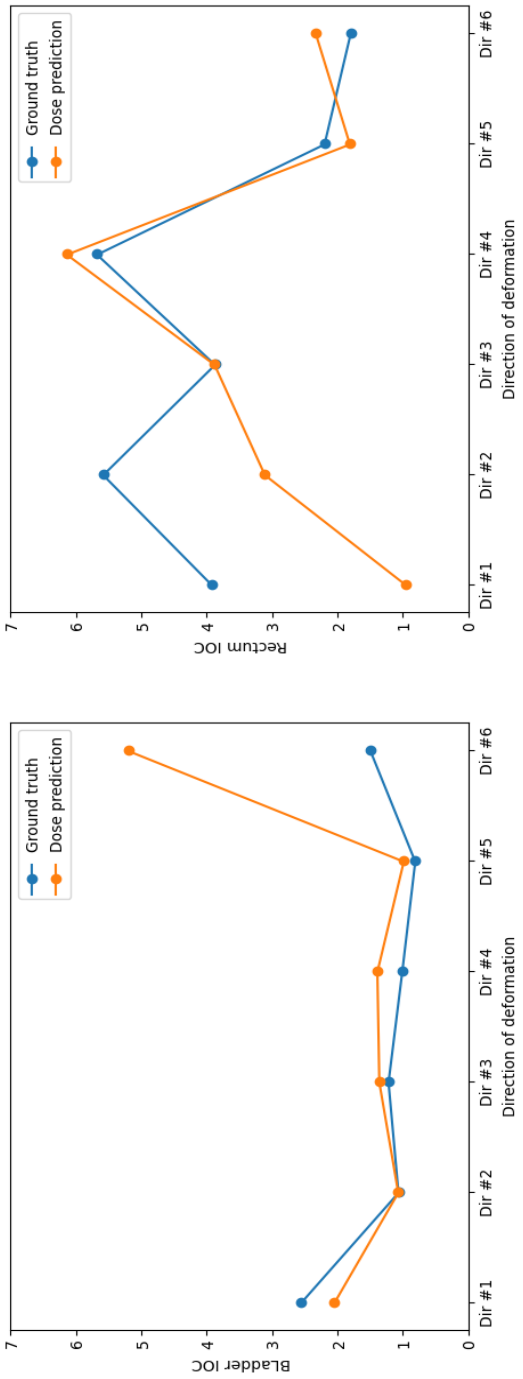


Figure S14: Illustration for one adaptive fraction of the quality of the predicted IOCs and problems that can occur with our dose prediction. The ground truth IOCs are generated with Ethos IOE using the deformed rectum/bladder. The amplitude of the deformation is the largest used to generate the DIVE-map in this study. The 6 directions of deformation consist in the conventional anatomical axis. Dir #1: direction posterior; Dir #2: direction anterior; Dir#3: direction left; Dir#4: direction right; Dir#5: direction cranial; Dir#6: direction caudal.

C. Supplementary materials Chapter 4

1. S1 Details on robust evaluation

Table S1 present the worst-case value on the Dmean for the different anatomical volumes.

Volume	Metric	**Dosimetric-adaptation minus Reference**	
		Mean (Gy)	**Range (Gy)**
CTV_p	D_{98}	-0.43	[-0.76, -0.19]
CTV_p	D_2	0.42	[-0.12, 0.95]
CTV_n	D_{98}	-0.42	[-0.68, 0.00]
$CTV_{n,high}$	D_{98}	-0.31	/
CTV_{n_high}	D_2	0.37	/
SC	D_{max}	-1.02	[-3.12, 1.29]
Body	D_{1cc}	0.78	[-0.36, 1.93]
Mandible	D_{mean}	0.26	[-8.06, 4.11]
Parotid Left	D_{mean}	0.51	[-1.70, 1.88]
Parotid Right	D_{mean}	0.48	[-0.60, 2.33]
Thyroid	D_{mean}	-0.47	[-3.00, 1.45]
Esophagus	D_{mean}	-0.88	[-4.67, 3.66]
Larynx	D_{mean}	0.83	[-0.51, 2.37]
Oral cavity	D_{mean}	-0.41	[-3.57, 1.27]

Table S1: Critical dose volume-metrics difference between efficiently-adapted plans and reference plans on CT2 for worst case. SC: spinal cord.

2. S2 Details on spots selection criteria

Table S2.1 provides the size of the selected subset of spots for the different selection criteria. S1 refers to our first selection criterion, i.e. of the smallest subset of spots carrying at least 65% of the total spot weight; S2 refers to our second selection criterion, $\frac{\partial \mathrm{F}}{\partial w_i} > \mathrm{t}$; original criterion (OC) refers to the one use in the original paper from *Botas et al.,* the smallest subset of spots carrying at least 50% of the total spot weight. We can notice that the combination of both S1 and S2 (S1+S2) led to a larger spot selection than with OC.

Table S2.1: Spot selection according the different criteria. First column refers to the selection of the smallest subset of spots carrying at least 65% of the total spot weight. Second column refers to the spot selection when adding the second criterion to the first one. Third column refers to the spots that would be selected if the same criterion than reference paper from Botas et al. would have been chosen, i.e. the smallest subset of spots carrying at least 50% of the total weight.

	% spots kept after selection of initial weight (S1)	% spots kept after S1 and objective function gradient constraint (S1+S2)	% spots selection *Botas et al.* criterion (OC)
P1	55.6	46.6	39.8
P2	57.5	46.8	42.1
P3	53.6	47.4	38.1
P4	50.7	48.0	35.5
P5	55.2	49.5	39.7

Table S20.2: Dose volume-metrics difference between efficiently-adapted plans and reference plans on CT2 for nominal case. SC: spinal cord

Volume	**Metric**	smart-adaptation minus Reference		OC-adaptation minus Reference	
		Mean (Gy)	**Range (Gy)**	**Mean (Gy)**	**Range (Gy)**
CTV_p	D_{98}	-0.25	[-0.59, 0.06]	-0.27	[-0.65, 0.10]
CTV_p	D_2	0.24	[-0.05, 0.66]	0.21	[-0.04, 0.79]
CTV_n	D_{98}	-0.38	[-0.50, -0.07]	-0.44	[-0.60, -0.29]
$CTV_{n,high}$	D_{98}	-0.41	/	-0.20	/
$CTV_{n,high}$	D_2	-0.03	/	0.03	/
SC	D_{max}	-1.74	[-4.97, 1.16]	-0.95	[-4.75, 2.22]
SC	D_{mean}	-0.07	[-1.60, 0.80]	0.35	[-0.50,1.87]
Mandible	D_{mean}	-0.29	[-8.70, 4.39]	0.22	[-8.84,4.61]
Parotid_L	D_{mean}	0.00	[-3.56, 1.87]	-0.07	[-3.46, 2.26]
Parotid_R	D_{mean}	0.66	[-0.75, 2.76]	-0.08	[-1.90, 2.11]
Thyroid	D_{mean}	-0.12	[-2.70, 1.67]	-0.42	[-2.71, 1.22]
Esophagus	D_{mean}	-1.11	[-3.87, 2.06]	-1.7	[-3.94, 1.40]
Larynx	D_{mean}	0.67	[-1.34, 2.49]	0.96	[-1.25, 3.42]
Oral Cavity	D_{mean}	-0.40	[-4.13, 1.20]	-0.38	[-3.95, 1.14]

Table S2.3: Critical dose volume-metrics difference between efficiently-adapted plans and reference plans on CT2 for worst case. SC : spinal cord

Volume	Metric	smart-adaptation minus Reference		OC adaptation minus Reference	
		Mean (Gy)	**Range (Gy)**	**Mean (Gy)**	**Range (Gy)**
CTV_p	D_{98}	-0.43	[-0.76, -0.19]	-0.61	[-0.90, -0.32]
CTV_p	D_2	0.42	[-0.12, 0.95]	0.56	[-0.17, 1.27]
CTV_n	D_{98}	-0.42	[-0.68, 0.00]	-0.64	[-1.12, -0.27]
$CTV_{n,high}$	D_{98}	-0.31	/	-0.49	/
CTV_{n_high}	D_2	0.37	/	0.49	/
SC	D_{max}	-1.02	[-3.12, 1.29]	-0.32	[-1.88, 1.90]
Body	D_{1cc}	0.78	[-0.36, 1.93]	0.85	[-0.29, 2.05]

Table S2.2 is the extension of table 2 with additional dosimetric value achieved with the OC criterion for nominal case. Table S2.3 is the table 3 with additional dosimetric value achieved with the OC criterion for worst-case.

While there is not much difference between our selection criteria (S1+S2) and the original one (OC) on the nominal case as illustrated on table S2.2, for the worst-case scenario on table S2.3, dosimetric gap between both approaches is a bit larger. OC consistently provided a less homogeneous target coverage, a higher D_{max} to SC and D1cc to body.

