UNIVERSIDADE D
COIMBRA

Natália Bortolami Colaço Alves

ASSESSING AND PREDICTING THE NEED FOR ADAPTIVE RADIOTHERAPY IN HEAD & NECK CANCER PATIENTS

Dissertação no âmbito do Mestrado Integrado em Engenharia Biomédica no ramo de Imagem Médica e Radiação, orientada pela Professora Doutora Maria do Carmo Lopes e pela Professora Doutora Joana Matos Dias, apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologias da Universidade de Coimbra

Julho de 2020

Natália Bortolami Colaço Alves

Assessing and predicting the need for Adaptive Radiotherapy in Head & Neck cancer patients

Dissertation presented to the University of Coimbra to obtain the Master's degree in Biomedical Engineering

Supervisors:

Joana Matos Dias (Faculty of Economics, University of Coimbra)
Maria do Carmo Lopes (Medical Physics Department, IPOCFG, E.P.E.)

Coimbra, 2020

This work was developed in collaboration with:

Instituto Português de Oncologia Francisco Gentil – E.P.E.

Instituto de Engenharia de Sistemas e Computadores de Coimbra

Agradecimentos

Os meus sinceros agradecimentos às minhas orientadoras Doutora Maria do Carmo Lopes e Doutora Joana Matos Dias que se dispuseram a idealizar e planear este projeto para mim. Foi uma grande honra poder trabalhar com duas mulheres tão extraordinárias e realizadas, que todos os dias me inspiraram a ser uma melhor engenheira e pessoa. Muito obrigada por toda a atenção, os conselhos e o carinho que tiveram para comigo, não poderia ter pedido melhores orientadoras para me acompanharem nesta fase final do meu percurso académico.

Um especial obrigada ao Engenheiro Tiago Ventura, pelas horas que despendeu do seu trabalho para me apoiar na concretização deste projeto, por todas as suas sugestões e acima de tudo pela paciência, disponibilidade e boa disposição.

Agradeço também à Doutora Leila Khouri pela disponibilização dos dados clínicos, à engenheira Josefina Mateus, pela sua disponibilidade para ajudar, preocupação e alegria e ao Engenheiro Miguel Capela por toda a ajuda e esclarecimentos prestados. Um obrigada muito especial à Engenheira Tânia Santos, por ter estado sempre lá para me ajudar e apoiar em tudo o que precisei, estou infinitamente grata por os nossos caminhos se terem cruzado. A estes aqui referidos e a todos os outros funcionários e funcionárias do IPO com os quais me cruzei ao longo destes meses, obrigada por terem proporcionado um ambiente descontraído e alegre, mas ao mesmo tempo extremamente profissional, que me proporcionou uma imensa aprendizagem.

À Laura e à Adriana, que ao longo destes cinco anos estiveram sempre lá para os bons e maus momentos, vocês são a família que Coimbra me deu e levo-vos comigo para a vida.

Ao Lar Teresiano e a todas as irmãs que se cruzaram no meu caminho, obrigada por nos acolherem e proporcionarem uma casa longe de casa.

Ao Vasco, pela compreensão, carinho, amor e companheirismo. Obrigada por acreditares sempre em mim e nunca me deixares esquecer das minhas capacidades.

Finalmente à minha família, aos meus pais e irmãos por terem feito com que tudo isto fosse possível, por todo amor incondicional e apoio que sempre me ofereceram e por serem o meu maior exemplo.

Financiamento:

Este trabalho foi parcialmente financiado pelo fundo POCI-01-0145-FEDER-028030.

Abstract

Currently the majority of patients with locally advanced head and neck (H&N) cancer require radiotherapy, with intensity modulated radiation therapy (IMRT) being the gold standard for treatment. The highly conformal dose distributions produced by IMRT lead to steep dose gradients in the borders of the target volumes, which are extremely sensitive to positional errors and anatomic changes, especially in H&N cases, since there are many critical structures close to the target volumes.

Several factors can lead to anatomic alterations of both the target volumes and organs at risk (OAR) during radiotherapy treatments which may originate discrepancies in dose delivery, namely underdosage of the target volume and/or to overdosage of the normal structures, potentially originating unexpected side effects.

A possible solution to this problem is Adaptive Radiotherapy (ART). Recently a great deal of scientific attention has been drawn to this topic due to its encouraging potential benefits, but questions remain about whether patients benefit from a dosimetric improvement by ART, and about the identification of the patients that are more likely to need it. The purpose of this dissertation is to address these two issues.

In the first part of this dissertation, the need for ART is assessed in a cohort of thirty patients from IPOCFG. This was done by using an automatic planning tool to generate plans simulating the scenarios with and without adaptation. Statistically significant dosimetric differences were reported for the target volumes, which presented lower near minimum dose in the non-adaptive scenario, and for the spinal cord, that showed a significant larger maximum dose in the non-adaptive scenario.

The second part of the dissertation focuses on developing machine learning models to predict which patients will require adaptive radiotherapy prior to the beginning of treatment. Three support vector machine models were developed: 1) considering only pre-treatment clinical data from the patient; 2) considering only radiomic features extracted from pre-treatment computed tomography images; 3) using a combination of features from 1 and 2. The best classification results were obtained considering 6 features (4 semantic and 2 radiomic) with median accuracy and area under the receiver operating characteristic curve of 0.821 and 0.843, respectively.

Keywords: Radiotherapy, Adaptive Radiotherapy, Head and Neck cancer, Radiomics, Machine Learning

Resumo

Hoje em dia a grande maioria de doentes com cancro da cabeça e pescoço localmente avançado requerem radioterapia, sendo a radioterapia de intensidade modulada (IMRT) o padrão para tratamento. As distribuições de dose altamente conformacionais obtidas com IMRT originam gradientes de dose acentuados nas margens dos volumes alvo, que por sua vez são extremamente sensíveis a erros posicionais e mudanças anatómicas, especialmente em casos de cabeça e pescoço uma vez que existem várias estruturas críticas muito próximas dos volumes alvo.

Vários fatores podem contribuir para alterações anatómicas tanto nos volumes alvo como nos órgãos de risco (OAR) durante tratamentos de radioterapia, que podem originar discrepâncias na administração de dose, nomeadamente subdosagem dos volumes alvos e/ou sobredosagem dos órgãos de risco, potencialmente originando efeitos secundários não esperados.

Uma possível solução para este problema é Radioterapia Adaptativa. Recentemente muita atenção tem sido direcionada a este tópico, devido aos seus encorajadores potenciais benefícios, no entanto, questões acerca do seu impacto dosimétrico e da identificação dos doentes com maior probabilidade de beneficiar de adaptação ainda permanecem por responder. O objetivo desta dissertação é abordar estes dois problemas.

Na primeira parte da dissertação, a necessidade de radioterapia adaptativa é avaliada numa amostra de trinta doentes de cabeça e pescoço do IPOCFG. Recorreu-se a uma ferramenta de planeamento automático para gerar planos que simulam os cenários com e sem adaptação. Diferenças dosimétricas estatisticamente significativas foram encontradas para os volumes alvo, que apresentaram doses quase mínimas inferiores no cenário não adaptativo, e para a espinal medula, que apresentou uma dose máxima significativamente maior no cenário não adaptativo.

A segunda parte da dissertação concentra-se no desenvolvimento de modelos de machine learning para prever, antes do início do tratamento, que doentes vão necessitar de radioterapia adaptativa. Foram desenvolvidos três modelos de *support vector machine*: 1) considerando apenas dados clínicos pré-tratamento; 2) considerando apenas atributos radiómicos extraídos de imagens de tomografia computorizada pré-tratamento. 3) considerando uma combinação de 1 e 2. Os melhores resultados foram obtidos considerando 6 atributos (4 semânticos e 2 radiómicos) com precisão e área abaixo da curva ROC de 0.821 e 0.843 respetivamente.

Palavras chave: Radioterapia, Radioterapia Adaptativa, Cabeça e pescoço, Radiomics, Machine Learning

Contents

List of abbreviations

2D 2-dimentional

3D 3-dimentional

3D-CRT 3D conformal radiotherapy

ART Adaptive radiotherapy

AUC Area under the curve

CBCT Cone-beam Computed Tomography

CERR Computational Environment for Radiological Research

CT Computed Tomography

CTV Clinical target volume

DNA Deoxyribonucleic acid

DVH Dose-volume histogram

FDG Fluorodeoxyglucose

FN False negative

FNR False negative rate

FP False positive

FPR False positive rate

GTV Gross target volume

H&N Head and neck

HNSCC Head and neck squamous cell carcinoma

IAEA International Atomic Energy Agency

IBSI Imaging Biomarker Standardization Initiative

IGRT Image guided radiotherapy

IMRT Intensity modulated radiation therapy

KV Kilovoltage

LASSO Least Absolute Shrinkage and Selection Operator

LINAC Linear accelerator

LOO Leave-one-out

MRI Magnetic resonance image/imaging

MV Megavoltage

NSCLC Non-small cell lung cancer

NTCP Normal tissue control probability

NPV Negative predictive value

MLC Multi-leaf collimator

OAR Organs at risk

PET Positron emission tomography

PPV Positive predictive value

PTV Planning target volume

RO Radiation oncologist

ROC Receiver operating characteristic

RT Radiotherapy

TCP Tumour control probability

TN True negative

TNR True negative rate

TP True positive

TPR True positive rate

TPS Treatment planning system

List of Figures

List of Tables

Blank Page

Chapter 1

Introduction

The present dissertation focusses in the field of Adaptive Radiotherapy (ART). Since its introduction by Yan et al. [1] in 1997 a great deal of attention and research was drawn to the topic, due to its encouraging potential benefits for both tumoral control and ability to attenuate radiation induced side effects.

In this work both the necessity and benefit of ART in Head and Neck cancer patients, as well as the ability to predict which patients would need this approach prior to the beginning of treatment will be addressed.

1.1. Contextualization

By 2025 it is estimated that there will be about 4 million new cancer patients in Europe, which represent an increase of 15.9% when compared to the 3.4 million diagnosed cases in 2012. The number of patients that will require radiotherapy (RT) at least once during their oncological treatment is also expected to increase from approximately 1,700,000 patients in 2012 to 2,000,000 in 2025 (16.1% increase) [2].

In Portugal, oncological diseases are the second leading cause of death, preceded only by cardiovascular diseases, but they are also the cause of death with the highest increase rate over the last years [3].

Currently, RT is one of the most used cancer treatment modalities, with a rapidly growing number of patients worldwide. It is estimated that more than 50% of all oncological patients in Europe require radiotherapy at some point of the treatment, and that by 2025 the number of patients recommended for radiotherapy treatments in Portugal will be of 57 436, representing a 16.8% increase compared with 2012 [2].

1.2. Motivation and goals

Due to its magnitude and importance for the treatment of oncological diseases, radiotherapy has been a field with rapid and growing technological development through the years.

From the advent of intensity modulated radiotherapy, to the development of new treatment delivery techniques and the integration of image guidance in treatment units, the goal has always remained the same: to achieve the maximum possible tumour control with the least possible compromise of the surrounding normal tissues and organs.

Recently there has been a growing tendency to shift the approach of healthcare in general from standardized processes to individualized applications, which are optimized to the specificities of each patient thus maximizing the treatment benefit. This is referred to as precision medicine, and the radiotherapy field is beginning as well to incorporate it into its workflows, namely with the implementation of adaptive radiotherapy. Since its introduction there has been a great deal of development in the field of ART, powered by advances in technologies for automatic segmentation, deformable image registration and automatic planning [4].

These developments, aligned with the growing interest in applying machine learning techniques to the realm of health care (such as radiomics), potentiate the optimization of adaptive radiotherapy application and consequently the benefits it can bring to the patients' treatment. Nevertheless, the implementation of ART into clinics' workflow is still a rather expensive process, as it requires extra human and equipment resources, making it impossible for centres to perform it regularly for every patient. Furthermore, as will be exposed later, it has been shown that not all patients benefit equally from ART, making it an unnecessary effort for some.

Currently, there are four main questions which still remain to be solved regarding the implementation of adaptive radiotherapy [5]:

- Do patients actually *need* adaptive radiotherapy? Does it increase local control and/or decrease toxicity?
- Which patients need adaptive radiotherapy?
- What are the optimal number and timings for replanning?
- Is there a positive trade-off between the increase in patients' quality of life and the human and economic resources required for the implementation of adaptive radiotherapy?

This dissertation is divided in two parts, addressing the first two questions. In the first part the goal is to assess the need for ART in Head and Neck patients. The dosimetric impact of the implementation vs non-implementation of ART for both the target volumes and surrounding organs at risk is investigated, using an automatic planning tool. The objective of the second part of this work is to develop models to predict which patients will require ART during treatment, in other words, which patients are more likely to

achieve the benefits found in the first part of the work through adaptive radiotherapy. Three different approaches to this problem were investigated: 1) using only patients' pre-treatment clinical data; 2) using radiomic features extracted from the target volumes; 3) using a combination of 1 and 2.

1.3. Organization of the dissertation

This dissertation is organized into 6 chapters:

- **Chapter 1 – Introduction**: define the scope of the dissertation and corresponding goals.

- **Chapter 2 – Background Knowledge**: in this chapter some important concepts for the full understanding of the work developed in this dissertation will be defined. Firstly, a brief section about the clinical rationale for radiotherapy is presented, moving to some of the more relevant technological developments, the implementation of image guided radiotherapy and finally adaptive radiotherapy. A section on head and neck cancer is also presented to contextualize the pathology.

- **Chapter 3 – State of the art:** in this chapter a brief literature review is presented, considering the topics of anatomic modifications of both target volumes and organs at risk during radiotherapy treatments, the dosimetric impact of these alterations, the impact of adaptive radiotherapy and the prediction of adaptive radiotherapy in Head and Neck patients.

- **Chapter 4 – Assessing the need of adaptive radiotherapy in H&N cancer patients:** In this chapter a retrospective planning study to assess the need for ART is presented. The work was based on an automated planning platform to generate plans in the adaptive and non-adaptive scenarios. The chapter is sub-divided into 5 sections: purpose, material and methods, results, discussion, and conclusion.

- **Chapter 5 – Predicting the need for adaptive radiotherapy in H&N patients:** in this chapter the work developed in predicting the need for ART from pre-

treatment data and CT-based radiomics is presented. The chapter is sub-divided into 5 sections: purpose, material and methods, results, discussion, and conclusion.

- **Chapter 6 – General conclusions and future work**: the global conclusions of the dissertation are presented, as well as the next steps need to complement the achieved results.

The work described in chapters 5 and 6 was submitted to two relevant journals in the field of Medical Physics in the form of original research papers.

In addition to these chapters there is a section of appendixes in the end of the dissertation which contain complementary information about the developed work, and which are appropriately referred to in the text.

Chapter 2

Background Knowledge

2.1. Radiotherapy

Radiotherapy (RT) is a cancer treatment modality that relies on the usage of ionizing radiation, such as photons, electrons, protons, or heavy ions, to deliver toxic levels of energy to tumorous cells. The aim of the radiation is to cause damage to the cancer cells' genetic material (namely to the deoxyribonucleic acid (DNA) molecules), leading to their controlled destruction as they lose the ability to proliferate [6].

RT can be delivered externally, through linear accelerators (LINACs), or internally, by the placement of radioactive sources inside the patient (brachytherapy). For the purpose of this work, the term radiotherapy will be used to refer to external beam radiotherapy.

RT can be used either as the only treatment approach, as a complement to surgery (adjuvant radiotherapy) or with concurrent chemotherapy. Furthermore, it can be performed either with a curative/definitive intention (radical radiotherapy) or with the goal of improving quality of life in patients for which curative treatment is not possible (palliative radiotherapy).

The aim of RT is to deliver enough radiation to the tumour in order to achieve local control, whilst maintaining the dose to normal tissues sufficiently low to avoid serious complications. This goal is achieved by balancing tumour control probability (TCP) and normal tissue complication probability (NTCP) with respect to delivered dose. Although these are highly tumour-specific, both TCP and NTCP curves have sigmoid shapes which are shown in figure 1, being the optimum endpoint of radiotherapy to maximize TCP while simultaneously minimizing NTCP for a given tumour [7]. This corresponds to achieving the maximum of the curve represented in green in figure 1 (P_+), which is the probability of achieving tumour control without normal tissue complications [8].

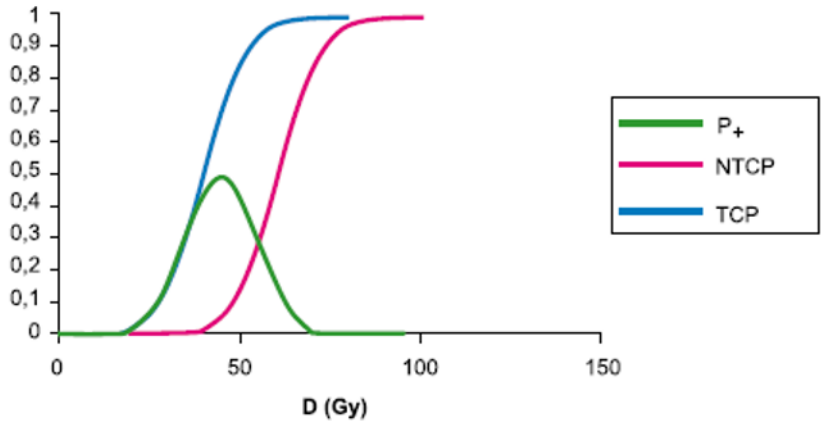

Figure 1. The blue curve represents the tumour control probability and the pink curve the normal tissue complication probability with respect to dose (Gy). The green curve is the probability of tumour control without normal tissue complications (Adapted from [8]).

RT is based on the idea that, generally speaking, the DNA repair capacity is greater for healthy cells when compared to tumorous ones [6]. The separation between the NTCP and TCP curves defines what is called the *therapeutic window*, which is the dose range in which treatment may successfully eradicate tumour while maintaining normal tissue tolerance. The therapeutic ratio, or therapeutic index, has been defined as the ratio between TCP and NTCP at a specified level of probability of response (typically 0.5) [7]. Both a larger therapeutic window and a larger therapeutic ratio are desirable, since a more narrow distance between the TCP and NTCP curves translates into the need for more strict and precise determination of administered dose to secure the success of treatment [8].

There are several ways to widen the therapeutic window, ranging from more conformal treatment delivery techniques (which will be discussed later on), to the application of radiosensitisers or concomitant chemotherapy [6].

Radiobiological studies have shown that delivery of the dose in multiple fractions spread out over a period of weeks (fractionated RT), improves the therapeutic ratio when compared to the delivery of a large dose in one single session. The rationale is that for fractionated RT the time span between the delivery of each treatment fraction allows for the repair of sub-lethal damage and repopulation of the normal cells, while increasing tumour damage by allowing reoxygenation of the tumour cells and their reassortment into a radiosensitive phase of the cell cycle [6].

Currently, the standard fractionation regimen consists of five daily treatments per week, with a two day pause on the weekends. This system integrates practical aspects of the dose delivery to the patient, a successful treatment outcome and the convenience for the staff delivering the treatment [7].

2.1.1. Radiotherapy Workflow

A schematic representation of the typical general workflow of an oncological patient treated with fractionated radiotherapy is shown in figure 2.

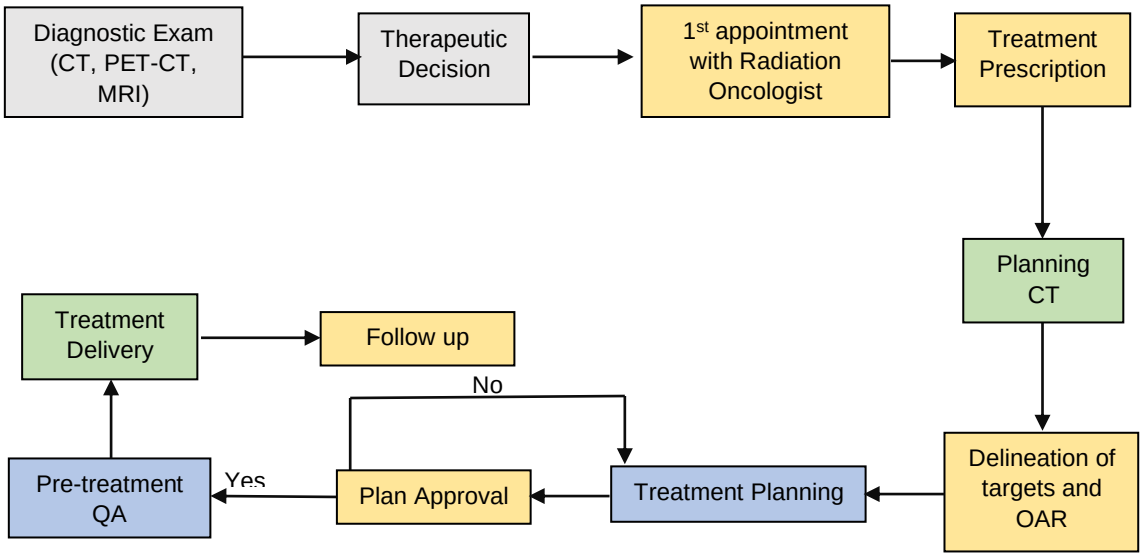

Figure 2. General RT workflow. The yellow, green and blue boxes represent tasks of the responsibility of the Radiation Oncologists, the Radiation Treatment Technologists and the Medical Physicists, respectively.

After the diagnosis of the oncologic disease, there is a therapeutic decision appointment where the treatment modality or modalities to be used are defined by a qualified clinical board. If RT is the recommended treatment, the patient is referred for a first appointment with the radiation oncologist (RO), that evaluates the individual case and prescribes the RT treatment: dose to be delivered to the tumour and/or adenopathies and lymphatic nodes, as well as the fractionation regimen.

After this first appointment, the patient is submitted to a planning computed tomography (CT) scan, on which the RO delineates the target volumes (tumour or tumoral site, and/or lymphatic nodes, adenopathies), and the surrounding organs at risk (OAR).

There are 3 target volumes defined in RT, according to the International Commission on Radiation Unites and Measurements reports 50 and 62 [9,10]:

1) The gross target volume (GTV), which is the gross palpable or visible/demonstrable extent and location of the malignant growth.
2) The clinical target volume (CTV), which is a tissue volume that contains the GTV and/or subclinical microscopic malignant disease, which must be eliminated.
3) The Planning Target Volume (PTV), which is a geometrical concept defined to select appropriate beam sizes and beam arrangements, taking into consideration the net effect of all the possible geometrical variations and inaccuracies (positioning, motion, and anatomical changes) in order to ensure that the prescribed dose is actually absorbed in the CTV.

A schematic representation of these volumes is shown in figure 3. In addition to these target volumes, the organs at risk that surround the tumour, and which should be spared to the maximum possible extent from incoming radiation also need to be delineated either manually or with the help of automatic contour software.

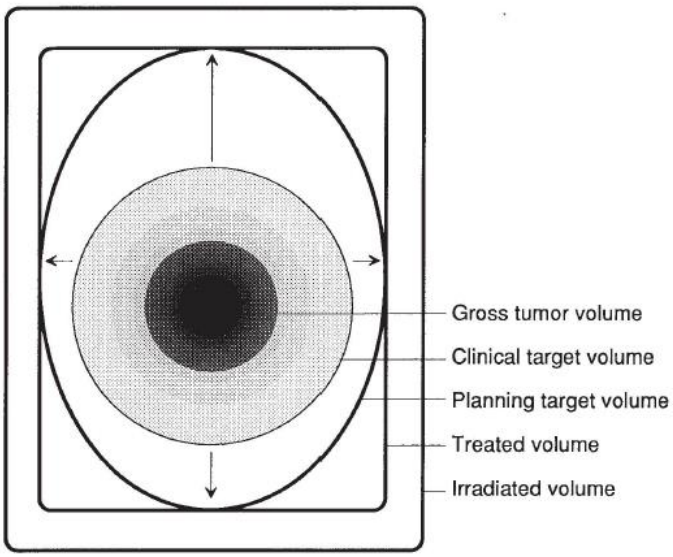

Figure 3. Schematic representation of GTV, CTV and PTV, taken from ICRU report 50 [5]. The treated volume is the volume that receives a dose that is considered important for local control, and the irradiated volume is the volume that receives a dose that is considered important for normal tissue tolerance

After the delineation of the volumes, the medical physicist is able to produce a treatment plan based on the planning CT, with the objective of delivering the radiation dose

prescribed by to the PTV(s), whilst reducing the dose received by the adjacent OAR to clinically acceptable values.

The goal of radiotherapy treatment planning is to generate beam shapes and distributions that maximize tumour control and minimize normal tissue complications. This is done by the medical physicist or dosimetrist using computerized treatment planning systems (TPS), in which the patient's tumour and anatomy are represented in three-dimensional (3D) models and dose distributions can be computed [7]. There is a complex multi-objective optimization problem inherit to treatment planning, making it still a rather time-consuming task, and the end result highly dependent on the planners' experience and skill [11].

This plan is then assessed by the RO, whose responsibility is to verify that the dose constraints to the normal structures are being met and that the prescribed dose is delivered to the target volumes. This is routinely done through the analysis of dose volume histograms (DVH) for each structure, which can be either differential, showing the fractional volume that receives a specific percentage of the prescribed dose, or cumulative, showing the fractional volume receiving at least a specific percentage of the prescribed dose (figure 4) [12]. If the radiation oncologist approves the plan, there is a pre-treatment quality assurance performed by the medical physicist, following a protocol that is defined by each institution. After this step, the patient can start the treatment.

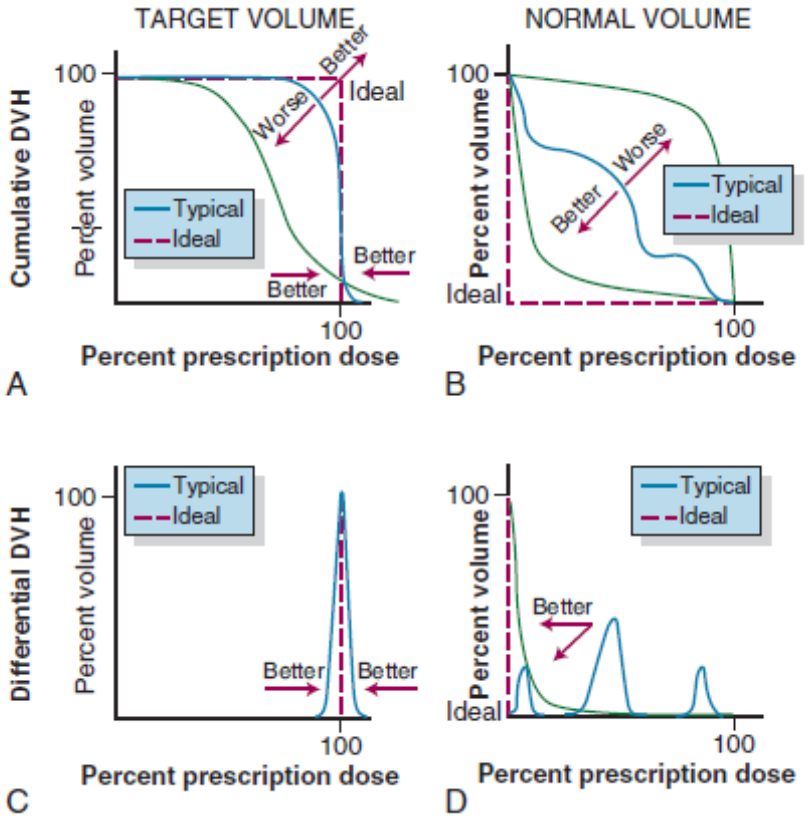

Figure 4. Cumulative and differential dose volume histograms. A) Cumulative DVH for target volume; B) Cumulative DVH for normal tissue or organ; C) Differential DVH for target volume; D) Differential DVH for normal tissue or organ. The typical curves are represented in blue and the ideal curves in dashed red. ([12] page 130).

2.1.2. 3D conformal and intensity modulated RT

The goal of technological advances in RT has always been to improve clinical outcome, by scaling dose to the target volumes and reducing toxicity in normal tissues [13]. This can be achieved by conforming the radiation dose to the tumour as closely as possible through beam shaping, which in the early days of RT was done through custom-designed metal blocks mounted in the head of the treatment machines (figure 5) [14]. Since the early 1990's, these traditional shaping methods started to be replaced by multi-leaf collimators, which consist of small metallic leaves located in the head of the LINAC that can move independently in and out, blocking specific fractions of the radiation field and resulting in highly conformal dose distributions while maintaining versatility [7,14].

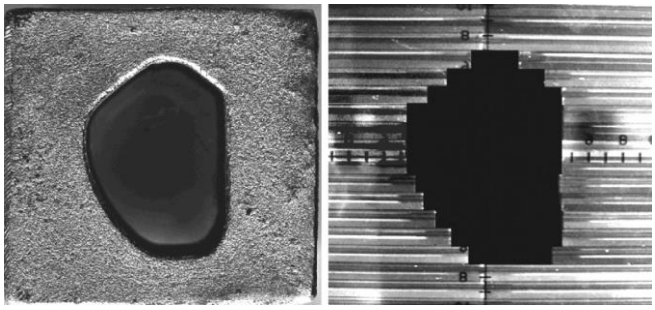

Figure 5. Shaped block (left) and multileaf collimator (right). ([12] page 296).

From the 1950s to the late 1980s, the two-dimensional (2D) radiation therapy largely dominated the field. Conventional planar radiographies of the patient were acquired, on which bony anatomy reference points could be identified and used as clues for the volumes of interest. The plans were then created on a very limited range of images, using standardized beam arrangements, and only rectangular and symmetrical collimation was performed by manually applying shielding blocks [15].

The technological advances of the 1990's, namely the advent of 3D imaging modalities such as CT, positron emission tomography (PET) and magnetic resonance imaging (MRI), which allowed for 3D target location, as well as advances in TPS technologies and delivery techniques, allowing for 3D treatment planning and delivery, potentiated the transition from 2D radiotherapy to a 3D, highly conformal paradigm [7,15].

According to the International Atomic Energy Agency (IAEA), 3D conformal radiotherapy (3D CRT) is *"the term used to describe the design and delivery of radiotherapy treatment plans based on 3-D image data with treatment fields individually shaped to treat only the target tissue"* [16]. In 3D-CRT, the beam arrangement (number, directions and intensities of the irradiating beams) is determined by the planner in a direct, trial and error approach, called forward planning. The radiation intensity is uniform within each beam, being the end goal to obtain a dose distribution as conformal and homogenous within the target as possible [14]. The only way to modulate the radiation intensities is through wedges, which are angled pieces of lead or steel placed in the beam, leading to modulation in just one direction [7].

Intensity modulated radiation therapy (IMRT) differs from traditional 3D-CRT since the beams have non-uniform radiation fluences, which create highly conformal dose distributions, with steep dose gradients in the tumour borders [9-12] . It has been shown that this technique can produce homogenous dose distributions within the target comparable to 3D-CRT, but it is able to achieve greater conformality, namely for concave or other complex-shaped target volumes, thereby further sparing surrounding normal

tissues [18]. Furthermore, it facilitates the achievement of non-uniform dose distributions in the cases where a volume is partially or completely contained in another volume [18].

This increase in dose conformation can be seen in figure 6, which shows the dose distributions for a head and neck patient achieved with 2D-RT, 3D-CRT and IMRT. It is clear that not only the total volume which receives dose is severely reduced from 2D-CRT to IMRT, but also that the volume outside the target (yellow contour) which receives a high dose (represented in red) is significantly smaller.

In practice, this illustrates the potential of IMRT to improve clinical outcome by widening the therapeutic window (see figure 1), both by increasing TCP and decreasing NTCP, with multiple studies showing its dosimetric and clinical benefits [16]. Since the mid 1990's, there has been a massive growth in the development and implementation of IMRT worldwide, replacing 3D-CRT as the gold standard for radiotherapy treatment [7].

The complex beam intensity patterns made possible by IMRT greatly increase the number of possible tuning parameters in planning, making forward planning no longer possible. Instead, an inverse approach was taken where the planner feeds the doses and dose/volume constraints for the tumour and surrounding normal organs to a dose-optimization algorithm, which iteratively tunes all parameters (shape, direction, weights, etc.) and determines the fluence of each field in order to meet the objectives set by the planner [14,17].

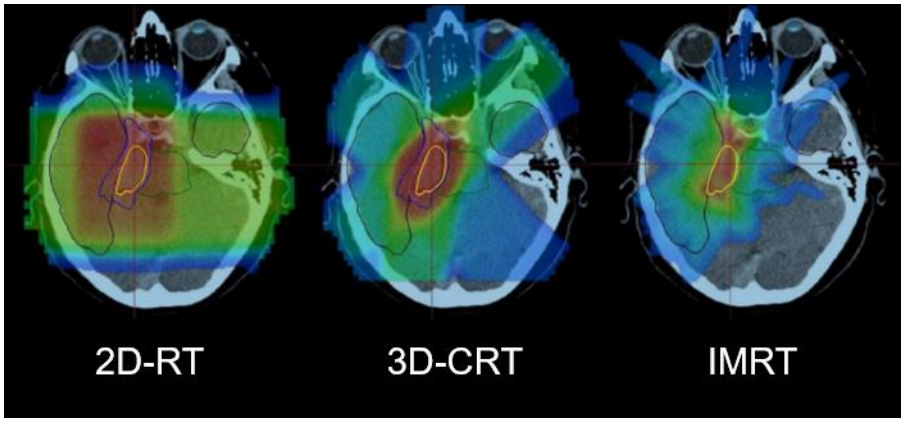

Figure 6. Dose distributions in an axial CT slice of the same patient using 2D-RT, 3D-CRT and IMRT. The red areas represent higher doses and the blue areas lower doses. It can be seen the progressively larger high-dose conformation to the target volume (yellow contour), as well as normal tissue sparing [19].

2.1.2.1. IMRT delivery

The key element for the delivery of IMRT in LINACs is the MLC, which can operate in one of 2 ways: step and shoot (static delivery), where the leaves are fixed in a certain position, irradiate, stop irradiating, move to a different position, and so on, creating an intensity profile, or dynamic delivery, where the leaves are continuously in motion during delivery [20].

IMRT delivery by LINACs can be done in a fixed gantry angle or volumetric arc approach. In the fixed gantry angle approach the gantry sequentially moves to different incidence angles, only irradiating when it arrives at a fixed position. Alternatively, in volumetric arc therapy (VMAT) the LINAC's gantry is continuously rotating around the patient while emitting a cone beam that can be modulated by the multileaf collimator, variable dose rate and variable gantry speed [15]. This allows for significantly reduced treatment times, as well as lower integral dose when compared to the fixed-gantry approach [20].

A more recent technology which was developed specifically for the delivery of IMRT is Helical Tomotherapy. The helical tomotherapy unit consists of 6 megavoltage (MV) LINAC mounted onto a CT-type ring gantry, which rotates with constant speed around the patient, while the patient's couch slides through it [21]. In helical tomotherapy, the radiation is delivered continuously by the rotating 6 MV intensity modulated fan beam, in a "slice by slice" manner. This approach allows for very conformal dose distributions and treatment of larger fields when compared to VMAT or fixed gantry IMRT [15]. A picture of a LINAC and a schematic representation of a tomotherapy unit are shown in figure 7.

Figure 7. LINAC [22] at the left and schematic representation of a helical tomotherapy unit [23] at the right.

Each one of these treatment techniques has its own advantages and disadvantages, being the choice of which to use highly dependent on the specific patient anatomy, the intention of treatment and the clinic's available resources.

2.2. Image Guided Radiotherapy

Radiotherapy treatments typically have a duration of several weeks, during which there can be modifications to the patient's anatomy (such as weight gain/loss, or tumour regression) that lead to discrepancies from the planning CT acquired prior to the start of treatment. Furthermore, the daily positioning of the patient can be itself a source of error, as a mismatch with the position in the planning CT would result in deviations on the delivered dose to the patient.

In order to correct for day-to-day positioning errors relative to the planning CT, most centres perform daily position verification through imaging. This process is called Image Guided Radiotherapy (IGRT) and its general workflow is shown in figure 8. Imaging techniques such as in room kilovoltage CT (kV-CT), kilovoltage (kV) or megavoltage (MV) cone beam CT (CBCT) and helical MV-CT (in the case of Tomotherapy) are used to obtain a daily image of the treated area, which is then compared to the planning CT, and, if necessary, the patient's and/or couch position is corrected before the delivery of the fraction.

However, one of the main limitations of IGRT is that internal changes in size, shape or relative position of the target volumes or OARs compared with the initial planning CT cannot be corrected by rigid translational and/or rotational couch shifts, and may lead to deviations between the actual dose delivered to the patient and the planned dose. These dose deviations may result in underdosing of the target volume and/or overdosing of the OARs, leading to lower tumour control and/or more severe radiation-induced complications [24].