Such dosimetric differences were sometimes enough to fall below expected minimal target coverage i.e >66.5Gy for CTV_p and >51.54Gy for CTV_n, while it was not the case for the S1+S2 selection. Indeed, for CTV_p coverage 1/5 patient met the target coverage goal, for CTV_n 2/5 and for $CTV_{n,high}$ 0/1.

While the OC selection could theoretically lead to faster optimization due to fewer selected spots, the dosimetric quality seemed to be a bit on the edge of clinical acceptability according to our investigation. This explains why we slightly increased the selection threshold to 65% of the total spot weights, while still trying to refine it a bit with the second criterion we proposed.

Although S1+S2 showed improved dosimetric quality, future studies should evaluate whether other thresholds for S1 and S2 might be more appropriate.

References

1. Bray, F., et al., *Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: a cancer journal for clinicians, 2024. **74**(3): p. 229-263.
2. Mehta, S., et al., *Radiotherapy: basic concepts and recent advances.* Medical Journal Armed Forces India, 2010. **66**(2): p. 158-162.
3. Group, I.M.R.T.C.W., *Intensity-modulated radiotherapy: current status and issues of interest.* International Journal of Radiation Oncology* Biology* Physics, 2001. **51**(4): p. 880-914.
4. Otto, K., *Volumetric modulated arc therapy: IMRT in a single gantry arc.* Medical physics, 2008. **35**(1): p. 310-317.
5. Yu, C.X., *Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy.* Physics in Medicine & Biology, 1995. **40**(9): p. 1435.
6. Brock, K.K., *Adaptive Radiotherapy: Moving Into the Future.* Semin Radiat Oncol, 2019. **29**(3): p. 181-184.
7. Hunt, A., et al., *Adaptive Radiotherapy Enabled by MRI Guidance.* Clin Oncol (R Coll Radiol), 2018. **30**(11): p. 711-719.
8. van Vulpen, M. and L. Wang, *1.12. Within the next five years, adaptive hypofractionation will become the most common form of radiotherapy.* Controversies in Medical Physics: a Compendium of Point/Counterpoint Debates Volume 3, 2017: p. 75.
9. Hickey, B.E., et al., *Hypofractionated radiation therapy for early breast cancer.* Cochrane Database of Systematic Reviews, 2016(7).
10. Koontz, B.F., et al., *A systematic review of hypofractionation for primary management of prostate cancer.* European urology, 2015. **68**(4): p. 683-691.
11. Yan, D., et al., *Adaptive radiation therapy.* Physics in Medicine & Biology, 1997. **42**(1): p. 123.
12. Green, O.L., L.E. Henke, and G.D. Hugo. *Practical clinical workflows for online and offline adaptive radiation therapy*. in *Seminars in radiation oncology*. 2019. Elsevier.
13. Lim-Reinders, S., et al., *Online adaptive radiation therapy.* International Journal of Radiation Oncology* Biology* Physics, 2017. **99**(4): p. 994-1003.
14. McComas, K.N., et al., *Online adaptive radiation therapy and opportunity cost.* Advances in Radiation Oncology, 2023. **8**(3): p. 101034.
15. Newhauser, W.D. and R. Zhang, *The physics of proton therapy.* Physics in Medicine & Biology, 2015. **60**(8): p. R155.

16. Soukup, M., et al., *Study of robustness of IMPT and IMRT for prostate cancer against organ movement.* International Journal of Radiation Oncology* Biology* Physics, 2009. **75**(3): p. 941-949.
17. Góra, J., et al., *ART for head and neck patients: on the difference between VMAT and IMPT.* Acta Oncologica, 2015. **54**(8): p. 1166-1174.
18. Kooy, H.M., et al., *A case study in proton pencil-beam scanning delivery.* International Journal of Radiation Oncology* Biology* Physics, 2010. **76**(2): p. 624-630.
19. Unkelbach, J., et al., *Robust radiotherapy planning.* Physics in Medicine & Biology, 2018. **63**(22): p. 22TR02.
20. Fredriksson, A., *A characterization of robust radiation therapy treatment planning methods—from expected value to worst case optimization.* Medical physics, 2012. **39**(8): p. 5169-5181.
21. Paganetti, H., *Range uncertainties in proton therapy and the role of Monte Carlo simulations.* Physics in Medicine & Biology, 2012. **57**(11): p. R99.
22. Holmes, J., et al., *Fast Monte Carlo dose calculation in proton therapy.* Physics in Medicine & Biology, 2024. **69**(17): p. 17TR01.
23. Mairani, A., et al., *A Monte Carlo-based treatment planning tool for proton therapy.* Physics in Medicine & Biology, 2013. **58**(8): p. 2471.
24. Fredriksson, A., A. Forsgren, and B. Hårdemark, *Minimax optimization for handling range and setup uncertainties in proton therapy.* Medical Physics, 2011. **38**(3): p. 1672-1684.
25. Yang, M., et al., *Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration.* Physics in Medicine & Biology, 2012. **57**(13): p. 4095.
26. Thummerer, A., et al., *Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy.* Physics in Medicine & Biology, 2020. **65**(9): p. 095002.
27. Archambault, Y., et al., *Making on-line adaptive radiotherapy possible using artificial intelligence and machine learning for efficient daily re-planning.* Med Phys Intl J, 2020. **8**.
28. Winkel, D., et al., *Adaptive radiotherapy: the Elekta Unity MR-linac concept.* Clinical and translational radiation oncology, 2019. **18**: p. 54-59.
29. Klüter, S., *Technical design and concept of a 0.35 T MR-Linac.* Clinical and translational radiation oncology, 2019. **18**: p. 98-101.
30. Albertini, F., et al., *First clinical implementation of a highly efficient daily online adapted proton therapy (DAPT) workflow.* Physics in Medicine and Biology, 2024.
31. Mein, S., et al., *Particle arc therapy: Status and potential.* Radiotherapy and Oncology, 2024: p. 110434.

32. Favaudon, V., et al., *Ultrahigh dose-rate FLASH irradiation increases the differential response between normal and tumor tissue in mice.* Science translational medicine, 2014. **6**(245): p. 245ra93-245ra93.
33. Huang, D., et al., *Cost-effectiveness models of proton therapy for head and neck: Evaluating quality and methods to date.* International Journal of Particle Therapy, 2021. **8**(1): p. 339-353.
34. Verma, V., M.V. Mishra, and M.P. Mehta, *A systematic review of the cost and cost-effectiveness studies of proton radiotherapy.* Cancer, 2016. **122**(10): p. 1483-1501.
35. Borderías-Villarroel, E., A. Barragán-Montero, and E. Sterpin, *Time is NTCP: Should we maximize patient throughput or perform online adaptation on proton therapy systems?* Radiotherapy and Oncology, 2024. **198**: p. 110389.
36. Visak, J., et al., *Evaluating machine learning enhanced intelligent-optimization-engine (IOE) performance for ethos head-and-neck (HN) plan generation.* Journal of applied clinical medical physics, 2023: p. e13950.
37. Sibolt, P., et al., *Clinical implementation of artificial intelligence-driven cone-beam computed tomography-guided online adaptive radiotherapy in the pelvic region.* Phys Imaging Radiat Oncol, 2021. **17**: p. 1-7.
38. Khalifa, A., et al., *Machine learning automated treatment planning for online magnetic resonance guided adaptive radiotherapy of prostate cancer.* Physics and Imaging in Radiation Oncology, 2024: p. 100649.
39. Benitez, C.M., et al. *MRI-Guided Adaptive Radiation Therapy*. in *Seminars in Radiation Oncology*. 2024. Elsevier.
40. Stanescu, T., et al., *MRI-guided online adaptive stereotactic body radiation therapy of liver and pancreas tumors on an MR-linac system.* Cancers, 2022. **14**(3): p. 716.
41. Joiner, M.C. and A. van der KogeJ, *Basic clinical radiobiology*. 2025: CRC press.
42. Mohan, R., et al., *Radiobiological considerations in the design of fractionation strategies for intensity-modulated radiation therapy of head and neck cancers.* International Journal of Radiation Oncology* Biology* Physics, 2000. **46**(3): p. 619-630.
43. Gong, L., et al., *Application of radiosensitizers in cancer radiotherapy.* International journal of nanomedicine, 2021: p. 1083-1102.
44. Quan, E.M., et al., *A comprehensive comparison of IMRT and VMAT plan quality for prostate cancer treatment.* International Journal of Radiation Oncology* Biology* Physics, 2012. **83**(4): p. 1169-1178.
45. Sharfo, A.W.M., et al., *Comparison of VMAT and IMRT strategies for cervical cancer patients using automated planning.* Radiotherapy and Oncology, 2015. **114**(3): p. 395-401.