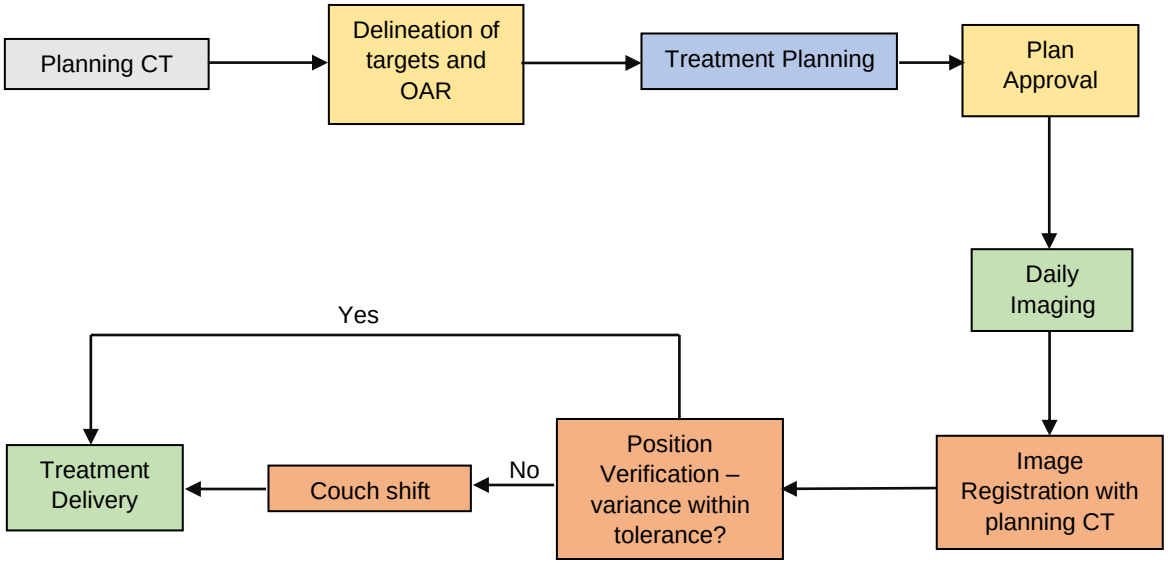

Figure 8. Typical RT workflow. The orange boxes mark the IGRT steps.

2.3. Adaptive Radiotherapy

One possible solution for this problem is adaptive radiotherapy (ART), which was first introduced by Di Yan, Frank Vicini, John Wong and Alvaro Martinez in 1997 as *"a closed-loop radiation treatment process where the treatment plan can be modified using a systematic feedback of measurements"* [1].

The goal of ART is to take into account any anatomical changes that might occur during the radiotherapy treatment and re-optimize the treatment plan in order to compensate for them, improving the overall treatment quality.

With the advances in technology for daily imaging, deformable image registration, auto-segmentation and dose re-calculation, ART has largely evolved since its appearance in 1997 and can be currently classified in three types, according to the time when the process of re-optimization is performed [25]:

1) Offline ART if the process occurs between fractions.
2) Online ART if it occurs immediately prior to a fraction.
3) Real-time ART if it occurs during a fraction.

IGRT is intimately connected to ART, as daily imaging is a fundamental part of all kinds of ART workflows, being the trigger that warns the radiotherapy technicians, the medical physicists and radiation oncologists that an adjustment must be made.

2.3.1. Offline ART

Offline ART is the most common type of ART due to its reduced complexity and to the fact that more advanced tools such as auto-segmentation and optimization are not routinely available for most clinics. An example of a typical workflow for offline ART is described in figure 9.

The first step is detecting the need for an adjustment in the treatment plan. If the treatment centre has access to daily imaging this can be done by performing a rigid registration between the planning CT and the image of the day, thus obtaining a transformation vector, which can be translated into a 3D couch shift. If there is no access or daily imaging cannot be performed for all patients, direct observation of the patient's anatomy (severe gain or loss of weight, poor fit of the thermoplastic mask) is used as an indication that replanning is probably needed.

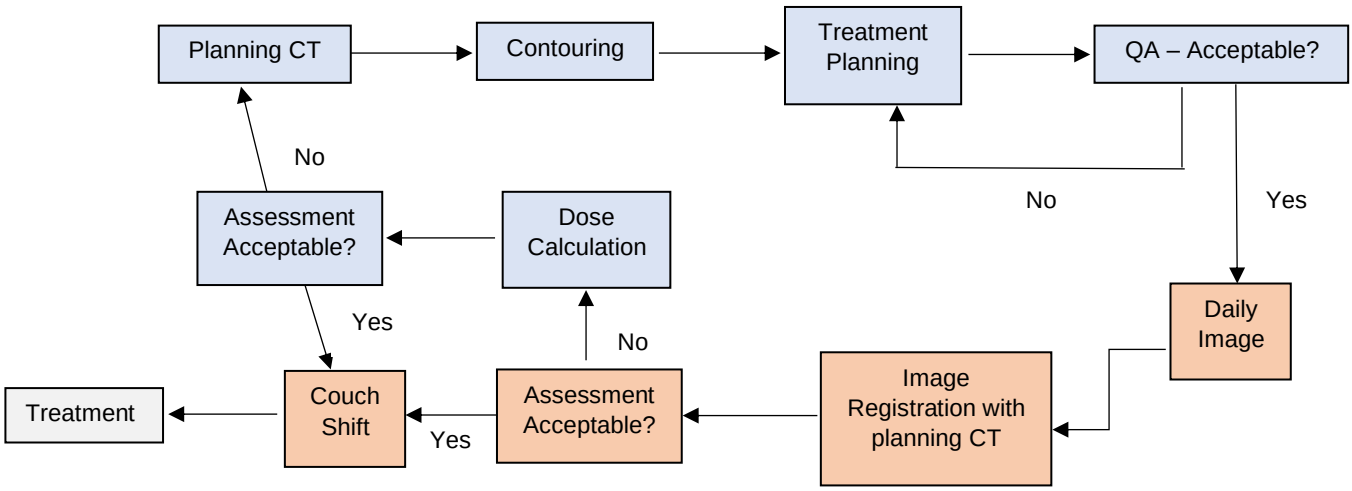

Figure 9. Typical offline ART workflow. The orange boxes mark the steps of a typical IGRT workflow (see figure 8) and the blue boxes those characteristic of offline ART.

After the verification that there is a mismatch between the planning CT and the daily image, calculation of the daily dose distribution in the new image or deformable dose mapping and estimation of cumulative dose can be performed in order to make the decision of whether a replan is necessary.

If the obtained dose distribution is not acceptable, the treatment is interrupted, a new planning CT scan (re-CT) is acquired, and all the planning process is repeated.

2.3.2. Online ART

The difference between offline and online ART is that not only the assessment for the need of replanning, but also the replanning itself occur immediately before the delivery of the fraction, while the patient is lying on the couch. In this way, there is no acquisition of a second CT scan and the daily image is used for the replanning, which needs to be fully automated as the overall process cannot take more than a few minutes. This requires the use of more advanced imaging technologies and techniques such as automatic segmentation and automatic planning, which are not commonly available for most centres [26].

Another ART approach that is a kind of combination between offline and online ART is the “plan of the day”. In this scenario, a library of plans is created for each patient, either in a generalized approach, by applying different population-based margins around the planning CTV to generate different PTVs, or in a patient-specific approach, where different images of the patient are obtained, representing different possible anatomical arrangements (for example the bladder with different levels of filling), and a plan is generated for each of them [27].

Then, at each fraction, the daily image is analysed, and the best fit plan is selected for treatment. This approach is more commonly used for bladder, rectum, prostate cancers, and other treatment sites that may present severe anatomical changes in a daily basis [27].

2.3.3. Real-time ART

The concept of real-time ART arises from tumour and OAR motion caused by the patient’s respiratory, circulatory, digestive and muscular systems on a time scale of seconds and minutes, which cannot be captured by a single pre-treatment image of the treatment site. The objective of real-time ART is to address this issue by detecting, measuring and correcting for anatomic changes during treatment delivery [28].

This is a very complex and technologically expensive process, which requires that high-quality imaging as well as automatic, almost instantaneous plan re-optimization occurs simultaneously to the delivery of each treatment fraction, with the plan to deliver the desired dose being continuously adjusted [28].

One of the main challenges for the implementation of real-time ART is that the images provided by the current in-room imaging techniques such as CBCT or MV-CT have low soft tissue contrast, making them unreliable for the definition of new target volumes.

Furthermore, these imaging techniques rely on the use of ionizing radiation, which translates into undesirable additional dose to the patient [29].

These limitations led, in recent years, to the investigation of magnetic resonance imaging as a possible alternative, leading to the development of the first MRI-LINAC (MRIdian) by ViewRay in 2014 [30]. This new system has the potential of providing high quality magnetic resonance imaging, with enough soft tissue contrast for target delineation, without the extra radiation dose of CTs. These images can be obtained simultaneously with irradiation by the integrated LINAC, making real-time ART possible.

However, there are major difficulties in the integration of a LINAC with an MRI machine, namely the geometrical problem of how to merge two very bulky systems and the physical problem of shielding the LINAC from the strong magnetic fields of the MRI machine, since many electronic components are susceptible to electromagnetic interference [29]. Currently, there are only two commercially available and clinically used models of the MRI-LINAC: the MRIdian from Viewray (USA) and the Unity from Elekta (UK), both of which are still not accessible for most clinics due to their cost.

Real-time ART may be very useful for tumours in areas with constant motion such as the lungs due to breathing. However, the implementation of real-time ART comes at the cost of significantly more advanced and automated tools, as well as more strict quality assurance, and it has not yet been generally implemented [25].

The choice of which ART regime to apply depends largely on tumour location, clinical benefit and resources availability, being its implementation still not a reality for all centres.

In the case of Head and Neck patients, as the main anatomical changes occur over the course of weeks, in a gradual manner, it is argued that online/real-time ART might be unnecessary for most patients, which has the benefit of making ART easier to incorporate in the workflow of most clinics [26,31].

2.4. Head and Neck cancer

Head and neck (H&N) cancer is currently the seventh most common type of cancer worldwide, with 550 000 new cases and 380 000 deaths every year [32]. The term H&N cancer refers to a heterogeneous group of tumours, arising from the upper aerodigestive tract (the oral cavity, lip, oropharynx, hypopharynx, nasopharynx, larynx), paranasal sinuses, and salivary and thyroid glands [33]. These tumours usually have their origin in

the squamous cells lying in the mucosal surfaces inside the head or neck, such as the mouth, nose and throat, and are classified as squamous cell carcinomas [34].

The most common risk factors for H&N cancer are tobacco and alcohol consumption, being the combined effect of both substances greater than one of them alone [35]. Another possible cause for H&N cancer is infection with the Human Papilloma Virus type 16 (HPV-16) [36].

Treatment of H&N cancer requires a multidisciplinary approach, often demanding a multimodal treatment, including surgery, radiotherapy, chemotherapy or targeted therapy. The choice of treatment depends on several factors, such as tumour grade, stage and location [36], surgery and radiotherapy being the most used modalities [33].

In current clinical practice, the vast majority of patients with locally advanced H&N cancer requires either definite, adjuvant or palliative radiotherapy, with or without concurrent chemotherapy [36].

Radiotherapy treatment of H&N cancer is a complex task as there are multiple structures, such as the spinal cord, the brainstem and the parotid glands, that are critical to the patient's quality of life, located close to, and often overlapping, the target volumes. Acute side effects of radiotherapy include acute mucositis, leading to soreness of the mouth and throat, deterioration of the taste buds lining the tong, oral candidiasis (oral thrush), acute skin reactions, dysphagia (difficulty to swallow), and laryngeal edema accompanied by hoarseness [33]. These acute side effects, which typically start during treatment and can persist up to 3 months after its completion, can severely impair the ability for oral food intake, and may lead to the need for nutritional supplements via nasogastric tube [33]. As a result, H&N patients are very prone to substantial weight loss during treatment as well as dehydration [35].

As for late complications, the most common effect is xerostomia, arising from degradation of the parotid glands (which are responsible for the salivary function) by the ionizing radiation. Xerostomia manifests as a dryness of the mouth, which can be associated with a change in the composition of saliva or reduced salivary flow, and it is known as being responsible for difficulty in swallowing, nutritional deficiency, compromised oral hygiene, poor dental condition, altered taste sensation, impaired speech function and poor sleep quality [33].

Other late effects of radiotherapy include osteoradionecrosis, trismus, subcutaneous fibrosis of the soft tissues and laryngeal edema. Neurologic complications, such as cranial nerve injuries, are the most severe complications of radiotherapy for H&N

patients. Furthermore, optic nerve injury or radiation retinopathy resulting in blindness can also occur if the orbit and optic chiasm are over irradiated [33].

The highly conformal dose distributions produced by IMRT can help reduce the dose to the OAR (and consequently dim radiation induced side effects) whilst assuring full target coverage in H&N cancer patients. Several studies have concluded that IMRT significantly reduces the dose to the parotid glands, which are often very close to the targets, resulting in lower levels of xerostomia compared to conventional techniques [37–40].

However, the steep dose gradients resulting from IMRT are extremely sensitive to positional errors and anatomic changes, since small changes in position can result in large changes in dose to the tumour and OARs.

Chapter 3

State of the art

3.1. Anatomical and dosimetric changes during RT for H&N patients

H&N radiotherapy treatments typically have a duration of 5-7 weeks. During this period, several factors can contribute to deviations between the dose distribution planned by the medical physicist, and approved by the radiation oncologist, and the actual dose delivered to the patient in subsequent fractions. These include daily setup variations, primary tumour or nodal volumes regression or progression, alteration in muscle mass and/or fat distribution, fluid shift within the body and weight loss, and may lead to changes in the locations, shapes and sizes of both tumours and OARs [41].

Figure 10 shows such an example. In this case, a CT scan was acquired 2 weeks after the beginning of the treatment and the anatomical changes to the target volumes and OAR, combined with severe patient weight loss (which is very common in H&N cases), led to a loss of conformity of the original tumour contour.

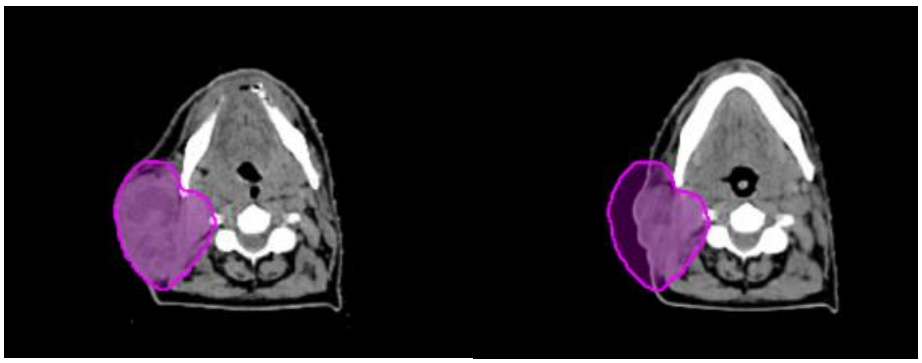

Figure 10. In the right side of the image, the planning CT and tumour contour for a given H&N patient is shown. In the left side, a repeat CT for the same patient acquired 2 weeks after the beginning of treatment.

These alterations need to be detected and corrected as soon as possible, in order to prevent target underdose and/or OAR overdose, especially when dealing with IMRT due to the sharp dose gradients in the target-organ interfaces [42].

3.1.1. Target volumes

One of the first studies to quantify the magnitude of anatomical changes to the GTV was conducted by Baker et al. (2004), using an on-board kV CB-CT to acquire 3 CT scans per week for each of the 14 patients in the study [42]. The GTV was manually delineated by a radiation oncologist in each of the CT scans and the volumes were then compared using a linear regression. The authors concluded that GTVs decreased in volume during treatment at a median rate of 1.8% per treatment day, corresponding to a median total relative loss of 69.5% of the initial GTV on the last day of treatment. They also concluded that this tumour loss was often asymmetrical, as the centre of mass of the GTV changed its position with time, reporting a median displacement of 3.3 mm at treatment completion. The authors observed that the absolute volume loss was larger for large initial tumours/nodes.

Another study by Geets et al. (2007), analysed 10 patients treated with helical tomotherapy and concomitant chemotherapy [33]. The patients were weekly submitted to a CT, T2-MRI, fat suppressed T2-MRI and static and dynamic fluorodeoxyglucose (FDG)-PET scans during treatment. The authors concluded that the GTVs (and consequently CTVs and PTVs) significantly decreased throughout the course of RT ($p<0.001$) for all imaging modalities, with a relative tumour reduction by the 4th week of treatment ranging from 54% to 70% in comparison to the pre-treatment GTV [43].

Bhide et al. (2010) conducted a prospective observational study of weekly volume changes during chemoradiotherapy with 20 H&N patients that had weekly CT scans in the course a 5-week treatment [44]. The authors reported that both tumoral and nodal CTVs reduced in volume during treatment, with the most significant reductions occurring at week 2, corresponding to a mean percentage volume reduction of 3.2% for tumoral CTV ($p=0.003$) and 10.5% for nodal CTV ($p<0.001$). They also observed that this reduction in volume resulted in a reduction of the minimum dose delivered to the PTVs throughout treatment, with a significant ($p=0.002$) mean reduction of 2 Gy for the tumoral PTV in the 2nd week of treatment. Finally, it was reported that the dose range across the PTVs also increased during treatment, indicating reduced dose homogeneity. This range increase was more significant from the beginning to week 2 of the treatment both for tumoral PTV ($p<0.001$) and nodal PTV ($p=0.008$).

A more recent study by Mnejja et al. (2020), included 20 nasopharyngeal carcinoma patients treated with concomitant chemoradiotherapy in 33 fractions [45]. For each patient, three treatment volumes were defined from the nodal and tumoral GTVs: one high-risk PTV (PTV-H) that was prescribed 69.96 Gy, one intermediate-risk PTV (PTV-I)

that was prescribed 60 Gy and a low-risk PTV (PTV-L) that was prescribed 54 Gy [45]. In addition to the planning CT, all patients had a 2nd CT scan at 38 Gy, around treatment fraction 20. The authors reported significant ($p<0.001$) reductions in both nodal GTV (58.56%) and tumoral GTV (29.52%) between the two CT scans. This reductions translated into a significant decrease in the mean dose received by 98% of the volume, corresponding to 1.4 Gy ($p=0.007$) for PTV-H, 0.3 Gy ($p=0.03$) for PTV-I, and 1.15 Gy ($p=0.007$) for PTV-L.

Conversely to these results, some studies report that, despite the existence of volume reduction, dosimetric coverage of target volumes tends to be robust to anatomical changes in the GTV/CTV/PTV [31,46,47]. For instance, Wu et al. (2009) [31] showed that anatomical and dosimetric changes observed in several structures of 11 H&N patients that underwent weekly CTs did not lead to significant differences in the dose delivered to the primary (tumoral) CTV, with small increases in the minimum dose delivered to the nodal CTV [31].

3.1.2. Organs at risk

Regarding the organs at risk, Brouwer et al. (2015) [48] published a comprehensive review article, where 51 papers that reported anatomical and dosimetric changes in several OARs during radiotherapy treatments were analysed. The majority of the studies focused on alterations in the parotid glands, due to the association between radiation induced effects on the parotids and xerostomia.

The average reduction of the parotid glands' volume was reported to be 26% ± 11%. One study by Sanguineti et al. (2013) [49] reported that this shrinkage was not linear, being more pronounced in the first half of the treatment. In addition to volume loss, a medial shift of the parotids was also reported by several studies.

As for the dosimetric impact of these volumetric alterations, a mean increase of 2.2 ± 2.6 Gy in the mean dose to the parotid glands was reported (24 papers). Furthermore, five studies found significant associations between these volume changes and complications such as increased xerostomia, reduced saliva production and increased mucositis, all of which leading to decreased patient's quality of life.

Three studies reported on the effects of RT on the submandibular glands, showing an average volume reduction of 22% (15–32%). One study also observed superior and medial shifts of the submandibular glands, as well as increased delivered dose when compared to the planned dose (52.8 Gy vs 51.9 Gy).

Regarding the spinal cord and the brainstem, several authors reported an increase in the maximum dose or in the $D_{1\%}$ (the dose that is received by 1% of the volume - the near-maximum dose), being the highest reported average increase of 0.2% per fraction for the spinal cord and 0.09 Gy per fraction for the brainstem, which then results in an accumulated excess dose of 5.6 and 2.5 Gy, respectively, for the entire treatment course.

The ultimate result of all these anatomical/dosimetrical changes is that highly conformal IMRT plans based on a single planning CT acquired several days prior to the beginning of radiotherapy treatments may lead to decreased tumour control and/or unexpected complications, since the dose that was planned to be delivered to the target volumes may be misplaced to other organs, if anatomical and positional uncertainties are not adequately taken into account [41].

3.2. Dosimetrical and Clinical benefits of ART for H&N patients

Deformable image registration, dose mapping and dose accumulation have made it possible to compare the dose distributions achieved by IGRT alone and ART [26].

The first prospective study reporting the dosimetric effects of ART was developed by Schwartz et al. (2012) [50] and included 17 patients, all of which had 1 replan in the course of treatment, with 4 requiring a second replan. The study compared 4 different planning scenarios: 1) original IMRT plan aligned daily to the marked isocentre, simulating treatment with no correction; 2) original plan rigidly aligned daily to bone, simulating IGRT; 3) IGRT with 1 adaptive replan mid-treatment; 4) actual treatment scheme received by the patient (IGRT + 1 or 2 replans) [50].

By using deformable dose mapping and dose accumulation, the authors were able to calculate the cumulative total delivered dose for each scenario. The results showed that one mid-treatment replan (scenario 3) reduced the mean delivered dose to the contralateral parotid by 2.8 %, corresponding to 0.6 Gy ($p=0.003$), and for the ipsilateral parotid by 3.9%, corresponding to 1.3 Gy ($p=0.002$) when compared to IGRT (scenario 2). It was also observed that for the 4 patients which had a second replan there was a further reduction of the mean dose to the contralateral parotid of 0.8 Gy ($p=0.026$) and of the ipsilateral parotid of 4.1 Gy ($p=0.001$) when compared to IGRT. The authors also reported that ART significantly reduced the integral body dose and did not lead to underdosing of the CTV [50].

In a second study regarding these patients clinical outcomes, the authors report a 100% local and 95% regional disease control at 2 years (with a median follow up of 31 months),

and an acute toxicity at 1 year equivalent to toxicity profiles observed under conventional IMRT [51].

Zhao et al. (2011) [52] conducted a study that retrospectively analysed 33 patients which had one repeat CT scan and replanning in the course of treatment (ART group) and a control group with 66 patients that did not have any replanning (non-ART group). For the patients in the ART group, a "hybrid plan" was generated to represent the situation in which no replanning occurred, by applying the beam configurations of the first IMRT plan (including the intensity profile of each beam) to the anatomy of the second CT scan.

By comparing the dosimetric results of the hybrid plan (no-replanning) with the adaptive replan for the ART group, no replanning showed decreased dose to the target volumes, with reductions to the dose received by 99% and 95% of the CTV, as well as the volume receiving at least 95% of the prescribed dose ($p=0.034$, $p=0.052$ and $p=0.047$ respectively). No replanning also demonstrated an increased dose to the normal structures, with increased maximum dose to the spinal cord ($p=0.002$) and brainstem ($p=0.007$), as well as increasing in all dosimetric endpoints for the parotid glands [52].

As for the clinical impact of ART, the 3-year local relapse free survival for the ART group was of 72.71%, and the median local relapse free survival was of 50 months, while for the non-ART group the values were 68.16% and 48 months respectively. Despite these discrepancies, the survival differences were not statistically significant between the 2 groups ($p=0.34$), but when looking at patients with tumour staging higher than T3[1] there was a significant improvement of the 3-year local relapse free survival for the ART group ($p=0.03$). Regarding early and late side effects, no significant differences were found between the two groups, despite a slight improvement in the severity of mucosa injury and xerostomia for the replanning patients ($p=0.05$ and $p=0.04$ respectively) [52].

A more recent study by Surucu et al. (2017) [53], included 34 patients which had a second CT scan and a replan during treatment and compared the dose achieved by the replan and with the original plan calculated in the second CT. The authors reported a median reduction of 4.5%, 3.0%, 6.2%, and 2.5% for the maximum dose to the spinal cord, brainstem, mean ipsilateral, and contralateral parotid respectively in the ART scenario.

[1] TNM system is the most widely used cancer staging system. The T stage refers to the size of the main tumour, ranging from 0 (cannot be found) to 4 (maximum stage), the N stage refers to the number and location of the lymph nodes and the M stage to whether or not the tumour is metastasized [107].

Zhang et al. (2016) [54] studied several different ART implementation alternatives, by varying the number and time point of replans, for 13 oropharyngeal cancer patients, in order to find the strategy that best optimized parotid sparing. Each of the patients had one planning CT and then 6 weekly CT scans during a 7-week IMRT treatment. For each of the weekly scans, a new plan was produced by the same physicist that conducted the original plan. A total of 63 ART alternatives were then simulated, by considering all combinations of number of replannings and their timings, and weekly non-ART distributions were estimated by calculating the original plan on each of the weekly CTs.

The authors found that the maximum benefit, considering the parotid glands mean dose, was found when 6 replannings (1 for each week of treatment) took place, corresponding to a mean and maximum benefit of 1.6 Gy and 6.4 Gy, respectively, when compared with the planned dose to the parotids, and of 3.3 Gy and 10.8 Gy when compared with the estimated delivered dose without ART. It was also reported that 94% of this benefit could be achieved with only 3 replans, at weeks 1, 2 and 5, indicating that replanning is more impactful in the first half of the treatment [54]. For this 3-replan scheme, 54% of the patients had a benefit superior to 4 Gy in at least one parotid, and 15% in both parotids. Regarding target coverage, 73% of non-ART fractions led to an underdosage of CTV, and increasing the number of replans was effective in improving CTV coverage [54].

In a study regarding the improvement in patient's quality of life with ART, Yang et al. (2013) followed a group of 129 patients, from which 43 did not undergo replanning during IMRT treatment and 86 had 1 mid-treatment replan [55]. By comparing global quality of life and other quality of life scales, the authors concluded that replanning had a significant impact for nasopharyngeal carcinoma patients. Additionally, the clinical outcome comparison indicated that ART significantly improved the 2-year local regional control (97.2% vs 92.4%, respectively) but did not improve the 2-year overall survival (89.8% vs 82.2%, respectively).

3.3. Predicting the need for ART

Although several studies have successfully established both the clinical and dosimetric benefits of ART application, the implementation of ART into the clinical practice is still a rather labour-intensive and time-consuming process, making it impossible for most clinics to implement it for every patient. Furthermore, evidence shows that ART is not equally beneficial for all patients, potentially being an unnecessary effort for some [56].

The percentage of patients that would benefit from ART has been reported to range from 21% to 66% [57,58], which means that electively scheduling adaptive replanning when initiating treatment would be beneficial in order to optimize clinical resources and the gain for the patient.

In a study by Capelle et al. (2012) [59], 20 randomly selected H&N patients treated with helical tomotherapy received a second CT scan at fraction 15, with a new plan being implemented from fraction 20 onwards [59]. This group of patients included both adjuvant (patients which had surgery prior to radiotherapy) and non-adjuvant (chemo)radiotherapy cases and, for each patient, the cumulative adapted and non-adapted dose distributions were calculated. The authors arrived at the conclusion that ART presented no benefit for adjuvantly treated patients, and that the benefit for the non-adjuvant group was minimal, being the patients with greater changes in weight loss the most benefited.

Patient selection for ART can be performed at two time periods: either prior to treatment, based on pre-treatment patient characteristics or imaging, or during treatment, based on geometric and/or dosimetric changes during treatment that can be imaging related factors (e.g. density changes) or non-imaging related factors such as weight loss [48].

It could be argued that the patients that would benefit the most from ART are the ones that will suffer greater anatomic changes in the course of treatment, and in this sense several studies have tried to identify pre-treatment factors that correlate with more severe anatomic modifications of both target volumes and OAR.

In a 2015 review, Brouwer et al. [48] analysed several studies that reported on factors correlating significantly with anatomic and/or dosimetric alterations to several structures during treatment. The majority of the studies reported on changes to the parotid glands (18 studies) and the most commonly identified factors were weight loss, planned dose to the parotid glands and parotid glands' volume loss. As for the spinal cord and the brainstem, some factors correlating with significant increase in dose parameters were changes in GTV volume, weight loss and changes in neck diameter at the level of the thyroid notch [48].

The main problem with this type of studies is that the selection of predictive factors is based on correlation tests, which only take into account the linear relationship between individual variables and the end-point (which in this case is anatomical modifications), not-considering multi-variable or non-linear interactions that could potentially improve prediction. To this date, there are relatively few studies that develop multi-variable models with the objective of predicting the need for ART.

3.3.1. Definition of performance metrics

Before discussing the published literature on models predicting the need for ART, it is important to define the performance metrics used to evaluate the quality of those models.

The classification performance of any model (classifier) regarding some test data can be summarized in a 2-dimensional matrix called the confusion matrix. The number of rows and columns of this matrix correspond to the number of different possible classes in the classification problem. For the purpose of this work, the dimensions of the confusion matrix will be 2x2, as represented in yellow in figure 11, as only two classes are possible: positive, if the patient needs ART, or negative otherwise. One dimension of the confusion matrix represents the true classification of the objects and the other dimension represents the class that the classifier assigns them to [60].

Outcome of the classifier	**Ground truth**		
	Positive	Negative	**Row Total**
Positive	**TP**	**FP**	TP+FP (total nº of objects with positive classification by the model)
Negative	**FN**	**TN**	FN+TN (total nº of objects with negative classification by the model)
Column Total	TP+FN (total nº of real positives in the testing sample)	FP+TN (total nº of real negatives in the testing sample)	N=TP+TN+FP+FN (total number of objects in the testing sample)

Figure 11. Example of a 2x2 confusion matrix (yellow).

As shown in figure 11, there are 4 basic cardinalities in the confusion matrix:

- True Positives (TP): the number of positive objects that are correctly classified by the model as being positive.
- False Positives (FP): the number of negative objects that are wrongfully classified by the model as being positive.
- False Negatives (FN): the number of positive objects that are wrongfully classified by the model as being negative.
- True Negatives (TN): the number of negative objects that are correctly classified by the model as being negative.

From these cardinalities, several metrics used to assess the performance of a given classifier can be derived. Some of the most commonly used ones are as follows:

- Accuracy, the proportion of true (both positive and negative) classifications calculated as:

$$\frac{TN+TP}{TN+TP+FN+FP} \quad (1)$$

- Specificity or true negative rate (TNR), the classifier's ability to correctly identify an object as negative, calculated as:

$$\frac{TN}{TN+FP} \quad (2)$$

- Sensitivity or true positive rate (TPR), the classifier's ability to correctly identify an object as positive, calculated as:

$$\frac{TP}{TP+FN} \quad (3)$$

- False positive rate (FPR), the probability that a positive result will be given by the classifier when the true value is negative, calculated as:

$$\frac{FP}{FP+TN} \quad (4)$$

- False negative rate (FNR), the probability that a negative result will be given by the classifier when the true value is positive, calculated as:

$$\frac{FN}{FN+TP} \quad (5)$$

- Positive predictive value (PPV), the proportion of positive results produced by the classifier that correspond to true positives:

$$\frac{TP}{TP+FP} \quad (6)$$

- Negative predictive value (NPV), the proportion of negative results produced by the classifier that correspond to true negatives:

$$\frac{TN}{TN+FN} \tag{7}$$

Typically, the output of a binary classifier is a certain score, which is a continuous value representing the probability that a given sample belongs to a specific class. In this way, the actual classification, and all the related performance metrics depend on a chosen threshold, which assigns the sample to one of the two possible classes according to whether their scores fall above or below the chosen threshold.

A good classifier should have both high values of sensitivity, being able to detect positive cases when they occur, and of specificity, having a low “false alarm” percentage by not identifying negative cases as positive.

The receiver operating characteristic (ROC) curve depicts the sensitivity versus the false positive rate which, as can be seen from equations 2 and 4 corresponds to (1 – specificity).

An example of the ROC curves for 2 models, A and B, are shown in figure 12. All ROC curves start at (0,0) and end at (1,1), with each point representing the balance between specificity and sensitivity for a different threshold. The dashed line in the image (diagonal) represents the uninformative model, where the true positive rate is equal to the false positive rate, both being 50% which is the same as randomly classifying the data.

Conversely, a model that perfectly classifies all data points for any thresholds has a ROC curve that is an horizontal line in sensitivity=1 [61]. In this way, the closer a model’s curve is to the diagonal line the worse its performance is, and the closer it is to the line representing the perfect classifier, the better it will classify the data.

In figure 12 we can see that, for a given classification task, model’s A ROC curve is closer to the dashed line, while model’s B curve is closer to the line of the perfect classifier, which means that model B is a better classifier.

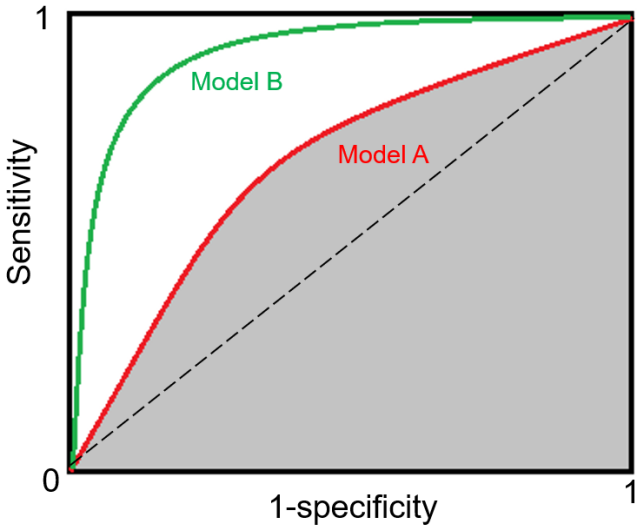

Figure 12. ROC curves for 2 models. The dashed line corresponds to random guessing. The grey area corresponds to the area under the curve for model A.

The area under the curve (AUC, depicted in grey for model A in figure 12), represents the probability that a model will correctly classify an object. A perfect model will have an AUC of 1 and for the random model the AUC is 0.5. This means that better models will have AUC values closer to 1, making the AUC a very popular metric to assess classifier's performances. Likewise, in figure 12 we can see that the AUC for model B is higher than for model A.

3.3.2. Training and testing the models

Regardless of the chosen method for classification, it is vital to keep in mind that the ultimate goal is not to produce a model which works perfectly for the available data, but one that will correctly classify new, previously unseen datapoints.

In this way, it is very important to use independent train and test cohorts for model building, and a separate validation cohort, preferably from an external environment, to assess the model performance, in order to produce robust and reliable predictive models [62–65].

Furthermore, the splitting of the data into training and testing sets should be repeated multiple times, in order to prevent the results from being biased by specific selected sets. One of the most popular methods to perform this repetition is cross-validation, which consists of dividing the data set into k subsets (folds) and then repeating the classification

process *k* times, each time using one different subset for testing and all the *k*-1 remaining ones for training [66].

A specific variation of cross-validation which is also very commonly used is leave-one-out cross-validation, in which the number *k* of folders is equal to the number of samples. In this case, in each repetition of the classification process, a different sample is used for testing, with all the remaining samples being used for training [67].

It is also very important that both classes are equally represented in the training and testing cohorts. If the data is highly imbalanced, with a significantly larger proportion of examples from one class, some of the metrics defined above can be misleading for the actual classification performance. For example, in a case where in the test set 90% of samples are from class A and 10% from class B, a model that classifies all samples as A would have 90% accuracy, despite misclassifying all examples from class B.

3.3.3. Predictive models for the need for ART in H&N cancer

Brown et al. (2015) [68] conducted a study with a sample of 110 H&N patients treated with definitive IMRT (with or without concurrent chemotherapy) in which all patients had a daily image prior to treatment and a repeated CT if significant differences (>1 cm at any point of the external contour) were detected between the daily image and the planning CT. The initial plan was then calculated in the repeat CT scan and, if the dosimetric discrepancies were perceived as detrimental by the radiation oncologist, a new plan was implemented. From the 110 patients, 21 had a repeated CT at some point during treatment, with 5 resulting in a new treatment plan.

The authors carried out univariate and multivariate analysis to compare various factors between the two groups, finding statistically significant differences in nodal disease staging (more advance for the replanned group), dominant nodal size (larger for the replanned group), and diagnosis, with the majority of replanned patients (3 out of 5) having a nasopharyngeal carcinoma. These variables plus a binary variable representing initial weight (<100 kg or >100 kg) were used to generate two logistic regression models to predict the need for ART: one not weighted (model 1) and another that was weighted for the proportion of replanned patients (model 2). The post estimation results for sensitivity and specificity were 60% and 100% respectively for model 1, and 100% and 92.31% for model 2 [68].

These predictive models were used to determine threshold values that would include a patient in an ART risk profile, using a process that could be implemented clinically. The high-risk class was defined as including patients with a probability of requiring a replan

greater than 80%. The intermediate risk class considered a probability greater than 60% [68].

This study presents some limitations, namely the imbalance of the data, since only 4.5% of patients had a replan, and the fact that the model is trained and tested on the same population, which could severely undermine its generalizability.

Brouwer et al. (2016) [69] developed a linear regression model to identify patients for ART based on the parotid glands dose deviation, using a training cohort of 113 patients. The endpoint for the linear regression was defined as the difference between the mean dose to the parotids in the planning CT and that at a post-RT CT, acquired after the completion of treatment (ΔDmean).