46. Lecchi, M., et al., *Current concepts on imaging in radiotherapy.* European journal of nuclear medicine and molecular imaging, 2008. **35**: p. 821-837.
47. Pereira, G.C., M. Traughber, and R.F. Muzic Jr, *The role of imaging in radiation therapy planning: past, present, and future.* BioMed research international, 2014. **2014**(1): p. 231090.
48. Kessler, M.L., et al., *Integration of multimodality imaging data for radiotherapy treatment planning.* International Journal of Radiation Oncology* Biology* Physics, 1991. **21**(6): p. 1653-1667.
49. Brock, K.K., et al., *Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132.* Medical physics, 2017. **44**(7): p. e43-e76.
50. Czajkowski, P. and T. Piotrowski, *Registration methods in radiotherapy.* Reports of Practical Oncology and Radiotherapy, 2019. **24**(1): p. 28-34.
51. Sharp, G., et al., *Vision 20/20: perspectives on automated image segmentation for radiotherapy.* Medical physics, 2014. **41**(5): p. 050902.
52. Vaassen, F., et al., *Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy.* Physics and Imaging in Radiation Oncology, 2020. **13**: p. 1-6.
53. Burnet, N.G., et al., *Defining the tumour and target volumes for radiotherapy.* Cancer Imaging, 2004. **4**(2): p. 153.
54. Van Herk, M., et al., *The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy.* International Journal of Radiation Oncology* Biology* Physics, 2000. **47**(4): p. 1121-1135.
55. Hansen, C.R., et al., *Plan quality in radiotherapy treatment planning–Review of the factors and challenges.* Journal of Medical Imaging and Radiation Oncology, 2022. **66**(2): p. 267-278.
56. Haas, O.C., *Radiotherapy treatment planning: new system approaches*. 2012: Springer Science & Business Media.
57. Dixon, P. and B. O'Sullivan, *Radiotherapy quality assurance: time for everyone to take it seriously.* European Journal of cancer, 2003. **39**(4): p. 423-429.
58. Lorenz, F., et al., *An independent dose calculation algorithm for MLC-based stereotactic radiotherapy.* Medical physics, 2007. **34**(5): p. 1605-1614.
59. Toscano, S., et al., *Impact of machine log-files uncertainties on the quality assurance of proton pencil beam scanning treatment delivery.* Phys Med Biol, 2019. **64**(9): p. 095021.
60. Waddle, M.R., et al., *Photon and proton radiation therapy utilization in a population of more than 100 million commercially insured patients.*

International Journal of Radiation Oncology* Biology* Physics, 2017. **99**(5): p. 1078-1082.
61. Stein, J., et al., *Dynamic X-ray compensation for conformal radiotherapy by means of multi-leaf collimation.* Radiotherapy and Oncology, 1994. **32**(2): p. 163-173.
62. Jeraj, M. and V. Robar, *Multileaf collimator in radiotherapy.* Radiology and Oncology, 2004. **38**(3).
63. Klish, M.D., G.A. Watson, and D.C. Shrieve, *Radiation and intensity-modulated radiotherapy for metastatic spine tumors.* Neurosurgery Clinics, 2004. **15**(4): p. 481-490.
64. Kryder, C.L., *Proton-beam therapy versus photon-beam therapy: The debate continues.* Radiation Oncology, 2017.
65. St James, S., C. Grassberger, and H.-M. Lu, *Considerations when treating lung cancer with passive scatter or active scanning proton therapy.* Translational lung cancer research, 2018. **7**(2): p. 210.
66. Mishra, M.V., et al., *Proton beam therapy delivered using pencil beam scanning vs. passive scattering/uniform scanning for localized prostate cancer: comparative toxicity analysis of PCG 001-09.* Clinical and translational radiation oncology, 2019. **19**: p. 80-86.
67. Chuong, M., et al., *Pencil beam scanning versus passively scattered proton therapy for unresectable pancreatic cancer.* Journal of gastrointestinal oncology, 2018. **9**(4): p. 687.
68. Paganetti, H., *Proton beam therapy*. 2017: IOP Publishing Bristol.
69. Wase, V., et al., *The proton arc therapy treatment planning problem is NP-Hard.* Computers in Biology and Medicine, 2024. **171**: p. 108139.
70. Wuyckens, S., et al., *Treatment planning in arc proton therapy: Comparison of several optimization problem statements and their corresponding solvers.* Computers in Biology and Medicine, 2022. **148**: p. 105609.
71. Bazalova-Carter, M. and N. Esplen, *On the capabilities of conventional x-ray tubes to deliver ultra-high (FLASH) dose rates.* Medical physics, 2019. **46**(12): p. 5690-5695.
72. Diffenderfer, E.S., et al., *The current status of preclinical proton FLASH radiation and future directions.* Medical physics, 2022. **49**(3): p. 2039-2054.
73. Zhang, Q., et al., *FLASH investigations using protons: design of delivery system, preclinical setup and confirmation of FLASH effect with protons in animal systems.* Radiation research, 2020. **194**(6): p. 656-664.
74. Hughes, J.R. and J.L. Parsons, *FLASH radiotherapy: current knowledge and future insights using proton-beam therapy.* International journal of molecular sciences, 2020. **21**(18): p. 6492.

75. Fippel, M. and M. Soukup, *A Monte Carlo dose calculation algorithm for proton therapy.* Medical physics, 2004. **31**(8): p. 2263-2273.
76. De Martino, F., et al., *Dose calculation algorithms for external radiation therapy: an overview for practitioners.* Applied sciences, 2021. **11**(15): p. 6806.
77. Sievinen, J., W. Ulmer, and W. Kaissl, *AAA photon dose calculation model in Eclipse.* Palo Alto (CA): Varian Medical Systems, 2005. **118**: p. 2894.
78. Ahnesjö, A., *Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media.* Medical physics, 1989. **16**(4): p. 577-592.
79. Mackie, T., et al., *Generation of photon energy deposition kernels using the EGS Monte Carlo code.* Physics in Medicine & Biology, 1988. **33**(1): p. 1.
80. Vassiliev, O.N., et al., *Validation of a new grid-based Boltzmann equation solver for dose calculation in radiotherapy with photon beams.* Physics in Medicine & Biology, 2010. **55**(3): p. 581.
81. Schaffner, B., E. Pedroni, and A. Lomax, *Dose calculation models for proton treatment planning using a dynamic beam delivery system: an attempt to include density heterogeneity effects in the analytical dose calculation.* Physics in Medicine & Biology, 1999. **44**(1): p. 27.
82. Van Herk, M. *Errors and margins in radiotherapy*. in *Seminars in radiation oncology*. 2004. Elsevier.
83. Fiorino, C., et al., *Intra-and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning.* Radiotherapy and oncology, 1998. **47**(3): p. 285-292.
84. Engelsman, M. and H.M. Kooy, *Target volume dose considerations in proton beam treatment planning for lung tumors.* Medical physics, 2005. **32**(12): p. 3549-3557.
85. Liu, W., et al., *Effectiveness of robust optimization in intensity-modulated proton therapy planning for head and neck cancers.* Medical physics, 2013. **40**(5): p. 051711.
86. Liu, W., et al., *Robust optimization of intensity modulated proton therapy.* Medical physics, 2012. **39**(2): p. 1079-1091.
87. Chu, M., et al., *Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty.* Physics in Medicine & Biology, 2005. **50**(23): p. 5463.
88. Ripsman, D.A., et al., *Light Pareto robust optimization for IMRT treatment planning.* Medical Physics, 2023. **50**(5): p. 2637-2648.
89. Verellen, D., et al., *Innovations in image-guided radiotherapy.* Nature Reviews Cancer, 2007. **7**(12): p. 949-960.

90. Freislederer, P., et al., *Recent advances in surface guided radiation therapy.* Radiation Oncology, 2020. **15**: p. 1-11.
91. Neicu, T., et al., *Synchronized moving aperture radiation therapy (SMART): improvement of breathing pattern reproducibility using respiratory coaching.* Physics in Medicine & Biology, 2006. **51**(3): p. 617.
92. Cervino, L.I., et al., *Using surface imaging and visual coaching to improve the reproducibility and stability of deep-inspiration breath hold for left-breast-cancer radiotherapy.* Physics in Medicine & Biology, 2009. **54**(22): p. 6853.
93. Van Ooteghem, G., et al., *Mechanically-assisted non-invasive ventilation: A step forward to modulate and to improve the reproducibility of breathing-related motion in radiation therapy.* Radiother Oncol, 2019. **133**: p. 132-139.
94. Vander Veken, L., et al., *Lung and Liver Stereotactic Body Radiation Therapy During Mechanically Assisted Deep Inspiration Breath-Holds: A Prospective Feasibility Trial.* Advances in Radiation Oncology, 2024. **9**(9): p. 101563.
95. Giraud, P. and A. Houle, *Respiratory gating for radiotherapy: main technical aspects and clinical benefits.* International Scholarly Research Notices, 2013. **2013**(1): p. 519602.
96. Giraud, P., et al., *Respiratory gating techniques for optimization of lung cancer radiotherapy.* Journal of thoracic oncology, 2011. **6**(12): p. 2058-2068.
97. Shirato, H., et al., *Physical aspects of a real-time tumor-tracking system for gated radiotherapy.* International Journal of Radiation Oncology* Biology* Physics, 2000. **48**(4): p. 1187-1195.
98. Buzurovic, I., et al., *A robotic approach to 4D real-time tumor tracking for radiotherapy.* Physics in Medicine & Biology, 2011. **56**(5): p. 1299.
99. Cervino, L.I., J. Du, and S.B. Jiang, *MRI-guided tumor tracking in lung cancer radiotherapy.* Physics in Medicine & Biology, 2011. **56**(13): p. 3773.
100. Tison, T., et al., *Online Adaptive Radiotherapy for Planning Target Volume (PTV) Reduction in Gastric Mucosa-Associated Lymphoid Tissue (MALT) Lymphoma.* Cureus, 2024. **16**(9): p. e68919.
101. Dupont, F., D. Dechambre, and E. Sterpin, *Evaluation of safety margins for cone beam CT-based adaptive prostate radiotherapy.* Physica Medica, 2024. **121**: p. 103368.
102. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. 2015. Springer.

103. Nguyen, D., et al., *Three-dimensional radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected u-net deep learning architecture.* arXiv preprint arXiv:1805.10397, 2018.
104. Isensee, F., et al., *nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation.* Nature methods, 2021. **18**(2): p. 203-211.
105. Choi, Y., et al., *Deep learning–based multimodal segmentation of oropharyngeal squamous cell carcinoma on CT and MRI using self-configuring nnU-Net.* European Radiology, 2024: p. 1-12.
106. Gay, S.S., et al., *Fully-automated, CT-only GTV contouring for palliative head and neck radiotherapy.* Scientific reports, 2023. **13**(1): p. 21797.
107. Xu, D., et al., *Fully automated segmentation of prostatic urethra for MR-guided radiation therapy.* Medical Physics, 2023. **50**(1): p. 354-364.
108. Goodfellow, I., et al., *Generative adversarial nets.* Advances in neural information processing systems, 2014. **27**.
109. Salimans, T., et al., *Improved techniques for training gans.* Advances in neural information processing systems, 2016. **29**.
110. Croitoru, F.-A., et al., *Diffusion models in vision: A survey.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023. **45**(9): p. 10850-10869.
111. Ma, Y., et al., *Structure and illumination constrained GAN for medical image enhancement.* IEEE Transactions on Medical Imaging, 2021. **40**(12): p. 3955-3967.
112. Singh, N.K. and K. Raza, *Medical image generation using generative adversarial networks: A review.* Health informatics: A computational perspective in healthcare, 2021: p. 77-96.
113. Feng, Z., et al. *Diffdp: Radiotherapy dose prediction via a diffusion model*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2023. Springer.
114. Zhang, Y., et al., *DoseDiff: distance-aware diffusion model for dose prediction in radiotherapy.* IEEE Transactions on Medical Imaging, 2024.
115. Sun, J., et al., *DiffRecon: Diffusion-based CT reconstruction with cross-modal deformable fusion for DR-guided non-coplanar radiotherapy.* Computers in Biology and Medicine, 2024. **179**: p. 108868.
116. Bao, P., et al., *Deep Reinforcement Learning for Beam Angle Optimization of Intensity-Modulated Radiation Therapy.* arXiv preprint arXiv:2303.03812, 2023.
117. Van der Veen, J., et al., *Benefits of deep learning for delineation of organs at risk in head and neck cancer.* Radiotherapy and Oncology, 2019. **138**: p. 68-74.

118. Aselmaa, A., et al., *The influence of automation on tumor contouring.* Cognition, Technology & Work, 2017. **19**(4): p. 795-808.
119. Fogliata, A., et al., *RapidPlan knowledge based planning: iterative learning process and model ability to steer planning strategies.* Radiation Oncology, 2019. **14**: p. 1-12.
120. Barragan-Montero, A.M., et al., *Three-dimensional dose prediction for lung IMRT patients with deep neural networks: robust learning from heterogeneous beam configurations.* Med Phys, 2019. **46**(8): p. 3679-3691.
121. Liu, S., et al., *A cascade 3D U-Net for dose prediction in radiotherapy.* Medical Physics, 2021. **48**(9): p. 5574-5582.
122. Babier, A., et al., *The importance of evaluating the complete automated knowledge-based planning pipeline.* Phys Med, 2020. **72**: p. 73-79.
123. McIntosh, C., et al., *Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method.* Phys Med Biol, 2017. **62**(15): p. 5926-5944.
124. Vandewinckele, L., et al., *Treatment plan prediction for lung IMRT using deep learning based fluence map generation.* Physica Medica, 2022. **99**: p. 44-54.
125. Vandewinckele, L., et al., *Deep learning based MLC aperture and monitor unit prediction as a warm start for breast VMAT optimisation.* Physics in Medicine & Biology, 2023. **68**(22): p. 225013.
126. Shen, C., et al., *Operating a treatment planning system using a deep-reinforcement learning-based virtual treatment planner for prostate cancer intensity-modulated radiation therapy treatment planning.* Medical physics, 2020. **47**(6): p. 2329-2336.
127. Wang, H., et al., *An integrated solution of deep reinforcement learning for automatic IMRT treatment planning in non-small-cell lung cancer.* Frontiers in Oncology, 2023. **13**: p. 1124458.
128. Liu, Y., et al., *CBCT-based synthetic CT generation using deep-attention cycleGAN for pancreatic adaptive radiotherapy.* Med Phys, 2020. **47**(6): p. 2472-2483.
129. Wodzinski, M., et al., *Semi-supervised deep learning-based image registration method with volume penalty for real-time breast tumor bed localization.* Sensors, 2021. **21**(12): p. 4085.
130. Salehi, S.S.M., et al., *Real-time deep pose estimation with geodesic loss for image-to-template rigid registration.* IEEE transactions on medical imaging, 2018. **38**(2): p. 470-481.
131. Eppenhof, K.A. and J.P. Pluim, *Pulmonary CT registration through supervised learning with convolutional neural networks.* IEEE transactions on medical imaging, 2018. **38**(5): p. 1097-1105.

132. Fan, J., et al., *BIRNet: Brain image registration using dual-supervised fully convolutional networks.* Medical image analysis, 2019. **54**: p. 193-206.
133. Ferreira, A., et al., *Improved Multi-Task Brain Tumour Segmentation with Synthetic Data Augmentation.* arXiv preprint arXiv:2411.04632, 2024.
134. Lei, Y., et al. *Image quality improvement in cone-beam CT using deep learning*. in *Medical imaging 2019: physics of medical imaging*. 2019. SPIE.
135. Zhang, Y., et al., *Improving CBCT quality to CT level using deep learning with generative adversarial network.* Medical physics, 2021. **48**(6): p. 2816-2826.
136. Rusanov, B., et al., *Deep learning methods for enhancing cone-beam CT image quality toward adaptive radiation therapy: A systematic review.* Medical Physics, 2022. **49**(9): p. 6019-6054.
137. Bohannon, D., et al., *Prediction of plan adaptation in head and neck cancer proton therapy using clinical, radiographic, and dosimetric features.* Acta Oncologica, 2023. **62**(6): p. 627-634.
138. Lee, D., et al., *Deep learning driven predictive treatment planning for adaptive radiotherapy of lung cancer.* Radiotherapy and Oncology, 2022. **169**: p. 57-63.
139. Zhen, X., et al., *Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study.* Physics in Medicine & Biology, 2017. **62**(21): p. 8246.
140. Desideri, I., et al., *Application of radiomics for the prediction of radiation-induced toxicity in the IMRT era: current state-of-the-art.* Frontiers in oncology, 2020. **10**: p. 1708.
141. McDonald, D.G., et al., *Validation of a modern second-check dosimetry system using a novel verification phantom.* Journal of applied clinical medical physics, 2017. **18**(1): p. 170-177.
142. Shen, C., et al., *Clinical experience on patient-specific quality assurance for CBCT-based online adaptive treatment plan.* Journal of applied clinical medical physics, 2023. **24**(4): p. e13918.
143. Zhao, X., et al., *Do we need patient-specific QA for adaptively generated plans? Retrospective evaluation of delivered online adaptive treatment plans on Varian Ethos.* Journal of applied clinical medical physics, 2023. **24**(2): p. e13876.
144. Zwart, L.G.M., et al., *Cone-beam computed tomography-guided online adaptive radiotherapy is feasible for prostate cancer patients.* Physics and Imaging in Radiation Oncology, 2022. **22**: p. 98-103.
145. Astrom, L.M., et al., *Online adaptive radiotherapy of urinary bladder cancer with full re-optimization to the anatomy of the day: Initial experience and dosimetric benefits.* Radiother Oncol, 2022. **171**: p. 37-42.