The authors performed both univariate and multivariate linear regression analysis between this endpoint and several pre-treatment factors. In the univariate analysis, multiple factors significantly correlated with the endpoint, namely initial body-mass index, chemotherapy, T-stage, N-stage, planned mean dose to the parotid glands, initial GTV volume, tumour location as well as overlap volume of the parotids with the target. However, from the multivariate analysis, the planned mean dose to the parotids was the only significant factor, which was then applied to the data to select patients for ART. In this study, it was defined that any patient with an absolute ΔDmean > 3 Gy would be referred for ART. The proposed model established a threshold of 22.2 Gy for the mean dose to the parotid glands, meaning that patients who exceeded this value would be more likely to need ART [69].

When validating this model in an independent patient cohort of 43 patients, the authors obtained a sensitivity of 80% but a positive predictive value of only 19%, meaning that a large percentage of patients that did not need ART would be selected for it. In fact, only for 18% patients of the test cohort the ΔDmean was higher than 3 Gy but the model selected 76% of the patients for ART [69].

Decision trees have also been used with the intent of identifying patients suitable for ART. In a study by Surucu et al. (2016) [70], two decision trees were developed to classify patients into high or low expected percentage of GTV volume change during therapy: one for primary and another for nodal tumours. Forty eight patients were included in the study and the whole dataset was used to develop the decision trees, both of which achieved an accuracy of 88%, but no test set was used to validate this result. Of the 14 predictive factors considered, the ones that were most discriminating between the two groups for both primary and nodal volume were: type of chemotherapy, age,

tumour appearance, site, Karnofsky performance status[2], HPV status, primary tumour growth pattern and both primary and nodal initial tumour volume.

In another study, Castelli et al. (2016) [71] developed a nomogram to predict parotid gland overdose from early treatment patient factors. Twenty locally advanced oropharyngeal cancer patients were included in the study and each of the patients had a weekly CT scan, in which the actual dose delivered weekly was estimated. These weekly dose estimations were then propagated to the planning CT using rigid and deformable image registration and then used to compute the cumulated dose on the planning CT, which was finally compared to the planned dose. Parotid gland overdose was calculated as the difference between the cumulated mean dose and the mean dose on the planning CT. Parameters from the first weekly CT were found to be more strongly correlated with parotid overdose and three factors were included in the model: the difference between the mean dose to the parotids from the planning CT to the first week CT, the difference between the CTV receiving 70 Gy in the planning CT and in the first week CT and finally the initial CTV volume percentage receiving 70 Gy [71].

Using leave-one-out cross validation on the 20-patient cohort, they reported a sensitivity of 80% and a specificity of 60% when predicting whether the cumulative parotid dose would increase or decrease [71].

3.4. Radiomics Approach

Two recent studies [72,73] employed a different, image-based, technique to predict the need for ART from pre-treatment factors: radiomics.

Radiomics is a relatively recent concept in the fields of computational oncology and image processing, and can be essentially defined as the high-throughput extraction of quantitative biomarkers (features) from medical images, which can then be applied in mathematical models for several purposes, such as aid in diagnosis or prediction of outcome [62,65,74].

The recent advances in medical imaging devices, imaging agents, image processing and analysis combined with the standardization of imaging protocols across different practices, made it possible to convert the imaging data, that so far had been used only

[2] According to the national cancer institute dictionary of cancer terms, the Karnofsky performance status is a standard way of measuring the ability of cancer patients to perform ordinary tasks. Its values range from 0 to 100, with a higher score meaning the patient is better able to carry out daily activities. It may be used to determine a patient's prognosis, to measure changes in a patient's ability to function, or to decide if a patient could be included in a clinical trial [108].

for qualitative purposes by qualified experts, into quantitative information [74], which can be mined resorting to datamining and machine learning techniques with the purpose of improving decision support [62].

This conversion of digital medical images into mineable high-dimensional data is motivated by the concept that biomedical images contain information that reflect the underlying pathophysiology, and which are not detectable by the human eye. Consequently, these “invisible” relationships could be revealed via quantitative image features based on intensity, shape, size or volume, and texture [62].

The concept of radiomics has gained a lot of popularity in the recent years due to its potential towards individualized medicine without any need for invasive or extra interventions on the patient’s typical workflow in the clinic, since imaging is already a major part of any treatment process. For this reason also, large amounts of data are available for radiomics studies, since every patient has one or more sets of images and each image contains millions of voxels and hundreds of features that can be potentially explored in a radiomics approach [65].

3.4.1. Radiomics Workflow

After the identification of the aim of the radiomics analysis and of a suitable patient cohort in which to perform the study, the general radiomics workflow can be broken down into 6 steps, which are represented in figure 13. This workflow can be applied to any image modality. Since this work is based on CT scans the specific steps will be discussed in the realm of this modality only.

3.4.1.1. Image Acquisition

The process begins with the acquisition of high-quality medical images, which could be CT, PET-CT or MRI scans. The modern technologies in all of these modalities allow for a wide range of image acquisition parameters and reconstruction protocols, which can have an impact on the value of the extracted features [63–65,72,74,75].

Using a standardized acquisition protocol across the different patients in the study is a way to eliminate this bias, but which can then lead to poor generalizability compared to cohorts in which the images are acquired with different parameters, with some authors defending the use of heterogeneous protocols in the training cohorts [64].

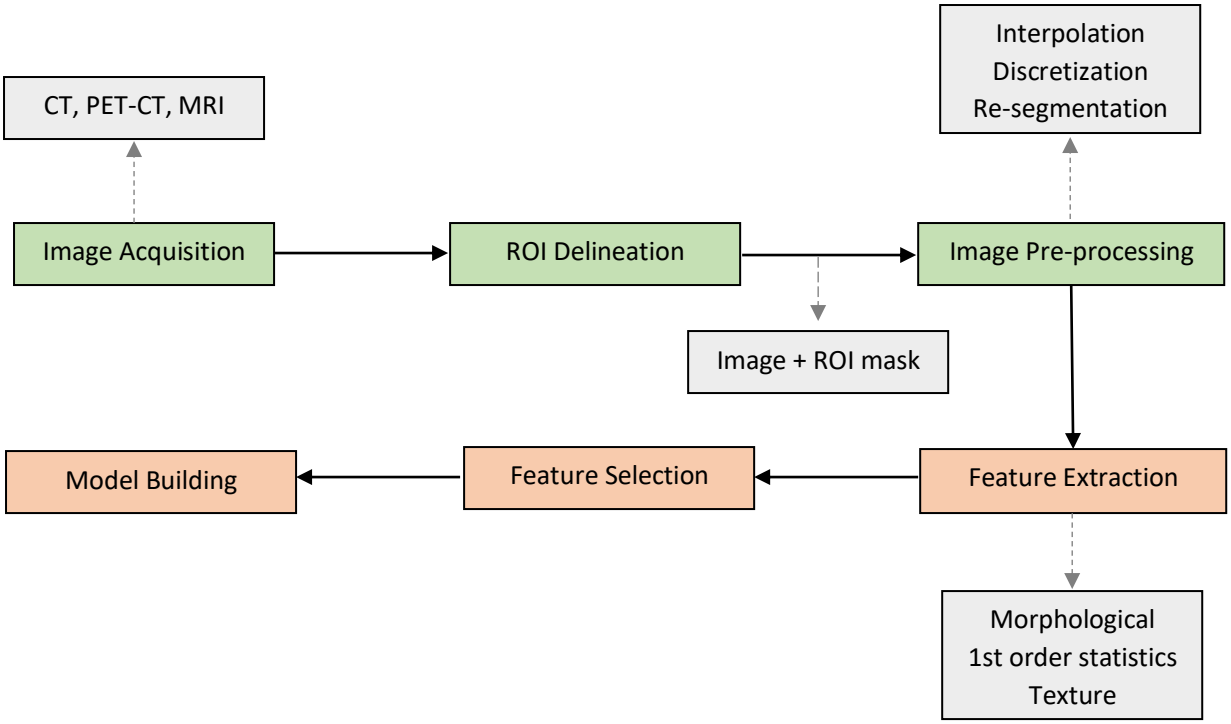

Figure 13. General Radiomic workflow. The boxes in green represent steps in which the actual images and/or masks of the ROI are considered. The orange boxes represent the steps on which vectors of the quantitative features extracted from the images are considered.

3.4.1.2. Region of Interest (ROI) Delineation

The delineation of the region of interest (typically the primary tumour (GTV) and/or nodal tumours) from which the radiomic features are to be extracted can be performed manually by a radiation oncologist, or automatic/semi-automatically by a software (such as Mirada RTx 1.6 and Workflow Box 1.4 by Mirada Medical Ltd., United Kingdom).

This step has been described as the most challenging of all the radiomics process [62], since in most cases tumours do not have distinct borders, making manual segmentation still an expert-dependent process.

Many studies have shown that both the inter and intra-operator variability of manual contouring of tumours is very high, and it can be reduced with the aid of automatic/semi-automatic tools, or multi-operator comparisons [76–78].

3.4.1.3. Image Pre-processing

Several image pre-processing steps must be followed before the extraction of the radiomic features. These steps may vary from study to study, according to the specificities of each dataset or the desired endpoint. In this work, the guidelines provided by the Imaging Biomarker Standardization Initiative (IBSI) [79] were followed and the general pre-processing steps are hereby succinctly presented.

Right after image acquisition, it can be necessary to perform some post-acquisition processing, in order to enhance the quality of the image and to do some de-nosing. For instance, metal objects such as pacemakers and tooth implants introduce artefacts in CT images and may require artefact suppression.

From the delineation of the ROI, a binary ROI mask is generated by assigning the value of 1 to all voxels inside the contour and 0 to all the other voxels in the image, which will then be used for feature extraction.

The image's voxel spacing has been shown to have a significant impact in texture features, so interpolation to an isotropic voxel spacing is an important step of image pre-processing, allowing for comparison between image data from different samples or cohorts [64]. The binary ROI mask must also be interpolated to the same voxel spacing as the image.

Rounding to the nearest integer value is usually done following interpolation, as some image voxels might be left with non-integer values, which in the case of CT scans is not possible since they represent Hounsfield Units (which are always integer).

Additionally, range re-segmentation may be performed to remove voxels that fall outside a specified range and are not useful to the desired analysis, such as voxels with Hounsfield Units indication air or bone in CT images, since they do not provide any information about the ROI.

Finally, it is necessary to discretize the range of intensity values inside the ROI in order to efficiently compute several texture features [80], which can be done either with a fixed number of bins or with a fixed bin width approach.

3.4.1.4. Feature Extraction

After all the previous steps have been executed, the image features, which consist of quantitative measures in the context of radiomics, can be extracted from the ROI in the image.

These features can be grouped in several families according to their characteristics. Several authors refer to first, second, or higher order statistical outputs with first order features describing the distribution of values of individual voxels without any concern for the spatial interactions between them, second order describing statistical interrelationships between voxels with similar (or dissimilar) contrast values and higher-order features imposing filter grids on the image to extract patterns [62,65,81].

In this work, the feature definition and grouping adopted by the Image Biomarker Standardization Initiative [79] will be used. The IBSI divides image features into 6 families:

a) **Morphological features**: which describe geometric aspects of a ROI, such as area and volume, as well as shape characteristics, such as sphericity, elongation, flatness, etc.
b) **Local intensity features**: which use voxel intensities within a defined neighbourhood around a centre voxel to compute local and global intensity peaks.
c) **Intensity-based statistical features**: which describe how intensities within a ROI are distributed, without requiring discretization, so that they can be used to describe a continuous intensity distribution. These features are only meaningful if the intensity values scale is not arbitrary.
d) **Intensity histogram features**: which are obtained by discretizing the intensity values in the image to a fixed number of bins and generating a frequency histogram (intensity histogram) from which several statistics can be calculated.
e) **Intensity volume histogram features**: which describe the relationship between each discretized intensity *i*, and the fraction of the volume that contains at least intensity *i*.
f) **Texture matrix-based features**: which group the information of the voxels in matrixes, based on different types of relationships between them, and then extract different kinds of values from those matrixes:

- Grey level co-occurrence matrix: which describes how combinations of discretized intensities of neighbouring pixels, or voxels in a 3D volume, are distributed along one of the image directions (figure 14 shows an example for a 2D matrix of grey levels).
- Grey level run length matrix: which assesses run lengths, the length of a consecutive sequence of pixels or voxels with the same grey level along direction.

- Grey level size zone matrix: which counts the number of groups (or zones) of linked voxels (voxels with identical discretized grey level).
- Grey level distance zone matrix: which counts the number of groups (or zones) of linked voxels which are at the same distance to ROI edge, capturing the relationship between position and grey level.
- Neighbourhood grey tone difference matrix: as defined by Amadasun and King [82].
- Neighbouring grey level dependence matrix: as defined by Sun and Wee [83].

(a)

0	0	0	0	0
4	1	6	4	2
3	1	3	2	7
5	1	2	7	7
4	2	7	7	7

(b)

	0	1	2	3	4	5	6	7
0	6	0	0	0	0	0	0	0
1	0	0	1	0	1	0	0	1
2	0	1	0	0	0	1	1	2
3	0	0	0	2	0	0	0	1
4	0	1	0	0	0	0	1	1
5	0	0	1	0	0	0	0	0
6	0	0	1	0	1	0	0	0
7	0	1	2	1	1	0	0	2

Figure 14. a) Digital Image; b) corresponding grey level co-occurrence matrices for the horizontal direction (Adapted from Castellano et al. (2004) [84]).

3.4.1.5. Feature Selection

The features mentioned in the previous section can amount to hundreds per image, and this number can grow even further if filters, which can be used to enhance certain image characteristic as edges or blobs, are applied prior to extraction. This massive number of features is in staggering contrast with the reduced sample sizes that are available in most studies, which are often well below 100 patients. Therefore, it is unthinkable to use all extracted features in model building, as several of these extracted features may not be useful for the particular goal of the study and using all of them would inevitably result in reduced generalization power (the so called overfitting of the data) [65].

In this way, effective feature selection or reduction techniques are imperative for the development of adequate and useful radiomics models. Several such techniques exist, and the choice of which to use highly depends on the task-specific goal, the size of the sample and the computational resources available.

Typically, feature selection methods can be divided into 3 groups: filter methods, wrapper methods and embedded methods. Filter methods are classifier independent and essentially rank the features according to a scoring criterion, which can depend on each given feature relevancy towards the outcome (univariate methods), or can take into account the interactions between different features, being sensitive to feature redundancy as well as relevancy (multivariate methods). This type of methods are very computationally efficient, which makes them very appealing for radiomics studies due to the large quantity of imaging features that need to be taken into account [85].

Wrapper methods, on the other hand, are a classifier dependent approach, which resembles a search method, scanning the whole feature space to define a set of relevant and non-redundant features for a given classifier, being the training/validation accuracy of the classifier the measure of utility for the candidate feature subset.

These methods can often produce better classification results when compared to filter methods, but they can also become very computationally expensive for a high-dimensional feature space and produce feature subsets that are somewhat specific to the chosen classifier [85].

At last, embedded methods are also classifier dependent methods which incorporate the feature selection in the training process. The most common example of this method is the decision tree algorithm, which selects a feature in each recursive step of the tree growth process, according to its ability to correctly divide the data.

These methods are more computationally efficient when compared to wrapper methods, but have also the disadvantage of producing classifier-specific feature subsets [85].

In addition to these feature selection techniques, dimensionality can be reduced by combining and/or transforming the original feature set into a transformed smaller set, using methods like principal component analysis or linear discriminant analysis. These methods are computationally very efficient, but have as disadvantage the total loss of physical meaning of the final set of features [65].

3.4.1.6. Model building

After a suitable set of features is identified a predictive model can be built. This can be done resorting to both supervised and unsupervised approaches. In unsupervised techniques, such as k-nearest neighbour clustering, the classifier does not know the true class of the training samples, but aims to discover structure within the data itself,

grouping similar samples (according to some similarity metric which can be defined by the user) into clusters of data.

Conversely, in supervised methods the classifier uses information about the true class of the training samples to produce an optimised mapping between the input features and the expected output classification of a new sample. Some examples of supervised techniques are neural networks, support vector machines, bayesian classifiers, random forests, among others. [65,85,86]. Here a brief description of the models used in the second part of this work, namely support vector machines, Bayesian classifier, decision trees and multi adaptive regression splines are presented.

Support vector machines (SVM) were first introduced by Dr. Vladmir Vapnik in 1982-86 as a pattern recognition method where input vectors are non-linearly mapped to a very high-dimension feature space in which a linear decision surface can be constructed [87].

The main idea behind SVM is the construction of a separating hyperplane (discriminative function) that optimally separates the boundary objects of each class, which are called support vectors. This is done by finding the hyperplane that maximizes the distance to these support vectors, which is called the margin of separation [61]. Figure 15 shows an example of the support vectors, optimal margin and optimal separation hyper-plane for a separable 2-dimensional case.

It is shown [87] that this problem can be solved by finding the weights α_i and bias b for which:

$$c = \sum_i \alpha_i k(s_i, x) + b \tag{8}$$

Being s_i the support vectors, $k(s_i, x)$ a kernel function (which can be linear, polynomial, gaussian, etc) x a given object and c its output classification.

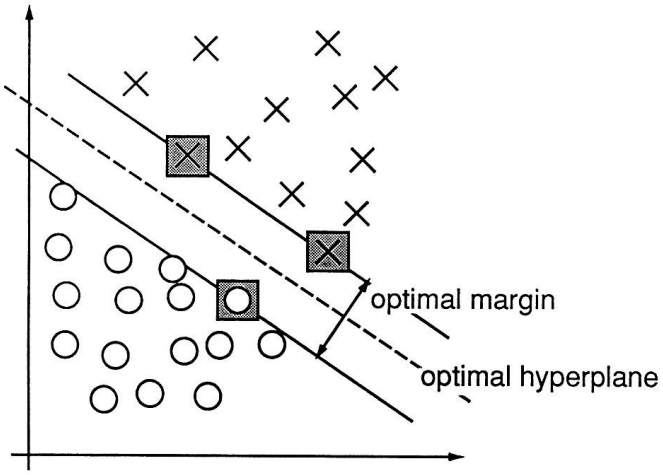

Figure 15. An example of a separable problem in a 2-dimensional space. The support vectors, marked with grey squares, define the margin of largest separation between the two classes [83].

Bayesian learning algorithms differ from SVM in the sense that they provide a statistical approach to classification, rooted on the assumption that the quantities of interest follow certain probability distributions which can be used to produce optimal decisions based on the observed data [88].

Another very popular classification approach are decision trees. This type of classifiers classifies the examples by sorting them down the tree starting from the root and ending on a leaf node, which provides the classification. Each node of the tree refers to a feature of the example, and each branch descending from a node corresponds to the several possible values that the feature can take. An example of a decision tree is shown in figure 16. Each example is classified by going through the decision tree, starting at the root node, testing the specified feature at each node and proceeding down the corresponding branch, until a leaf node with the classification is reached [88].