146. Yoon, S.W., et al., *Initial Evaluation of a Novel Cone-Beam CT-Based Semi-Automated Online Adaptive Radiotherapy System for Head and Neck Cancer Treatment - A Timing and Automation Quality Study.* Cureus, 2020. **12**(8): p. e9660.
147. Byrne, M., et al., *Varian ethos online adaptive radiotherapy for prostate cancer: Early results of contouring accuracy, treatment plan quality, and treatment time.* J Appl Clin Med Phys, 2022. **23**(1): p. e13479.
148. Bak, M.E., et al., *Clinical experiences with online adaptive radiotherapy of vulvar carcinoma.* Acta Oncologica, 2023. **62**(10): p. 1230-1238.
149. Nasser, N., et al., *A head and neck treatment planning strategy for a CBCT-guided ring-gantry online adaptive radiotherapy system.* Journal of Applied Clinical Medical Physics, 2023. **24**(12): p. e14134.
150. Mao, W., et al., *Evaluation of Auto-Contouring and Dose Distributions for Online Adaptive Radiation Therapy of Patients With Locally Advanced Lung Cancers.* Pract Radiat Oncol, 2022.
151. Moazzezi, M., et al., *Prospects for daily online adaptive radiotherapy via ethos for prostate cancer patients without nodal involvement using unedited CBCT auto-segmentation.* J Appl Clin Med Phys, 2021. **22**(10): p. 82-93.
152. Pierrard, J., et al., *Cone-beam computed tomography-guided online-adaptive radiotherapy for inoperable right colon cancer: First in human.* Technical innovations & patient support in radiation oncology, 2023. **28**: p. 100220.
153. Robar, J.L., et al., *Novel technology allowing cone beam computed tomography in 6 seconds: a patient study of comparative image quality.* Practical Radiation Oncology, 2024. **14**(3): p. 277-286.
154. Haertter, A., et al., *ACR benchmark testing of a novel high-speed ring-gantry linac kV-CBCT system.* Journal of Applied Clinical Medical Physics, 2024: p. e14299.
155. Lustermans, D., et al., *Image quality evaluation of a new high-performance ring-gantry cone-beam computed tomography imager.* Physics in Medicine & Biology, 2024. **69**(10): p. 105018.
156. Kunnen, B., et al., *The added value of a new high-performance ring-gantry CBCT imaging system for prostate cancer patients.* Radiotherapy and Oncology, 2024. **200**: p. 110458.
157. Sijtsema, N., et al., *Dose Calculation Accuracy of Hypersight Cone-Beam Ct for Prostate and Lung Cancer Patients.* Available at SSRN 4826897.
158. Duan, J., et al., *Enhancing Precision in Radiation Therapy for Locally Advanced Lung Cancer: A Case Study of Cone-Beam Computed Tomography (CBCT)-Based Online Adaptive Techniques and the Promise of HyperSight™ Iterative CBCT.* Cureus, 2024. **16**(8).

159. Sluijter, J.H., et al., *Does Hypersight Cbct Imaging Allow for Improved Artificial Intelligence-Based Autosegmentation in the Pelvic Region?* Available at SSRN 4786562.
160. MacDonald, R.L., et al., *The feasibility of CT simulation-free adaptive radiation therapy.* Journal of Applied Clinical Medical Physics, 2024: p. e14438.
161. Buchner, J.A., et al., *Development and external validation of an MRI-based neural network for brain metastasis segmentation in the AURORA multicenter study.* Radiotherapy and Oncology, 2023. **178**: p. 109425.
162. Nachbar, M., et al., *Automatic AI-based contouring of prostate MRI for online adaptive radiotherapy.* Zeitschrift für Medizinische Physik, 2024. **34**(2): p. 197-207.
163. Lorenzen, E.L., et al., *An open-source nnU-net algorithm for automatic segmentation of MRI scans in the male pelvis for adaptive radiotherapy.* Frontiers in Oncology, 2023. **13**: p. 1285725.
164. Naccarato, S., et al., *Automated planning for prostate stereotactic body radiation therapy on the 1.5 T MR-Linac.* Advances in radiation oncology, 2022. **7**(3): p. 100865.
165. Bijman, R., et al., *First system for fully-automated multi-criterial treatment planning for a high-magnetic field MR-Linac applied to rectal cancer.* Acta Oncologica, 2020. **59**(8): p. 926-932.
166. Chen, G.P., E. Ahunbay, and X.A. Li, *Development and performance of a software tool for quality assurance of online replanning with a conventional Linac or MR-Linac.* Medical Physics, 2016. **43**(4): p. 1713-1719.
167. Paulson, E.S., et al., *4D-MRI driven MR-guided online adaptive radiotherapy for abdominal stereotactic body radiation therapy on a high field MR-Linac: Implementation and initial clinical experience.* Clinical and translational radiation oncology, 2020. **23**: p. 72-79.
168. Chen, X., et al., *A daily end-to-end quality assurance workflow for MR-guided online adaptive radiation therapy on MR-Linac.* Journal of applied clinical medical physics, 2020. **21**(1): p. 205-212.
169. Olaciregui-Ruiz, I., et al., *Automatic dosimetric verification of online adapted plans on the Unity MR-Linac using 3D EPID dosimetry.* Radiotherapy and Oncology, 2021. **157**: p. 241-246.
170. Wegener, D., et al., *First experience and prospective evaluation on feasibility and acute toxicity of online adaptive radiotherapy of the prostate bed as salvage treatment in patients with biochemically recurrent prostate cancer on a 1.5 T MR-Linac.* Journal of Clinical Medicine, 2022. **11**(16): p. 4651.

171. de Leon, J., et al., *Early experience with MR-guided adaptive radiotherapy using a 1.5 T MR-Linac: first 6 months of operation using adapt to shape workflow.* Journal of Medical Imaging and Radiation Oncology, 2022. **66**(1): p. 138-145.
172. Uno, T., et al., *A new workflow of the on-line 1.5-T MR-guided adaptive radiation therapy.* Japanese Journal of Radiology, 2023. **41**(11): p. 1316-1322.
173. Gupta, A., et al., *Online adaptive radiotherapy for head and neck cancers on the MR linear Accelerator: Introducing a novel modified Adapt-to-Shape approach.* Clinical and translational radiation oncology, 2022. **32**: p. 48-51.
174. Ding, S., et al., *Dosimetric Advantages of Online Adaptative Radiotherapy for Cervical Cancer on 1.5 T MR-Linac.* International Journal of Radiation Oncology, Biology, Physics, 2023. **117**(2): p. e509.
175. Wang, Y., et al., *A GPU-accelerated Monte Carlo dose calculation platform and its application toward validating an MRI-guided radiation therapy beam model.* Medical physics, 2016. **43**(7): p. 4040-4052.
176. Gungor, G., et al., *Time Analysis of Online Adaptive Magnetic Resonance-Guided Radiation Therapy Workflow According to Anatomical Sites.* Pract Radiat Oncol, 2021. **11**(1): p. e11-e21.
177. Bernchou, U., et al., *Treatment planning and delivery workflow steps in MR-guided adaptive RT*, in *Advances in Magnetic Resonance Technology and Applications*. 2023, Elsevier. p. 153-167.
178. Weykamp, F., et al., *Dosimetric Benefit of Adaptive magnetic resonance-guided stereotactic body Radiotherapy of Liver Metastases.* Cancers, 2022. **14**(24): p. 6041.
179. Fink, C.A., et al., *Dosimetric benefit of online treatment plan adaptation in stereotactic ultrahypofractionated MR-guided radiotherapy for localized prostate cancer.* Frontiers in oncology, 2024. **14**: p. 1308406.
180. El-Bared, N., et al., *Dosimetric benefits and practical pitfalls of daily online adaptive MRI-guided stereotactic radiation therapy for pancreatic cancer.* Practical radiation oncology, 2019. **9**(1): p. e46-e54.
181. Cusumano, D., et al., *Artificial Intelligence in magnetic Resonance guided Radiotherapy: Medical and physical considerations on state of art and future perspectives.* Physica medica, 2021. **85**: p. 175-191.
182. Henke, L.E., et al., *Stereotactic MR-guided online adaptive radiation therapy (SMART) for ultracentral thorax malignancies: results of a phase 1 trial.* Advances in radiation oncology, 2019. **4**(1): p. 201-209.
183. Finazzi, T., et al., *Stereotactic MR-guided adaptive radiation therapy for peripheral lung tumors.* Radiotherapy and Oncology, 2020. **144**: p. 46-52.

184. Fallone, B.G., et al., *The Alberta Rotating Biplanar Linac-MR, aka, Aurora-RT™*, in *A Practical Guide to MR-Linac: Technical Innovation and Clinical Implication*. 2024, Springer. p. 193-215.
185. Chen, W., et al., *Including robustness in multi-criteria optimization for intensity-modulated proton therapy.* Physics in Medicine & Biology, 2012. **57**(3): p. 591.
186. Albertini, F., et al., *Online daily adaptive proton therapy.* The British journal of radiology, 2020. **93**(1107): p. 20190594.
187. Landry, G., et al., *Investigating CT to CBCT image registration for head and neck proton therapy as a tool for daily dose recalculation.* Medical physics, 2015. **42**(3): p. 1354-1366.
188. Matter, M., et al., *Intensity modulated proton therapy plan generation in under ten seconds.* Acta Oncologica, 2019. **58**(10): p. 1435-1439.
189. Nenoff, L., et al., *Daily adaptive proton therapy: is it appropriate to use analytical dose calculations for plan adaption?* International Journal of Radiation Oncology* Biology* Physics, 2020. **107**(4): p. 747-755.
190. Bernatowicz, K., et al., *Feasibility of online IMPT adaptation using fast, automatic and robust dose restoration.* Physics in Medicine & Biology, 2018. **63**(8): p. 085018.
191. Borderias-Villarroel, E., et al., *Dose mimicking based strategies for online adaptive proton therapy of head and neck cancer.* Physics in Medicine & Biology, 2023. **68**(10): p. 105002.
192. Borderias-Villarroel, E., et al., *Machine learning-based automatic proton therapy planning: Impact of post-processing and dose-mimicking in plan robustness.* Medical Physics, 2023. **50**(7): p. 4480-4490.
193. Borderías-Villarroel, E., et al., *Evaluation of the clinical value of automatic online dose restoration for adaptive proton therapy of head and neck cancer.* Radiotherapy and Oncology, 2022. **170**: p. 190-197.
194. Draguet, C., et al., *A comparative study on automatic treatment planning for online adaptive proton therapy of esophageal cancer: which combination of deformable registration and deep learning planning tools performs the best?* Physics in Medicine & Biology, 2024. **69**(20): p. 205013.
195. Botas, P., et al., *Online adaption approaches for intensity modulated proton therapy for head and neck patients based on cone beam CTs and Monte Carlo simulations.* Physics in Medicine & Biology, 2018. **64**(1): p. 015004.
196. Bobić, M., et al., *Comparison of weekly and daily online adaptation for head and neck intensity-modulated proton therapy.* Physics in Medicine & Biology, 2021. **66**(5): p. 055023.

197. de Jong, R., et al., *Online adaptive radiotherapy compared to plan selection for rectal cancer: quantifying the benefit.* Radiat Oncol, 2020. **15**(1): p. 162.
198. Ahunbay, E.E., et al., *Online adaptive replanning method for prostate radiotherapy.* Int J Radiat Oncol Biol Phys, 2010. **77**(5): p. 1561-72.
199. Vestergaard, A., et al., *Adaptive plan selection vs. re-optimisation in radiotherapy for bladder cancer: a dose accumulation comparison.* Radiother Oncol, 2013. **109**(3): p. 457-62.
200. Pokharel, S., A. Pacheco, and S. Tanner, *Assessment of efficacy in automated plan generation for Varian Ethos intelligent optimization engine.* J Appl Clin Med Phys, 2022. **23**(4): p. e13539.
201. Calmels, L., et al., *Evaluation of an automated template-based treatment planning system for radiotherapy of anal, rectal and prostate cancer.* Tech Innov Patient Support Radiat Oncol, 2022. **22**: p. 30-36.
202. Ge, Y. and Q.J. Wu, *Knowledge-based planning for intensity-modulated radiation therapy: a review of data-driven approaches.* Medical Physics, 2019. **46**(6): p. 2760-2775.
203. Wang, C., et al., *Artificial intelligence in radiotherapy treatment planning: present and future.* Technology in cancer research & treatment, 2019. **18**: p. 1533033819873922.
204. Babier, A., et al., *OpenKBP-Opt: An international and reproducible evaluation of 76 knowledge-based planning pipelines.* arXiv preprint arXiv:2202.08303, 2022.
205. Fan, J., et al., *Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique.* Medical physics, 2019. **46**(1): p. 370-381.
206. Petersson, K., et al., *Evaluation of dual-arc VMAT radiotherapy treatment plans automatically generated via dose mimicking.* Acta Oncol, 2016. **55**(4): p. 523-5.
207. Huet-Dastarac, M., et al., *Patient selection for proton therapy using Normal Tissue Complication Probability with deep learning dose prediction for oropharyngeal cancer.* Medical Physics, 2023. **50**(10): p. 6201-6214.
208. Nguyen, D., et al., *A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning.* Scientific reports, 2019. **9**(1): p. 1-10.
209. Kandalan, R.N., et al., *Dose prediction with deep learning for prostate cancer radiation therapy: model adaptation to different treatment planning practices.* Radiotherapy and Oncology, 2020. **153**: p. 228-235.
210. Gronberg, M.P., et al., *Dose prediction for head and neck radiotherapy using a three-dimensional dense dilated U-net architecture.* Medical physics, 2021. **48**(9): p. 5567-5573.

211. Zimmermann, L., et al., *Technical Note: Dose prediction for radiation therapy using feature-based losses and One Cycle Learning.* Med Phys, 2021. **48**(9): p. 5562-5566.
212. Draguet, C., et al., *Automated clinical decision support system with deep learning dose prediction and NTCP models to evaluate treatment complications in patients with esophageal cancer.* Radiotherapy and Oncology, 2022. **176**: p. 101-107.
213. Nguyen, D., et al., *A comparison of Monte Carlo dropout and bootstrap aggregation on the performance and uncertainty estimation in radiation therapy dose prediction with deep learning neural networks.* Physics in Medicine & Biology, 2021. **66**(5): p. 054002.
214. Villarroel, E.B., X. Geets, and E. Sterpin, *Online adaptive dose restoration in intensity modulated proton therapy of lung cancer to account for inter-fractional density changes.* Physics and imaging in radiation oncology, 2020. **15**: p. 30-37.
215. McNiven, A.L., M.B. Sharpe, and T.G. Purdie, *A new metric for assessing IMRT modulation complexity and plan deliverability.* Medical physics, 2010. **37**(2): p. 505-515.
216. Varian, *Ethos Algorithms Reference Guide*. 2019.
217. McIntosh, C., et al., *Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer.* Nat Med, 2021. **27**(6): p. 999-1005.
218. Cha, E., et al., *Clinical implementation of deep learning contour autosegmentation for prostate radiotherapy.* Radiotherapy and Oncology, 2021. **159**: p. 1-7.
219. Fu, Y., et al., *A review of deep learning based methods for medical image multi-organ segmentation.* Physica Medica, 2021. **85**: p. 107-122.
220. Chapman, J.W., et al., *Robustness and reproducibility of an artificial intelligence-assisted online segmentation and adaptive planning process for online adaptive radiation therapy.* Journal of applied clinical medical physics, 2022. **23**(8): p. e13702.
221. Rigaud, B., et al., *Automatic segmentation using deep learning to enable online dose optimization during adaptive radiation therapy of cervical cancer.* International Journal of Radiation Oncology* Biology* Physics, 2021. **109**(4): p. 1096-1110.
222. Tetar, S.U., et al., *Clinical implementation of magnetic resonance imaging guided adaptive radiotherapy for localized prostate cancer.* Physics and imaging in radiation oncology, 2019. **9**: p. 69-76.
223. van Herk, M., et al., *Magnetic resonance imaging–guided radiation therapy: a short strengths, weaknesses, opportunities, and threats*

analysis. International journal of radiation oncology, biology, physics, 2018. **101**(5): p. 1057-1060.

224. Van der Veen, J., et al., *Deep learning for elective neck delineation: More consistent and time efficient.* Radiotherapy and Oncology, 2020. **153**: p. 180-188.
225. Chen, X., et al., *CNN-based quality assurance for automatic segmentation of breast cancer in radiotherapy.* Frontiers in Oncology, 2020. **10**: p. 524.
226. Luan, S., et al., *Machine Learning-Based Quality Assurance for Automatic Segmentation of Head-and-Neck Organs-at-Risk in Radiotherapy.* Technology in Cancer Research & Treatment, 2023. **22**: p. 15330338231157936.
227. Isaksson, L.J., et al., *Quality assurance for automatically generated contours with additional deep learning.* Insights into Imaging, 2022. **13**(1): p. 1-10.
228. Claessens, M., et al., *Machine learning-based detection of aberrant deep learning segmentations of target and organs at risk for prostate radiotherapy using a secondary segmentation algorithm.* Physics in Medicine & Biology, 2022. **67**(11): p. 115014.
229. Hann, E., et al., *Deep neural network ensemble for on-the-fly quality control-driven segmentation of cardiac MRI T1 mapping.* Medical image analysis, 2021. **71**: p. 102029.
230. Balagopal, A., et al., *A deep learning-based framework for segmenting invisible clinical target volumes with estimated uncertainties for post-operative prostate cancer radiotherapy.* Medical image analysis, 2021. **72**: p. 102101.
231. Claessens, M., et al. *Quality assurance for AI-based applications in radiation therapy*. in *Seminars in radiation oncology*. 2022. Elsevier.
232. Kaderka, R., et al., *Geometric and dosimetric evaluation of atlas based auto-segmentation of cardiac structures in breast cancer patients.* Radiotherapy and Oncology, 2019. **131**: p. 215-220.
233. Fung, N.T.C., et al., *Automatic segmentation for adaptive planning in nasopharyngeal carcinoma IMRT: time, geometrical, and dosimetric analysis.* Medical Dosimetry, 2020. **45**(1): p. 60-65.
234. Tezzele, M., et al., *PyGeM: Python geometrical morphing.* Software Impacts, 2021. **7**: p. 100047.
235. Roberfroid, B., et al., *Comparison of Ethos template-based planning and AI-based dose prediction: General performance, patient optimality, and limitations.* Physica Medica, 2023. **116**: p. 103178.
236. Challen, R., et al., *Artificial intelligence, bias and clinical safety.* BMJ Quality & Safety, 2019. **28**(3): p. 231-237.