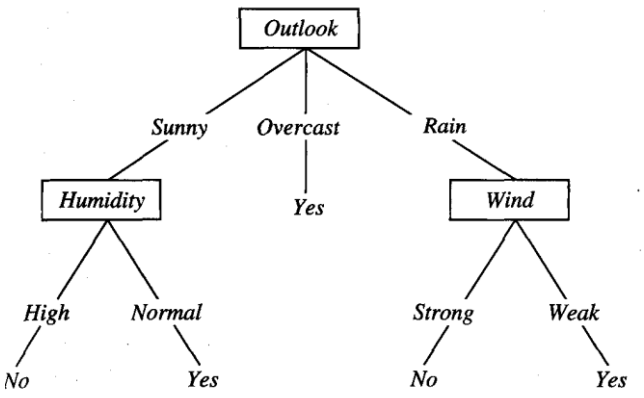

Figure 16. Example of decision tree. This tree classifies Saturday mornings according to whether or not they are suitable for playing tennis ([88] page 53).

Another classification method is Multivariate Regression Splines (MARS). This is a nonlinear and nonparametric regression method, in which the features in the training data set are divided into separate piecewise linear segments (splines) of differing gradients (slopes). Essentially, the predictors are broken into groups according to a certain cut point, and the linear relationships between the predictors and the outcome are modelled within each group. This break originates two new features, one which has values of zero greater than the cut point, while the second feature is zero less than the cut point. The new features are then added to a basic linear regression model, creating a piecewise linear model where each new feature models an isolated portion of the original data [89].

3.4.1.7. Radiomics applications – ART

In the past decade, there have been many applications of radiomics and texture analysis techniques in oncology, such as in tumour segmentation and pathologic classification, risk stratification and monitoring of alterations in normal tissue as a result of radiotherapy [90]. Recently, two studies were published exploring the capacity of using radiomics to predict the need for adaptive radiotherapy.

The first study, by Ramella et all. [72], was published in November 2018 and investigated the feasibility of using radiomic features of the patient's initial imaging to predict tumour reduction during chemoradiotherapy for non-small cell lung cancer (NSCLC) patients. The authors used a cohort of 91 NSCLC patients treated with concurrent chemoradiotherapy, of which 50 had an ART plan. Using an in-house developed software based on MATLAB, 12 statistics and 252 texture features were extracted from the Clinical Volume, which was manually delineated in the planning CT by a radiation oncologist.

In addition to the radiomics features, nine semantic features, namely patients' age, sex, smoking attitude, tumour staging scores and histology/gene mutation information, were included in the analysis.

Feature selection was done resorting to a wrapper method with random forests. A leave-one-out cross-validation loop was used to generate several feature subsets selected by the wrapper approach, and the frequency of selection of each feature was assessed. All the features that were selected in at least 10% of loop interactions were included in the model, which resulted in 12 features: 5 semantic (sex, N stage, histology, epidermal

growth factor mutation and smoking attitude), 2 GLCM, 4 Local Binary Patterns on three orthogonal planes (LBP-TOP)[3] features and 1 statistical feature.

These features were then used in a random forest method to classify the data into ART and non-ART groups, using LOO cross-validation to assess performance. The authors reported AUC of 82% (95% CI of 73%-91%), an accuracy of 78% (95% CI of 69.5%-86.5%), a precision of 77.8% (95% CI 69.95%-86%), a sensitivity of 84% (95% CI of 75.7%-92.2%), and a PPV and NPV of 65.7% (95% CI of 60.7%-70.7%) and 86.9% (95% CI 83.4%-90.4%) respectively. Furthermore, it was reported that the scores obtained using the semantic features alone were considerably lower than with the radiomic signature, showing a decrease of 4% in AUC and 6% in accuracy.

The second study, by Yu et al. [73], was published in October 2019 and had the objective of predicting ART eligibility for advanced nasopharyngeal carcinoma patients (NPC), using radiomic features extracted from multi-parametric magnetic resonance images. The study cohort included 70 NPC patients treated with radical radiotherapy from which 13 had a replan sometime during treatment. Pre-treatment contrast enhanced T1 weighted (CET1-w) and T2 weighted (T2-w) MR images were acquired for each patient, and for each image 490 radiomic features were extracted, including 14 shape, 90 1st order intensity and 375 texture features.

In order to reduce the number of features, the authors started by removing highly inter-correlated features based on the computation of Pearson correlation coefficient between pairs of features, reducing the number of features to 53 out of the original 490.

After this preliminary reduction, the authors applied the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm to both further perform classification and reduce the number of features. For each of three cases: 1) CET1-w image features, 2) T2-w image features, and 3) combination of features extracted from both CET1-w and T2-w images, a three step process was defined, during which several models with different ratios between planned and replanned patients were used for training. Eight sets of radiomic features with a number of variables ranging from 3 to 10 were identified and analysed for prediction capability.

For each set of features, 100 resampled iterations of 20-repeated 3-fold cross validation were performed for cases 1, 2 and 3, and the distribution of AUC values was computed.

The authors identified 8 features for case 1 (CET1-w model), including 2 shape features and 6 Laplacian of Gaussian (LoG) based features, which yielded an average AUC value

[3] LBP-TOP are descriptors which assign to each pixel of the image a label comparing it with its neighborhood matrix computed from three orthogonal planes (top) [72].

of 85.2% (95% CI of 84.7%-85.7%) in the testing sets. For case 2, 6 features were selected, including 2 shape and 4 LoG-based features, with an average AUC value in the testing sets of 75% (95%CI of 74.5%-75.5%). Finally, the best result was achieved for case 3, where 6 features were also identified, including one 1st order feature (kurtosis) and 5 LoG-based features, resulting in an average AUC value in the testing sets of 93% (95%CI of 92.8%-93.3%). No semantic features were included in the predictive models for this study.

Both studies show that there is potential for the application of radiomics in the prediction of ART.

3.4.1.8. Main challenges in Radiomics

Despite the many areas where radiomics can be potentially applied, there are still some factors that are keeping it from being routinely implemented in clinical practice.

For instance, several radiomic features have shown to be sensitive to a number of technical factors, such as image acquisition parameters and different reconstruction algorithms, which contributes to low reproducibility of the radiomic features across different machines or centres [80].

Furthermore, different pre-processing techniques, may also affect the value of extracted features and, even for the same set of features, different feature-selection and classification methods can yield highly different results [63].

Another limitation of the radiomics process is that the extracted features are highly dependent on the delineated region of interest, but manual delineation of tumours is both time-consuming and prone to intra-observer variability. It is thus recommended that semi-automatic delineation tools are used for radiomic studies, but these tools are currently not available for all clinics [80].

These issues make it very important to include robustness analysis in radiomics studies, in order to assess whether the extracted features are not significantly influenced by variations in patient positioning, image acquisition and segmentation. This analysis can be done by test-retest imaging, which means imaging the patient twice at different timepoints (minutes or days of difference), usually with the same acquisition parameters. In this way, two slightly different images of the ROI are acquired and the level of resemblance of the extracted features between both can be assessed [91].

Unfortunately, test-retest protocols demand the utilization of additional material and human resources, as well as produce additional radiation to the patient, which makes it

a difficult procedure to routinely implement for every radiomic study. Furthermore, since the robustness of the features highly depends on the imaging modality, the type of cancer, voxel size, type of discretization, etc., the results cannot be easily transferred from one study to another [91].

To overcome these difficulties, Zwanenburg et al. (2019) propose a different approach inspired by the deep learning computer vision field in which feature robustness is assessed through applying a series of perturbations to the images, which simulate differences in patient positioning, contour and levels of noise in the image [91]. In this way, by extracting the features from all perturbed images and calculating the intraclass correlation coefficient for each feature, considering all patients in the study, a good idea of the feature robustness can be obtained.

As mentioned before, most radiomics studies have relatively small patient cohorts when compared to the enormous amount of extracted radiomic features. This imbalance results in a high-dimensional feature space with very sparse data points, which can easily lead to an overfitting of the data and large false positive rates, making it imperative that reliable feature selection methods are applied prior to model building [63].

All these aspects make it more difficult to produce stable models that have good generalizability and can be consistently reproduced across multiple cohorts from different institutions, which is necessary for the implementation of these models into clinical practice. For this reason, there has been an increasing number of initiatives for the standardization of the different steps of the radiomics pipeline, such as the Image Biomarker Standardization Initiative [79], in order to improve reproducibility of radiomic studies.

Chapter 4

Assessing the need for Adaptive Radiotherapy in H&N patients

4.1. Purpose

The purpose of this study is to assess the need for ART in H&N patients by evaluating the dosimetric impact of performing a single replan in the course of treatment both on the target volumes and on seventeen OARs, using an automatic planning tool. The use of an automatic planning tool allows an unbiased analysis of the impact of ART considering the inter/intra user variability that is inevitable when the second plan is produced under different circumstances and/or by a different planner.

4.2. Material and Methods

4.2.1. Sample description

The study sample consisted of 30 H&N patients treated with helical IMRT at IPOCFG. The treatment was delivered by a Tomotherapy HD (Accuray) unit, with (22 patients) or without (8 patients) concomitant chemotherapy.

The vast majority (25 out of 30) of patients were prescribed 69.96 Gy to the tumour planning target volume (PTV-T), with two patients being prescribed 59.4 Gy and three being prescribed 50 Gy. The prescription to the lymphatic nodes' PTVs (PTV-N) was 50.4 Gy, 54 Gy or 59.4 Gy. There were 3 patients which were not prescribed dose to the lymph nodes. Depending on the prescription scheme, the treatment was delivered in either 33 (26 patients), 28 (1 patient) or 20 (3 patients) fractions.

All patients had a planning CT (p-CT) scan in treatment position acquired at a median of 14.5 days before the start of treatment, on which the target volumes were manually delineated by a radiation oncologist. This CT scan was acquired using the Somatom Sensation Open scanner from Siemens, with a 3 mm slice thickness.

A summary of the patients' characteristics can be seen in table 1.

Table 1. Summary of patients' characteristics. The mean and standard deviation is shown for the age and initial weight.

Characteristic	Value
Age	61.4 ± 10.3
Gender (M/F)	27/3
T stage (1/2/3/4)	1/1/6/22
N stage (0/1/2/3/)	4/1/20/5
Initial Weight (Kg)	62.0 ± 15.1

4.2.2. Adaptive Radiotherapy scheme

All patients underwent a daily MV-CT scan, before each treatment fraction, to assess for differences between the current position and the one in the p-CT. A rigid registration was then performed between the MV-CT and the p-CT, and a transformation vector calculated, which translated into rigid couch and roll angle shifts to correct for positional errors.

In the case where even after the rigid transformation significant discrepancies could be observed, the medical physicist was alerted, and a dose calculation was performed on the daily scan using the Planned Adaptive software module version 5.1.0.6 from Accuray. This dose distribution was then revised by the responsible radiation oncologist, which determined the need for a new treatment plan. If a replan was required, a second CT (verification CT) was acquired, and both the target volumes and OAR propagated from the p-CT to the verification CT by rigid and deformable image registration using Velocity AI, version 3.2. These volumes were then manually corrected by the radiation oncologist and a new plan was designed by the responsible medical physicist.

4.2.3. Plan generation

An automatic planning tool, *iCycle* [92], was used to generate three dose distributions for each patient in the cohort: 1) the dose distribution corresponding to the optimized plan for the planning CT (original plan); 2) the dose distribution corresponding to the optimized plan for the verification CT (ART-plan); 3) the dose distribution obtained with plan 1 but considering the verification CT (non-ART plan).

All plans were generated by the Erasmus-iCycle IMRT multicriterial optimization engine [92]. In this tool, plan generation is guided by a user defined wish-list, which contains both clinical constraints that must be strictly met, and prioritized objectives to be optimized. The optimization occurs through a constraint-based method, 2pεc method, which generates a single Pareto optimal IMRT solution for a given set of beams. The wish-list were built according to the guidelines established by Ventura, et al. (2018) [93].

For the calculation of the non-ART plan, the isocentres of the two CTs (planning and verification CT) were aligned and the dose distribution obtained by the fluence and beam arrangement resulting from the original plan was computed on the verification CT.

4.2.4. Dosimetric analysis

The dosimetric endpoints for both target volumes and OARs were compared between the adaptive and the non-adaptive plan, to assess the impact of ART. The near minimum dose (D98%) was considered for the target volumes and the maximum and mean doses for OAR with serial and parallel architectures, respectively.

First the normality and homogeneity of the variance were assessed using the Shapiro-Wilk normality test and Levene's test respectively, in order to determine whether parametric or non-parametric tests were suitable. Then, since all variables were non-parametric the Wilcoxon sign ranked test, with a significance level of 0.05, was used to assess statistical significance.

4.2.5. Plans assessment and comparison

The dose distributions obtained for the ART-plan and the non-ART plan were then assessed and compared using the SPIDERplan quality assessment tool. This tool, described in detail by Ventura et al. (2018) [94], scores each structure of interest using a score function based on targets/OAR clinical constraints. These scores are inversely proportional to the quality of the plan regarding that structure, meaning that the closer the score is to zero the better the sparing of the OAR or coverage of the target volume. A score lower than unity is achieved when all planning aims are met, and a score higher than 1 means that the plan is not complying with the required dose constraints. The structures are organized within groups and weights, which reflect the radiation oncologist clinical preferences, are assigned to both groups and structures. The scores associated with each structure are then combined so that group scores and a global plan score

reflecting the clinical preferences are calculated. The results are presented in an intuitive graphical representation through customized radar plots.

In this study, the structures were divided into five groups (PTV group, Critical group, DigestOral group, Bone group, Optics group and Other group) which were assigned different weights (50%, 30%, 5%, 3.5%, 10% and 1.5%, respectively). The different group structures depend on factors like the pathology and clinical preferences, established based on institutional or international protocols, and the different weights were assigned according to the group clinical relevance [94].

4.3. Results

The results of the dosimetric and the SPIDERplan quality analysis for every target volume and each OAR are shown in table 2. The structures are organized in the table by the respective SPIDERplan groups, starting with the PTV group, and moving sequentially to the Critical, DigestOral, Bone, Optics and Other groups. The dosimetric analysis regarding the specified parameter (D98%, Dmax or Dmean) is presented first, followed by the group score for the respective group. Finally, the mean global plan scores for the ART and non-ART plans are also presented. The p-values lower than 0.05, representing statistical significance, are in bold and marked with a star.

Figure 17 shows the SPIDERplan plot for a representative patient. The red line represents the structures' scores for the ART plan and the yellow line the scores for the non-ART plan. It is clear that the scores for the ART plan are generally closer to the centre of the plot, meaning that they are closer to the aimed value. The largest differences are observed for the target volumes (PTV group), and for the DigestOral group. Furthermore, for this specific patient, a significant improvement is achieved in the spinal cord with the ART plan clearly complying with the dose constraint for this OAR (score below 1) which is not the case for the non-ART plan (score above the radius 1 circle in the diagram). In fact, the scores in the ART plan are all bellow 1 (inner circle) with the exception of the thyroid, while the majority of the structures in the non-ART plan presents a score exciding 1, meaning that the required clinical constraints for plan acceptance are not achieved in this scenario.

Table 2. Results of the dosimetric and SPIDERplan analysis of the ART and non-ART plans.

Parameter	Structure	ART		Non-ART		p-value
		Mean	std	Mean	Std	
D98% (Gy)	PTV69-T	67.3	0.729	58.00	8.323	**0.000***
	PTV59-N1	58.31	1.053	44.46	13.20	**0.000***
	PTV54-N2	52.88	0.385	40.60	10.39	**0.026***
	PTV59-N2	57.34	0.68	49.73	7.686	**0.002***
	PTV59-N	58.52	1.198	32.83	18.27	**0.000***
Group Score	PTV group	0.977	0.010	2.447	5.975	**0.000**
Dmax (Gy)	Spinal Cord	35.58	10.57	41.52	13.45	**0.017**
	Brainstem	15.36	13.74	19.19	17.07	0.474
Group Score	Critical group	0.537	0.195	0.645	0.254	**0.**070
Dmean (Gy)	Left Parotid	23.46	8.987	26.41	11.97	0.512
	Right Parotid	24.00	10.07	25.11	11.88	0.522
	Oral Cavity	29.36	9.668	28.45	9.843	0.695
	Oesophagus	24.81	10.39	24.18	12.21	0.821
	Larynx	33.33	8.561	31.64	11.76	0.678
	Right Submandibular Gland	41.47	9.615	44.67	12.11	0.223
	Left Submandibular Gland	41.31	15.39	41.32	17.39	0.709
	Lips	17.54	5.833	17.77	5.637	0.915
	Constrictor Muscle	40.826	12.118	40.714	13.776	0.726
Group Score	DigestOral group	0.791	0.275	0.822	0.289	0.773
Dmax (Gy)	Mandible	63.111	13.546	65.748	14.412	0.069
Dmean (Gy)	Right Ear	6.578	9.200	9.623	12.846	0.959
	Left Ear	6.097	9.431	7.483	11.907	0.959
	Right Cochlea	16.820	18.970	16.085	20.874	0.559
	Left Cochlea	10.380	17.871	12.129	22.371	0.515
Group Score	Bone group	0.583	0.350	0.610	0.363	0.751
Dmean (Gy)	Right Lens	1.058	2.858	1.568	2.993	0.655
	Left Lens	1.500	3.752	1.455	2.821	1.000
	Right Optic Nerve	4.756	11.609	6.916	15.011	0.655
	Left Optic Nerve	6.567	14.655	8.900	17.238	0.634
	Right Retina	1.661	4.345	2.498	5.133	0.490
	Left Retina	2.058	5.087	2.497	5.014	0.743
	Chiasm	2.727	6.093	5.543	12.365	0.884
Group Score	Optics group	0.178	0.451	0.241	0.489	0.490
Dmean (Gy)	Right Lung	2.366	1.702	2.739	3.079	0.848
	Left Lung	2.257	1.690	2.534	2.873	0.763
	Thyroid	35.808	12.646	37.029	13.175	0.750
	Pituitary	3.026	10.012	3.967	10.847	0.961
Group Score	Other group	0.710	0.351	0.779	0.511	0.554
Global score	**Plan score**	**0.773**	**0.091**	**1.621**	**3.359**	**0.000**

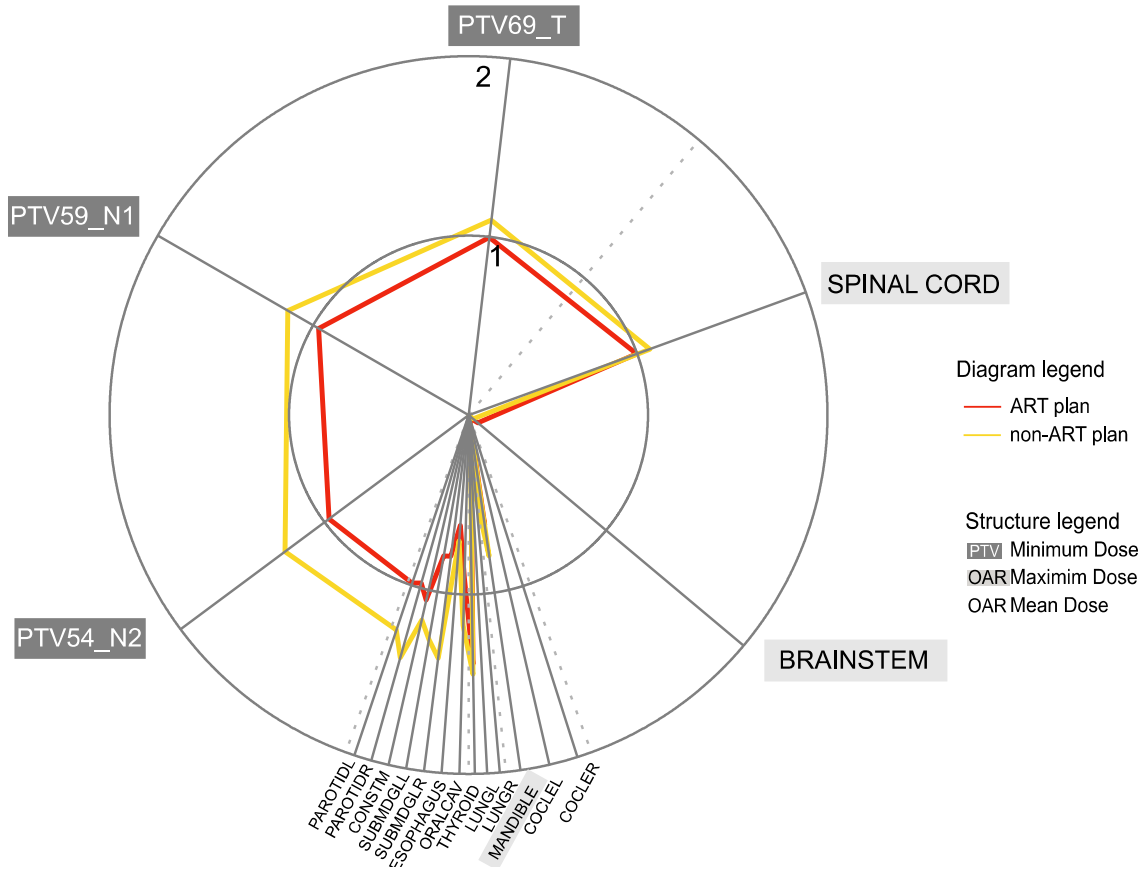

Figure 17. SPIDERplan radar plot for one patient. The red line represents the scores for the adaptive plan for each structure and target volume while the yellow line shows the scores for the non-adaptive plan.