237. Bond, R.R., et al., *Automation bias in medicine: the influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms.* Journal of electrocardiology, 2018. **51**(6): p. S6-S11.
238. Philpotts, L.E., *Can computer-aided detection be detrimental to mammographic interpretation?* Radiology, 2009. **253**(1): p. 17-22.
239. de Jong, B.A., et al., *Proton arc therapy increases the benefit of proton therapy for oropharyngeal cancer patients in the model based clinic.* Radiotherapy and Oncology, 2023. **184**: p. 109670.
240. Ding, X., et al., *Improving dosimetric outcome for hippocampus and cochlea sparing whole brain radiotherapy using spot-scanning proton arc therapy.* Acta Oncologica, 2019. **58**(4): p. 483-490.
241. Chang, S., et al., *Feasibility study: spot-scanning proton arc therapy (SPArc) for left-sided whole breast radiotherapy.* Radiation Oncology, 2020. **15**: p. 1-11.
242. Liu, G., et al., *Improve the dosimetric outcome in bilateral head and neck cancer (HNC) treatment using spot-scanning proton arc (SPArc) therapy: a feasibility study.* Radiation Oncology, 2020. **15**: p. 1-11.
243. De Jong, B.A., et al., *Spot scanning proton arc therapy reduces toxicity in oropharyngeal cancer patients.* Medical Physics, 2023. **50**(3): p. 1305-1317.
244. Engwall, E., et al., *Partitioning of discrete proton arcs into interlaced subplans can bring proton arc advances to existing proton facilities.* Medical Physics, 2023. **50**(9): p. 5723-5733.
245. Unkelbach, J., T.C. Chan, and T. Bortfeld, *Accounting for range uncertainties in the optimization of intensity modulated proton therapy.* Physics in Medicine & Biology, 2007. **52**(10): p. 2755.
246. Van de Water, S., et al., *Anatomical robust optimization to account for nasal cavity filling variation during intensity-modulated proton therapy: a comparison with conventional and adaptive planning strategies.* Physics in Medicine & Biology, 2018. **63**(2): p. 025020.
247. Lalonde, A., et al., *Anatomic changes in head and neck intensity-modulated proton therapy: Comparison between robust optimization and online adaptation.* Radiotherapy and Oncology, 2021. **159**: p. 39-47.
248. Wuyckens, S., et al., *Bi-criteria Pareto optimization to balance irradiation time and dosimetric objectives in proton arc therapy.* Physics in Medicine & Biology, 2022. **67**(24): p. 245017.
249. Ding, X., et al., *Spot-scanning proton arc (SPArc) therapy: the first robust and delivery-efficient spot-scanning proton arc therapy.* International Journal of Radiation Oncology* Biology* Physics, 2016. **96**(5): p. 1107-1116.

250. Engwall, E., et al., *Fast robust optimization of proton PBS arc therapy plans using early energy layer selection and spot assignment.* Physics in Medicine & Biology, 2022. **67**(6): p. 065010.
251. Zhao, L., et al., *An evolutionary optimization algorithm for proton arc therapy.* Physics in Medicine & Biology, 2022. **67**(16): p. 16NT01.
252. Gu, W., et al., *A novel energy layer optimization framework for spot-scanning proton arc therapy.* Medical Physics, 2020. **47**(5): p. 2072-2084.
253. Zhang, G., et al., *A treatment plan optimization method with direct minimization of number of energy jumps for proton arc therapy.* Physics in Medicine & Biology, 2023. **68**(8): p. 085001.
254. Borderias Villarroel, E., X. Geets, and E. Sterpin, *Online adaptive dose restoration in intensity modulated proton therapy of lung cancer to account for inter-fractional density changes.* Physics and Imaging in Radiation Oncology, 2020. **15**: p. 30-37.
255. Jagt, T., et al., *Near real-time automated dose restoration in IMPT to compensate for daily tissue density variations in prostate cancer.* Physics in Medicine & Biology, 2017. **62**(11): p. 4254.
256. Wuyckens, S., et al., *OpenTPS--Open-source treatment planning system for research in proton therapy.* arXiv preprint arXiv:2303.00365, 2023.
257. Souris, K., J.A. Lee, and E. Sterpin, *Fast multipurpose Monte Carlo simulation for proton therapy using multi-and many-core CPU architectures.* Medical Physics, 2016. **43**(4): p. 1700-1712.
258. Friedl, A.A., et al., *Radiobiology of the FLASH effect.* Medical Physics, 2022. **49**(3): p. 1993-2013.
259. Montay-Gruel, P., et al., *Irradiation in a flash: Unique sparing of memory in mice after whole brain irradiation with dose rates above 100 Gy/s.* Radiotherapy and Oncology, 2017. **124**(3): p. 365-369.
260. Montay-Gruel, P., et al., *Hypofractionated FLASH-RT as an effective treatment against glioblastoma that reduces neurocognitive side effects in mice.* Clinical Cancer Research, 2021. **27**(3): p. 775-784.
261. Vozenin, M.-C., et al., *The advantage of FLASH radiotherapy confirmed in mini-pig and cat-cancer patients.* Clinical Cancer Research, 2019. **25**(1): p. 35-42.
262. Montay-Gruel, P., et al., *Long-term neurocognitive benefits of FLASH radiotherapy driven by reduced reactive oxygen species.* Proceedings of the National Academy of Sciences, 2019. **116**(22): p. 10943-10951.
263. Bogaerts, E., et al., *Potential molecular mechanisms behind the ultra-high dose rate "FLASH" effect.* International Journal of Molecular Sciences, 2022. **23**(20): p. 12109.

264. Krieger, M., et al., *A quantitative FLASH effectiveness model to reveal potentials and pitfalls of high dose rate proton therapy.* Medical physics, 2022. **49**(3): p. 2026-2038.
265. Vozenin, M.-C., J. Bourhis, and M. Durante, *Towards clinical translation of FLASH radiotherapy.* Nature Reviews Clinical Oncology, 2022. **19**(12): p. 791-803.
266. Habraken, S., et al., *Trade-off in healthy tissue sparing of FLASH and fractionation in stereotactic proton therapy of lung lesions with transmission beams.* Radiotherapy and Oncology, 2022. **175**: p. 231-237.
267. Breneman, J., *Feasibility study of FLASH radiotherapy for the treatment of symptomatic bone metastases (FAST-01).* Oncology R, ed.: NIH, 2020.
268. Van De Water, S., et al., *Towards FLASH proton therapy: the impact of treatment planning and machine characteristics on achievable dose rates.* Acta oncologica, 2019. **58**(10): p. 1463-1469.
269. Rothwell, B., et al., *Treatment planning considerations for the development of FLASH proton therapy.* Radiotherapy and Oncology, 2022. **175**: p. 222-230.
270. Patriarca, A., et al., *Experimental set-up for FLASH proton irradiation of small animals using a clinical system.* International Journal of Radiation Oncology* Biology* Physics, 2018. **102**(3): p. 619-626.
271. Wei, S., et al., *Dose rate and dose robustness for proton transmission FLASH-RT treatment in lung cancer.* Frontiers in Oncology, 2022. **12**: p. 970602.
272. Deffet, S., K. Souris, and E. Sterpin, *Optimization of patient-specific range modulators for conformal FLASH proton therapy.* Medical Physics (submitted).
273. Zhang, G., W. Gao, and H. Peng, *Design of static and dynamic ridge filters for FLASH–IMPT: A simulation study.* Medical physics, 2022. **49**(8): p. 5387-5399.
274. Bélanger-Champagne, C., et al., *Delivery of proton FLASH at the TRIUMF Proton Therapy Research Centre.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2023. **1052**: p. 168243.
275. Roddy, D., et al., *Design, optimization, and testing of ridge filters for proton FLASH radiotherapy at TRIUMF: The HEDGEHOG.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2024. **1063**: p. 169284.
276. Barragan-Montero, A., et al., *Artificial intelligence and machine learning for medical imaging: A technology review.* Phys Med, 2021. **83**: p. 242-256.