4.4. Discussion

In this work, the impact of plan adaptation (ART) in H&N patients was assessed by investigating the dosimetric differences between performing one replan in the course of treatment as opposed to delivering the initial plan with no modification.

The dosimetric analysis was performed for all target volumes as well as seventeen OARs, providing a holistic view of the consequences of ART for the quality of the delivered treatment.

Looking at the dosimetric differences reported in table 2, it is obvious that the most significant dosimetric impact of ART is observed in the target volumes, all of which presented much lower D98% (or near minimum doses) in the non-ART scenario. The

differences are so important that the requirement of the coverage of 95% of the PTV by 95% of the prescribed dose is not met by any target, leading to the conclusion that, in the absence of a replan, dose coverage and tumoral control would be significantly impaired.

Furthermore, it is clear that the standard deviation is much higher in the non-ART scenario for all structures. This greatly increased variance implies that not performing ART could contribute to more unpredictable side effects. The actual dosimetric consequences are highly patient specific, reinforcing the need for an individual assessment and identification of patients for ART. These results are corroborated by the SPIDERplan score for the PTV group, with both mean value and standard deviation significantly higher for the non-ART scenario. This score is also the only of the six groups which is greater than one, meaning that the minimum clinical criteria are not met for the target volumes in the absence of a replan.

Regarding the OAR, most structures show an increase in the mean/maximum dose in the non-ART scenario, as can be seen in table 2. The majority of the published literature on the dosimetric impact of ART in OAR focuses on the parotid glands, due to their well known radiosensitivity which is associated with impaired salivary production, xerostomia and reduced quality of life [48,95]. There are currently relatively few studies which focus on other OAR that can also have a significant impact in the patients' life if overly irradiated, including highly critical organs such as the spinal cord and the brainstem [48,95]. Our results clearly show that, although there is an increase in the mean dose to the parotids in the absence of ART, this is not statistically significant, and the more affected OAR is the spinal cord. The maximum dose to the spinal cord increased approximately 6 Gy in average, but a very high dose variation across patients was observed. This is in line with the results from other studies which report excess dose to the spinal cord, with Hansen et al. [58] reporting an increase in the maximum dose ranging from 0.2 to 15.5 Gy in all patients of a 13 patient cohort, and Chitapanarux et al. [96] reporting a decrease in the Dmax to the spinal cord with ART in 95% of the patients in their cohort, with a dose difference ranging from 1.6 to 5.9 Gy. The average maximum dose to the spinal cord is still below the clinical threshold of 45 Gy in the ART plans, even with the observed increase. However, it is reasonable to infer that this value could escalate above the limit in the upcoming weeks of treatment, if the increase is not detected, or be a drawback in case of a future tumour recurrence needing a re-irradiation.

Despite the most significant discrepancies emerging for the target volumes and the spinal cord, the SPIDERplan global score, which gives a measure of the overall plan quality by performing a weighted score of the six different groups of structures, is also

significantly higher for the non-ART scenario, with a much higher standard deviation compared to the ART case. This shows that the individual differences observed for each structure translate, in fact, into an overall loss of plan quality, which could potentially lead to a lower treatment outcome. It is also important to point out that using an automatic planning tool to generate the dose distributions eliminates other sources of differences between the plans, such as inter or intra planner variability. Additionally, the iCycle tool is a powerful optimizer that inherently produces high quality dose distributions, which could be the reason why no statistically significant differences were found in most OARs.

Although these results highlight the advantages of ART in head and neck patients, further research on the long term impact of ART in patient survival and disease progression is still necessary to fully understand the impact and necessity of incorporating ART into routine clinical practice.

4.5. Conclusion

This study shows that the introduction of one adaptive replan during treatment translates into statistically significant differences in the coverage of the target volumes as well as the dose to OAR. Of the seventeen analysed OAR, most of them showed increased doses without ART, with the spinal cord presenting the only statistically significant differences.

Overall plan quality was also impaired without ART, as is shown by a significant reduction in the global plan score obtained using the SPIDERplan assessment tool when compared with ART.

Further work is needed to investigate the long-term effects of ART in head and neck patients.

Chapter 5

Predicting the need for adaptive radiotherapy in H&N patients

5.1. Purpose

Adaptive radiotherapy can reduce the negative dosimetric and clinical impacts of anatomical changes during head and neck (H&N) treatments. Evidence shows that ART is not equally important for all patients, making it valuable to electively schedule ART to optimize both clinical resources and the benefit for the patient.

The purpose of this study is to assess the feasibility of using both pre-treatment patient features and radiomic features from a pre-treatment contrast enhanced CT scan to predict the need for ART in H&N patients

5.2. Methodology

5.2.1. Sample description

The study sample consisted of 72 (66 male and 6 female) H&N patients from IPOCFG, E.P.E., which were treated with helical IMRT delivered by a Tomotherapy HD (Accuray) unit, with (58 patients) or without (14 patients) concomitant chemotherapy.

The vast majority (63 out of 72) of patients were prescribed 69.96 Gy to the tumour planning target volume (PTV-T), with 2 patients being prescribed 59.4 Gy. The prescription to the lymphatic nodes' PTVs (PTV-N) was either 50.4 Gy, 54 Gy or 59.4 Gy. Regarding the adenopathies, 48 patients were prescribed 69.96 Gy, 4 patients were prescribed 66 Gy, one patient was prescribed 50 Gy and one patient 59,4 Gy. The remaining patients did not present any adenopathies. There were 3 patients with a prescription of 50.4 Gy for both the PTV-T and the PTV-N, and 4 patients with dose prescribed to the PTV-T only (50 Gy for 3 patients and 70Gy for 1 patient). Depending on the prescription scheme, the treatment was delivered in either 33 (65 patients), 28 (4 patients) or 20 (3 patients) fractions.

5.2.2. Patient Imaging

All patients had a planning CT (p-CT) scan in treatment position acquired at a median of 14.5 days before the start of treatment, on which the target volumes and OAR were manually delineated by a radiation oncologist. This CT scan was acquired using the scanner Somatom Sensation Open from Siemens, with a 3 mm slice thickness.

For diagnostic purposes, patients undergo a contrast enhanced CT (d-CT) prior to the first appointment with the radiation oncologist (at a median of 27 days before the p-CT). This CT scan is often acquired at different external centres, leading to varying acquisition parameters, namely slice thickness and reconstruction kernels. The d-CT was available for 67 patients.

5.2.3. Adaptive Radiotherapy scheme

All patients underwent a daily MV-CT scan, before each treatment fraction, to assess for differences between the current position and the one in the p-CT. A rigid registration was then performed between the MV-CT and the p-CT, and a transformation vector calculated, which translated into rigid couch and roll angle shifts to correct for positional errors.

In the case where even after the rigid transformation significant discrepancies could be observed, the medical physicist was alerted, and a dose calculation was performed on the daily scan using the Planned Adaptive software module from Accuray. This dose distribution was then revised by the responsible radiation oncologist, which determined the need for a new treatment plan. If a replan was required, a second CT (re-CT) was acquired, and both the target volumes and OAR propagated from the p-CT to the re-CT by rigid and deformable image registration. These volumes were then manually corrected by the radiation oncologist and a new plan was designed by the responsible medical physicist.

From the 72 patients included in the study, 36 had at least one replan during treatment (ART group), with 3 requiring 2 replans. The remaining 36 patients were used as the control group for comparison and model development.

5.2.4. Semantic Features

Thirty-six pre-treatment factors (features), which could potentially be predictive for the need of ART were acquired for each patient.

Firstly, statistical tests were performed, to determine whether there were significant differences between the replan and control groups for each factor. For continuous features (like initial weight, GTV volume/surface area, parotid glands volume/surface area, prescription doses, etc) the normality and homogeneity of the variance were assessed first, using the Shapiro-Wilk normality test and Levene's test respectively in order to determine whether parametric or non-parametric tests were suitable. Then, the t-test and Wilcoxon sign ranked test were used to assess statistical significance in the parametric and non-parametric continuous features respectively.

For categorical features (as drinking habits, smoking habits, TMN staging, chemotherapy, etc) the chi-squared test was used. As a real clinical dataset was considered in this study, there was some missing data and not all parameters had the information for all patients.

5.2.5. Radiomics

All the radiomic analysis was performed in MATLAB R2019b. The adopted workflow was based on the guidelines provided by the Image Biomarker Standardization Initiative (IBSI) [79], regarding both image pre-processing and feature extraction, as implemented by the Standardized Environment for Radiomics Analysis (SERA) [97]. SERA is a MATLAB-based framework developed at John Hopkins University in 2019, that can handle CT, PET, SPECT or MRI images, and extract standardized radiomic features in compliance with the IBSI's guidelines. The adopted workflow was the one described in chapter 3, section 3.4.1.

Image segmentation was done from the manual contour of the GTV by the radiation oncologist in the p-CTs. Rigid and deformable image registration of the GTV from the p-CT was performed using the Velocity AI v. 3.2. software from Varian to define the region of interest (ROI) in the d-CT.

To correct for differences in voxel spacing and slice thickness (d-CTs), all images were interpolated to an isotropic voxel spacing of 1 mm x 1 mm x 1 mm, using the tri-linear interpolation algorithm. Furthermore, the intensity range of the ROI was re-segmented to include only voxels between -150 HU and 180 HU, eliminating all bone and air voxels [98], and voxels with outlier intensities (outside the range of ± 3 standard deviations)

were removed. Lastly, intensities inside the ROI were discretized in a fixed bin number approach considering 32 bins.

Having completed these pre-processing steps, 351 radiomic features (as defined by the IBSI [79]) were extracted from each ROI, which are summarized in table 3.

After extraction, all features were normalized (z-normalization) and hierarchical clustering was performed to prevent redundancy. The clustering was based on the Spearman correlation coefficient, meaning that highly correlated features were assigned to the same cluster, and all feature clusters with intra-cluster correlation > 0.9 were replaced by a meta-feature which corresponded to the average of all features in the cluster. This reduced set of non-redundant features was fed to the feature selection algorithm.

Table 3. Extracted Radiomic features.

Feature Family	Number of features
Morphological	29
Local Intensity	2
Intensity-based statistics	18
Intensity Histogram	23
Intensity-Volume Histogram	7
Gray Level Co-occurrence Matrix (GLCM) 2D/2.5D/3D	100
Gray Level Run Length Matrix (GLRLM) 2D/2.5D/3D	64
Gray Level Size Zone Matrix (GLSZM) 2D/3D	32
Gray Level Distance Zone Matrix (GLDZM) 2D/3D	32
Neighbourhood Grey Tone Difference Matrix (NGTDM) 2D/3D	10
Neighbouring Grey Level Dependence Matrix (NGLDM) 2D/3D	34
TOTAL	**351**

5.2.6. Feature selection

Feature selection was made by a greedy search algorithm that, at each iteration, eliminates one feature based on the performance of classification. This was done considering 4 different classifiers: support vector machines (SVM), naïve-Bayes, decision trees (DT) and multi adaptive regression splines (MARS). Two versions of the algorithm were implemented: 1) At each iteration the impact of the removal of each

available feature was tested by randomly splitting the data into 80% for training and 20% for testing, repeating the split 30 times and assessing the average accuracy, true positive rate (TPR) and true negative rate (TNR) of the classification; 2) The impact of the removal of each feature was assed using leave one out cross-validation for calculating the accuracy, TPR and TNR of classification.

In both versions, the feature set with better performance moves on to the next iteration of the algorithm and the whole process repeats until there is only 1 feature left.

Applying this method with different classifiers will produce possibly different feature selected sets, since the classification performance will be the best one obtained for the particular classifier used.

This process was done for the semantic features alone, for the radiomic features alone, and then for a combination of both, using the 20 last-standing features from each. For the semantic features, all the missing values were replaced by the mean value of the existing data for that factor.

5.2.7. Hyperparameter optimization

After a suitable set of variables was determined, hyper-parameter optimization was conducted for the SVM using only the selected variables.

A linear kernel was used, which means that the classification score $f(\boldsymbol{x})$ for each sample $\boldsymbol{x}$ was determined by:

$$f(\boldsymbol{x}) = \left(\frac{\boldsymbol{x}}{s}\right)' \boldsymbol{\beta} + \mathrm{b} \tag{9}$$

Where s is a constant referred to as kernel scale, $\boldsymbol{\beta}$ are the model's linear coefficients (the number of elements in $\boldsymbol{\beta}$ is the same as the number of predictors in $\boldsymbol{x}$), and b is the model bias.

The parameters considered for optimization were the kernel scale and the box-constraint, which is the parameter that controls the maximum penalty imposed on observations that violate the margin.

A Bayesian optimization algorithm implemented in the built-in SVM function of MATLAB (fitcsvm), considering a 5-fold cross-validation was used to find the optimum values for these two parameters [99].

5.2.8. Classification

The model building and classification was also performed in two ways: 1) using 30 splits of the data into 80% for training and 20% for testing, guaranteeing the balance of the two classes in both the training and testing cohort and 2) with 30 repetitions of leave-one-out cross-validation. The replacement of the missing values for scenario 1 was done based only on the training cohort at each split (average value in the training cohort). For both methods the accuracy, sensitivity, specificity, and area under the ROC curve were evaluated

In option 2 it was necessary to repeat the LOO cross-validation classification 30 times because the output of the SVM classifier is a classification score, which needs to be translated into a posterior probability to build the ROC curve. The MATLAB function *fitSVMposterior* was used to obtain the optimal score to posterior probability transformation, and for this purpose it uses a 10-fold cross-validation, leading to small differences at each run.

.

5.2.9. Robustness analysis

Feature robustness to variations in patient positioning, different degrees of noise in the image, and variation in the ROI delineation was assessed using the framework proposed by Zwanenburg et al. (2019) [91]. Based on the results obtained by the authors for a H&N squamous cell carcinoma cohort, four perturbations were selected: rotation, gaussian noise addition, volume adaption and contour randomisation, as shown in figure 18 [91].

The rotation perturbation affects both the image and the ROI mask and aims to simulate deviations in patient positioning. The *imrotate3* function from MATLAB 2019b [100] was used to rotate the 3D images and masks around the z-axis considering an angle θ of rotation, with $\theta \in [-10°, -6°, -2°, 2°, 6°, 10°]$. Trilinear interpolation was used to determine the intensities in the rotated image.

Reproducible features should also be robust to varying image noise. In order to implement the noise perturbation, first the noise variance existing in the image was determined using the method described by Chang et al. (2000) [101,102], as implemented in the *getImageNoise* function in the *PlanMetrics* toolbox of the Computational Environment for Radiological Research (CERR) [103]. Then, random noise drawn from a Gaussian distribution with the same variance was added to each image voxel.

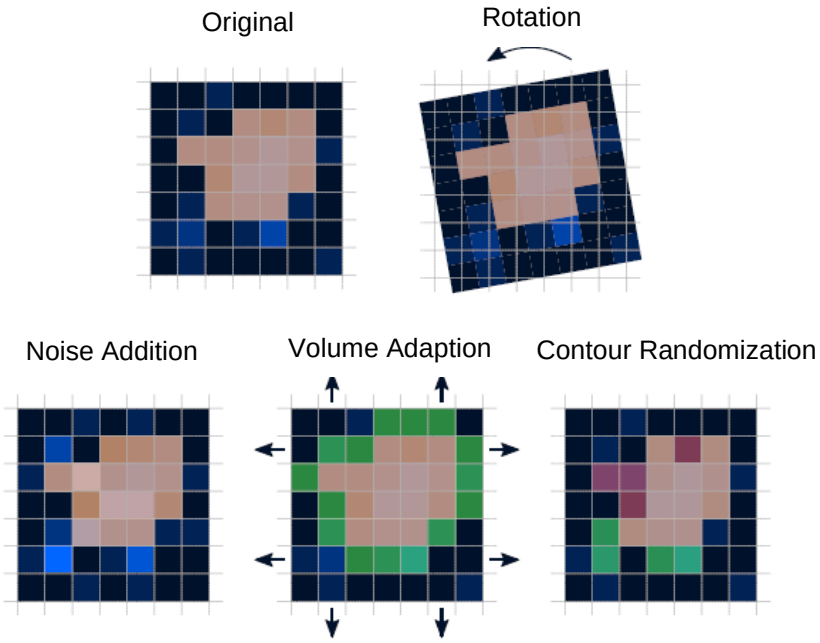

Figure 18. Image perturbations. The image is represented in blue and the region of interest mask in orange. The green squares represent added voxels to the mask, and the purple squares represent removed voxels from the mask. Adapted from [91].