277. Harrison, K., et al., *Machine learning for auto-segmentation in radiotherapy planning.* Clinical Oncology, 2022. **34**(2): p. 74-88.
278. Roberfroid, B., et al., *DIVE-ART: A tool to guide clinicians towards dosimetrically informed volume editions of automatically segmented volumes in adaptive radiation therapy.* Radiotherapy and Oncology, 2024. **192**: p. 110108.
279. Diez, P., et al., *UK 2022 consensus on normal tissue dose-volume constraints for oligometastatic, primary lung and hepatocellular carcinoma stereotactic ablative radiotherapy.* Clinical Oncology, 2022. **34**(5): p. 288-300.
280. Deffet, S., V. Hamaide, and E. Sterpin, *Definition of dose rate for FLASH pencil-beam scanning proton therapy: A comparative study.* Medical Physics, 2023. **50**(9): p. 5784-5792.
281. Hotoiu, L., et al., *Open source TPS for Bragg peak proton FLASH treatment plans.* Medical Physics, 2021. **48**(6).
282. Pin, A., et al., *Pencil beam proton flash therapy, field size limit with ConformalFLASH.* Physica Medica: European Journal of Medical Physics, 2022. **94**: p. S66-S67.
283. Folkerts, M.M., et al., *A framework for defining FLASH dose rate for pencil beam scanning.* Medical physics, 2020. **47**(12): p. 6396-6404.
284. Vozenin, M.-C., J.H. Hendry, and C. Limoli, *Biological benefits of ultra-high dose rate FLASH radiotherapy: sleeping beauty awoken.* Clinical oncology, 2019. **31**(7): p. 407-415.
285. Böhlen, T.T., et al., *Normal tissue sparing by FLASH as a function of single-fraction dose: A quantitative analysis.* International Journal of Radiation Oncology* Biology* Physics, 2022. **114**(5): p. 1032-1044.
286. Gao, H., et al., *Simultaneous dose and dose rate optimization (SDDRO) for FLASH proton therapy.* Medical Physics, 2020. **47**(12): p. 6388-6395.
287. Ma, J., et al., *Simultaneous dose and dose rate optimization via dose modifying factor modeling for FLASH effective dose.* Medical Physics, 2024.
288. Pross, D., et al., *Beamlet-free optimization for Monte-Carlo-based treatment planning in proton therapy.* Medical Physics, 2024. **51**(1): p. 485-493.
289. Bowman, E., et al., *Dynamic three-dimensional beam modification for radiation therapy*. 2021, Google Patents.
290. Putz, F., et al., *The Segment Anything foundation model achieves favorable brain tumor auto-segmentation accuracy in MRI to support radiotherapy treatment planning.* Strahlentherapie und Onkologie, 2024: p. 1-11.

291. Elguindi, S., et al., *AI Generated Clinical Target Volumes for Oropharyngeal Cancers Using a General Foundation Model.* International Journal of Radiation Oncology, Biology, Physics, 2024. **120**(2): p. e620-e621.
292. Pai, S., et al., *Foundation model for cancer imaging biomarkers.* Nature machine intelligence, 2024. **6**(3): p. 354-367.
293. Barragán-Montero, A., et al., *Towards a safe and efficient clinical implementation of machine learning in radiation oncology by exploring model interpretability, explainability and data-model dependency.* Physics in Medicine & Biology, 2022. **67**(11): p. 11TR01.
294. Luo, Y., et al., *Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling.* BJR open, 2019. **1**(1).
295. Reyes, M., et al., *On the interpretability of artificial intelligence in radiology: challenges and opportunities.* Radiology: artificial intelligence, 2020. **2**(3): p. e190043.
296. Brouwer, C.L., et al., *Machine learning applications in radiation oncology: Current use and needs to support clinical implementation.* Physics and imaging in radiation oncology, 2020. **16**: p. 144-148.
297. Vandewinckele, L., et al., *Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance.* Radiotherapy and Oncology, 2020. **153**: p. 55-66.
298. Babier, A., et al., *OpenKBP: the open-access knowledge-based planning grand challenge and dataset.* Medical Physics, 2021. **48**(9): p. 5549-5561.
299. Clark, K., et al., *The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository.* Journal of digital imaging, 2013. **26**: p. 1045-1057.
300. Hito, M., et al., *Assessing the robustness of artificial intelligence powered planning tools in radiotherapy clinical settings—a phantom simulation approach.* Quantitative Imaging in Medicine and Surgery, 2021. **11**(12): p. 4835.
301. Huet-Dastarac, M., et al., *Can input reconstruction be used to directly estimate uncertainty of a dose prediction U-Net model?* Medical physics, 2024.
302. Lakshminarayanan, B., A. Pritzel, and C. Blundell, *Simple and scalable predictive uncertainty estimation using deep ensembles.* Advances in neural information processing systems, 2017. **30**.
303. Gal, Y. and Z. Ghahramani. *Dropout as a bayesian approximation: Representing model uncertainty in deep learning*. in *international conference on machine learning*. 2016. PMLR.

304. Kushibar, K., et al. *Layer ensembles: A single-pass uncertainty estimation in deep learning for segmentation*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2022. Springer.
305. Hysing, L.B., et al., *Planning organ at risk volume margins for organ motion of the intestine.* Radiotherapy and oncology, 2006. **80**(3): p. 349-354.
306. Muren, L.P., et al., *Testing the new ICRU 62 'Planning Organ at Risk Volume'concept for the rectum.* Radiotherapy and oncology, 2005. **75**(3): p. 293-302.
307. Buti, G., et al., *Introducing a probabilistic definition of the target in a robust treatment planning framework.* Physics in Medicine & Biology, 2021. **66**(15): p. 155008.
308. Buti, G., et al., *Exploring trade-offs in treatment planning for brain tumor cases with a probabilistic definition of the clinical target volume.* Medical Physics, 2023. **50**(1): p. 410-423.
309. Nenoff, L., et al., *Integrating structure propagation uncertainties in the optimization of online adaptive proton therapy plans.* Cancers, 2022. **14**(16): p. 3926.
310. Smolders, A., et al., *Robust optimization strategies for contour uncertainties in online adaptive radiation therapy.* Physics in Medicine & Biology, 2024. **69**(16): p. 165001.
311. Fankhauser, L., et al., *PP14. 10 EFFICIENT AUTOMATIC CONTOUR QUALITY ASSURANCE (CAT-QA) FOR SAFE ONLINE ADAPTIVE RADIOTHERAPY.* Physica Medica, 2024. **125**: p. 103918.
312. Hui, C.B., et al., *Quality assurance tool for organ at risk delineation in radiation therapy using a parametric statistical approach.* Medical physics, 2018. **45**(5): p. 2089-2096.
313. Miura, H., et al., *Evaluation of the accuracy of automated segmentation based on deep learning for prostate cancer patients.* Medical Dosimetry, 2024.
314. Thibeau-Sutre, E., et al. *Uncertainty-based quality assurance of carotid artery wall segmentation in black-blood MRI*. in *International Workshop on Uncertainty for Safe Utilization of Machine Learning in Medical Imaging*. 2023. Springer.
315. Ren, J., et al., *Enhancing the reliability of deep learning-based head and neck tumour segmentation using uncertainty estimation with multi-modal images.* Physics in Medicine & Biology, 2024. **69**(16): p. 165018.
316. Huet-Dastarac, M., et al., *Quantifying and visualising uncertainty in deep learning-based segmentation for radiation therapy treatment planning: What do radiation oncologists and therapists want?* Radiotherapy and Oncology, 2024. **201**: p. 110545.

317. Pierrard, J., et al., *On the trail of CBCT-guided adaptive rectal boost radiotherapy, does daily delineation require a radiation oncologist?* Technical Innovations & Patient Support in Radiation Oncology, 2024: p. 100284.
318. Wei, Z., et al., *Towards interactive deep-learning for tumour segmentation in head and neck cancer radiotherapy.* Physics and Imaging in Radiation Oncology, 2023. **25**: p. 100408.
319. Bai, T., et al., *A proof-of-concept study of artificial intelligence–assisted contour editing.* Radiology: Artificial Intelligence, 2022. **4**(5): p. e210214.
320. Castelluccia, A., et al., *Economic evaluations of magnetic resonance image-guided radiotherapy (MRIgRT): A systematic review.* International Journal of Environmental Research and Public Health, 2022. **19**(17): p. 10800.
321. Parikh, N.R., et al., *Time-driven activity-based costing of CT-guided vs MR-guided prostate SBRT.* Applied radiation oncology, 2021. **10**(3): p. 33.
322. Thorwarth, D. and D.A. Low, *Technical challenges of real-time adaptive MR-guided radiotherapy.* Frontiers in oncology, 2021. **11**: p. 634507.
323. Muurholm, C.G., et al., *Real-time dose-guidance in radiotherapy: Proof of principle.* Radiotherapy and Oncology, 2021. **164**: p. 175-182.
324. van der Bijl, E., et al., *Adaptive margins for online adaptive radiotherapy.* Physics in Medicine & Biology, 2022. **67**(19): p. 195016.
325. Dohopolski, M., et al., *Dosimetric impact of simulated daily adaptive radiotherapy with significantly reduced setup margins in the definitive treatment of head and neck cancer.* International Journal of Radiation Oncology, Biology, Physics, 2022. **114**(3): p. e590.
326. Lalonde, A., et al., *Evaluating the effect of setup uncertainty reduction and adaptation to geometric changes on normal tissue complication probability using online adaptive head and neck intensity modulated proton therapy.* Physics in Medicine & Biology, 2023. **68**(11): p. 115018.