Volume adaption and contour randomization aim to represent variance in the delineation of the ROI by different experts. In the volume adaption perturbation, the ROI mask was either augmented or shrunk by a given fraction $\beta \in [-0.2, -0.1, 0, 0.1, 0.2]$, through recursive dilations/erosions.

In the contour randomization perturbation, a super voxel-based segmentation algorithm was used to create an alternative contour of the ROI, using the Simple Linear Iterative Clustering algorithm as implemented by the function *superpixels3* from MATLAB 2019b [104]. The ROI was then randomised based on the overlap of the super-voxel segmentation with the original contour (the probability of a given super-voxel to be included in the final mask was equal to its overlap fraction with the original mask).

These perturbations were chained using the different values of θ for rotation and β for volume adaption, resulting in 30 perturbed images per patient. The features selected in the previous step were then calculated on each of these perturbed images and the intra-class correlation coefficient ICC(1,1) [105] was computed to assess feature robustness using the MATLAB function implemented by McGraw, et al [106].

The intraclass correlation coefficient ICC (1,1) is defined in equation (10) and essentially describes how strongly units (the same feature extracted from different perturbed

images) in the same group (patient) resemble each other. In equation (10), BMS stands for between-targets mean squares and WMS for within-target mean squares, with k being the number of judges rating each target. In this case the targets are the patients and the judges are the features extracted from the different perturbed images. So, WMS is a measure of how much a given feature varies for the same patient, given different perturbed images, and BMS is a measure of how similar this variance across the different patients is.

$$ICC(1,1) = \frac{BMS - WMS}{BMS + (k-1)WMS} \quad (10)$$

The ICC varies between 0 and 1, with 1 meaning that the feature is fully repeatable between the different image perturbations and lower values representing a higher variance in the feature value for the same patient considering different perturbations, which means lower repeatability.

5.3. Results

5.3.1. Statistical Analysis of Semantic Features

From all the analysed variables, the only ones that showed a statistically significant difference between the control and the ART groups were the planned minimum dose to the GTV (Dmin GTV) and both the volume and the surface area of the initial GTV. These results, as well as some other pre-treatment features that were considered relevant, are shown in table 4. The complete analysis of all extracted features can be found in appendix 1.

Table 4. Statistical analysis for pre-treatment factors. The mean ± standard deviation for each group is shown for the continuous variables age, initial weight, Dmin GTV, Dpresc ADNs and GTV volume (vol)/surface area (sa). The drinking habits correspond to 1- marked drinking habits, 2- moderate drinking habits, 3- no drinking habits and 4- former drinking habits.

Feature	Control	ART	p-value
Age	57.28 ± 8.77	61.19 ± 10.24	0.102
Gender (M/F)	32/2	32/4	0.157
T stage (1/2/3/4)	2/4/12/18	1/1/10/22	0.243
N stage (0/1/2/3/X)	4/5/25/1/1	4/1/26/5/0	0.242
Drinking Habits (1/2/3/4)	6/10/3/17	14/3/8/11	0.213
Initial Weight (Kg)	67.05 ± 14.14	62.39 ± 14.59	0.140
Dpresc ADNs (Gy)	52.14 ± 30,55	52.8± 29,50	0,889
Dmin GTV (Gy)	67.47 ± 4.44	64.78 ± 6.67	**0.031***
GTV vol (cm^3)	47.10 ± 46.02	85.56 ± 69.95	**0.006***
GTV sa (mm^2)	4284.08 ± 3157,49	6367.05 ± 3778.02	**0.007***

5.3.2. Feature selection and model building

5.3.2.1. Semantic features

The best results for the feature selection were obtained using the SVM classifier with LOO cross-validation (table 5). The results for the feature selection using Bayesian classifier, DT and MARS can be found in appendix 2.

Table 5. Results for the feature selection using SVM considering both LOO cross-validation and 80/20 splits.

SVM 30 reps 80/20 splits				SVM LOO			
Vars	Mean Accuracy	Mean TPR	Mean TNR	Vars	Accuracy	TPR	TNR
[1:10]	0,7077	0,627	0,796	[1:10]	0,833	0,833	0,833
[1:9]	0,7179	0,628	0,837	[1:9]	0,833	0,833	0,833
[1:8]	0,7513	0,665	0,840	[1:8]	0,764	0,694	0,833
[1:7]	0,6949	0,647	0,784	[1:7]	0,750	0,667	0,833
[1:6]	0,7385	0,668	0,821	[1:6]	0,750	0,722	0,778
[1:5]	0,6692	0,504	0,842	[1:5]	0,736	0,750	0,722
[1:4]	0,6641	0,629	0,728	[1:4]	0,667	0,556	0,778
[1:3]	0,6795	0,650	0,726	[1:3]	0,681	0,583	0,778
[1:2]	0,6846	0,655	0,720	[1:2]	0,681	0,583	0,778
1	0,6436	0,927	0,305	1	0,611	0,389	0,833
Var nº	Name			Var nº	Name		
1	DrinkingHistory_2			1	DrinkingHistory_1		
2	DrinkingHistory_4			2	Dmin GTV		
3	Nstage_3			3	Initial Weight		
4	Dmean GTV			4	ADNS Presc		
5	Nstage_2			5	Age		
6	Nstage_0			6	Tstage_1		
7	BMI			7	Nstage_3		
8	Initial Weight			8	BMI		
9	DrinkingHistory_3			9	DrinkingHistory_2		
10	Smoking_1			10	Tstage_2		

As can be seen in the previous table, the best balance between accuracy, TPR and TNR was found with six selected variables using the greedy search method with LOO cross-validation. These features were: age, initial weight, prescription dose to adenopathies (Dpresc ADNs), Dmin GTV, and the binary variables T stage=1 and marked drinking habits.

After optimization of the parameters box constraints and kernel scale for the SVM classifier with the selected variables, the results achieved for both 30 splits 80/20 and 30 repetitions of LOO cross validation for the accuracy, TPR, TNR, and area under the ROC curve can be seen in table 6.

The average ROC curves and respective upper and lower pointwise confidence bounds can be seen in figure 19.

Table 6. Results for the classification using 6 semantic features.

	Accuracy			TPR			TNR			AUC		
	Mean	Med	Std	Mean	Med	Std	Mean	Med	Std	Mean	Med	Std
30 splits 80/20	0.795	**0.788**	0.078	0.814	**0.857**	0.114	0.776	**0.714**	0.128	0.801	**0.801**	0.080
30 reps LOO	0.799	**0.806**	0.014	0.778	**0.778**	0.023	0.820	**0.819**	0.016	0.803	**0.803**	0.004

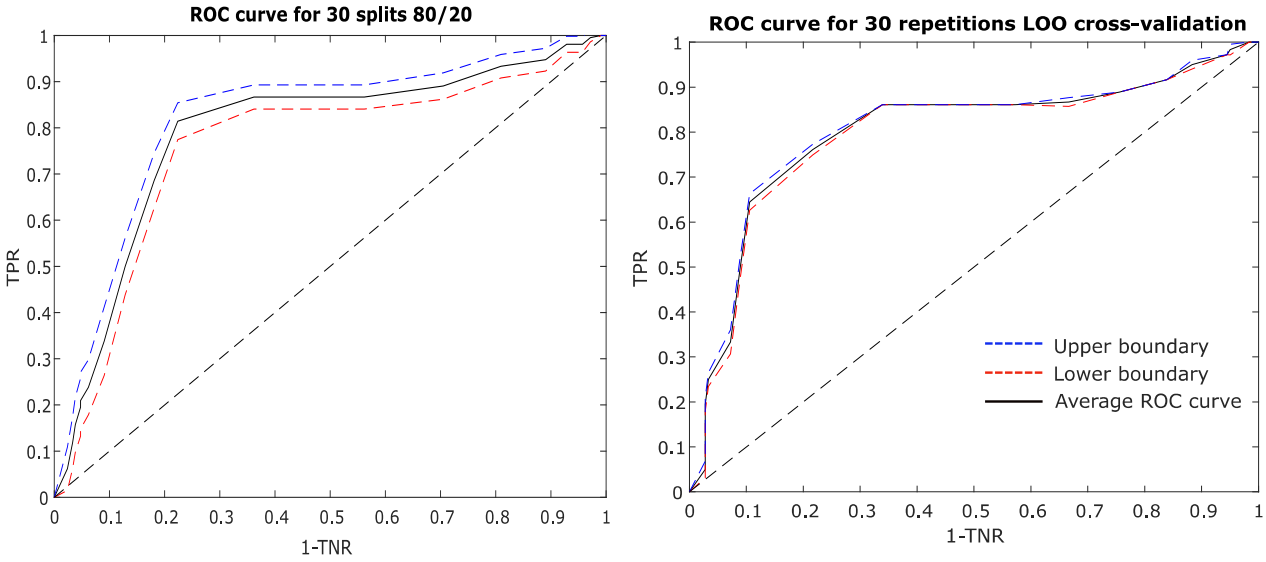

Figure 19. Average ROC curves for 30 repetitions of LOO cross-validation. The solid black curve represents the averaged ROC curve for all repetitions and the dashed blue and red curves the upper and lower pointwise confidence bounds respectively. The dashed black line is the plot diagonal.

5.3.2.2. Radiomics

The Radiomics pipeline was applied to both the d-CT and the p-CT. After the hierarchical clustering, the original set of 351 features extracted from the ROI was reduced to a non-redundant initial set of 88 features for the d-CT and 81 features for the p-CT.

Several tests for feature selection were done with 3 different classifiers: DT, Bayes and SVM, with the best results arising for SVM, which was the selected method. The MARS classifier was not considered as it was too computationally expensive. For the radiomics features, due to the high dimensionality of the problem, only the version of the feature selection algorithm considering the 30 repetitions of 80/20 splits was applied.

The best balance between accuracy, TPR and TNR was found using SVM for six selected features extracted from the ROI in the d-CT, which are shown in tables 7 and 8. From the six selected features, features R4, R5 and R6 corresponded to original extracted features and R1, R2 and R3 to clustered features. Feature R2 consists of a cluster containing eighteen highly correlated features, namely the low grey level run emphasis and the short run low grey level emphasis from the 2D averaged, 2.5D direction merged, 3D averaged and 3D direction merged GLRLM, the low grey level emphasis and the small zone low grey level emphasis from the 2D and 3D GLSZM, the low grey level emphasis from the 2D and 3D GLDZM and the low grey level count emphasis and low dependence low grey level emphasis from the 2D and 3D NGLDM.

Table 7. Results for the feature selection using SVM.

Diagnostic CT				Planning CT			
Vars	Mean Accuracy	Mean TPR	Mean TNR	Vars	Mean Accuracy	Mean TPR	Mean TNR
[1:10]	0,8071	0,804	0,815	[1:10]	0,6762	0,662	0,696
[1:9]	0,7405	0,745	0,761	[1:9]	0,7738	0,797	0,766
[1:8]	0,7786	0,760	0,806	[1:8]	0,7167	0,708	0,739
[1:7]	0,7714	0,722	0,819	[1:7]	0,7738	0,747	0,806
[1:6]	0,7762	0,738	0,819	[1:6]	0,7024	0,656	0,749
[1:5]	0,7738	0,777	0,783	[1:5]	0,6810	0,670	0,676
[1:4]	0,7357	0,712	0,761	[1:4]	0,6905	0,658	0,722
[1:3]	0,7286	0,638	0,806	[1:3]	0,6643	0,697	0,625
[1:2]	0,6976	0,790	0,646	[1:2]	0,6714	0,724	0,667
1	0,5571	0,598	0,545	1	0,6452	0,606	0,725

Table 8. Six selected features from the GTV of the d-CT.

	Feature(s)	Feature Families
R1	Grey level non uniformity normalised	GLSZM 2D GLDZM 2D
R2	Cluster with 18 features (described in the text)	GLRLM 2D/2.5D/3D (8 features) GLSZM 2D/3D (4 features) GLDZM 2D/3D (2 features) NGLDM 2D/3D (4 features)
R3	10th percentile Intensity at 90	Intensity-based statistics Intensity-volume histogram
R4	Centre of mass shift	Morphology
R5	Flatness	Morphology
R6	Small distance low grey level emphasis	GLDZM 3D

Since the results were superior for the d-CT only those images were considered for classification.

The results for the classification using the 30 splits after hyperparameter optimization are shown in figure 20 considering the SVM models with 3 to 6 features

Table 9 shows the results obtained for the models with 6 and 3 features, after hyperparameter optimization. The 3-feature model corresponds to features R1 to R3 from table 4, the 4-feature model to features R1 to R4, and so on for the models with five and six features. Figure 21 shows the ROC curves obtained for each set of variables considering LOO cross-validation.

Figure 20. Boxplots of the accuracy (right) and AUC (left) obtained for 30 repetitions of 80/20 train/test splits using 3 to 6 radiomic features in SVM.

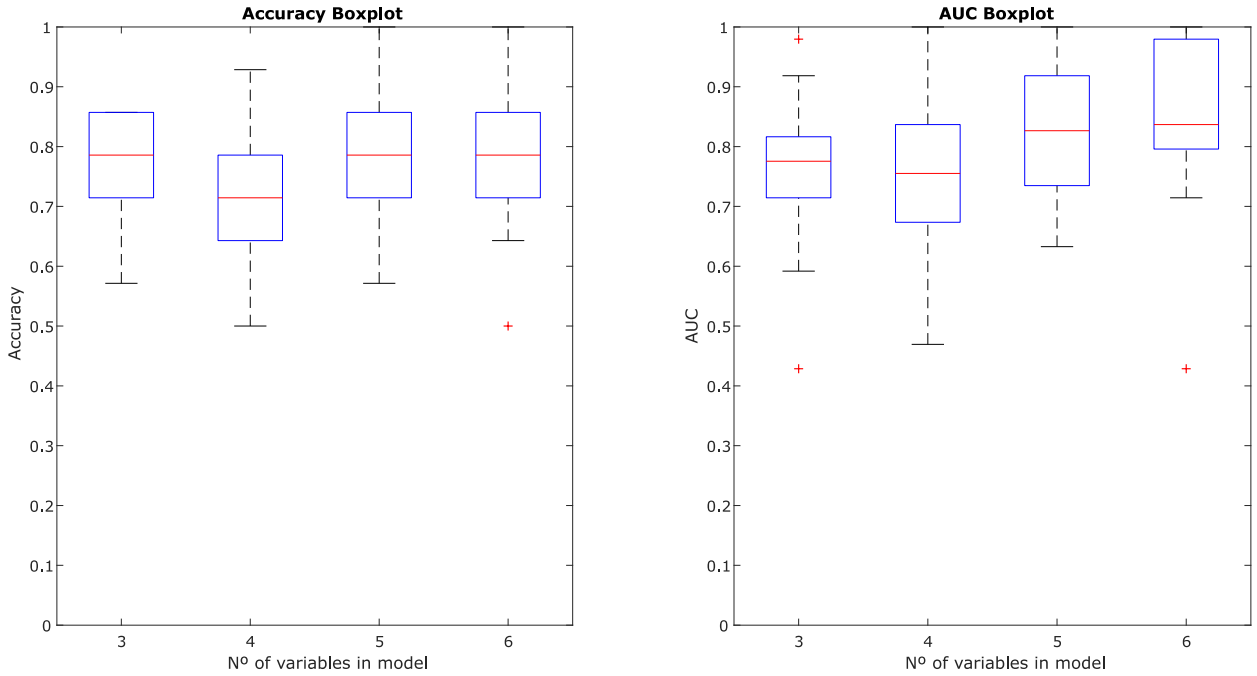

Table 9. Results obtained for the 6-feature and 3-feature radiomic models using 30 repetitions of LOO cross-validation.

	Accuracy			TPR			TNR			AUC		
	Mean	Med	Std	Mean	Med	Std	Mean	Med	Std	Mean	Med	Std
6-feature model	0.783	**0.776**	0.008	0.755	**0.758**	0.009	0.810	**0.824**	0.015	0.798	**0.799**	0.004
3-feature model	0.761	**0.761**	0.000	0.727	**0.727**	0.000	0.794	**0.794**	0.000	0.765	**0.765**	0.003

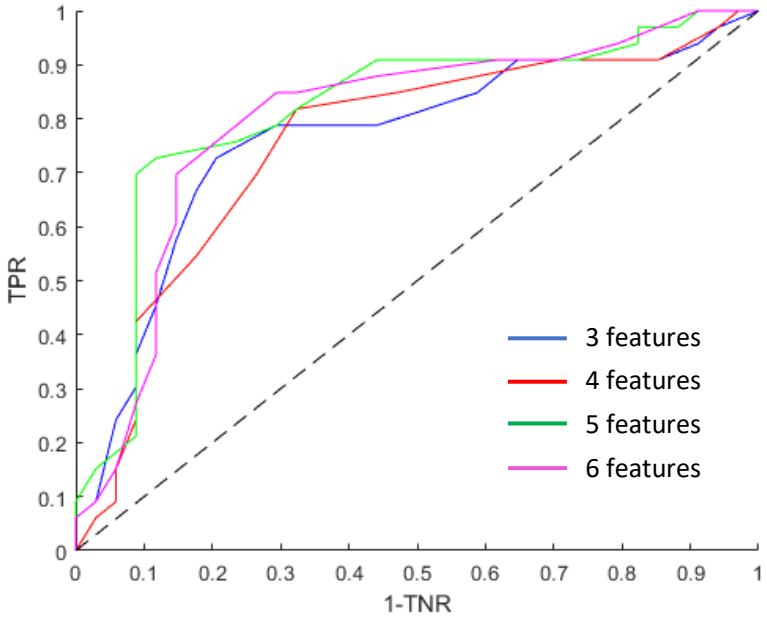

Figure 21. ROC curves obtained for the models with 3 to 6 variables using LOO cross-validation.

5.3.2.3. Combination of radiomic and semantic features

The best results from the feature selection were again achieved for six selected features, from which four were the semantic features: age, prescription dose to ADNs, Dmin to the GTV and marked drinking habits, and two were the radiomic features: a clustered feature comprising the zone-size non-uniformity normalized and the small-zone emphasis from the GLSZM 3D (feature R7), and feature R1 from table 4. The results obtained for 30 repetitions LOO cross-validation are shown in table 10. The last three features standing after the greedy search were the Dmin to the GTV, R1 and R7. The results for the 3-feature model are also shown in table 10.

Table 10. Results obtained for the 6-feature and 3-feature mixed models using 30 repetitions of LOO cross-validation.

	Accuracy			TPR			TNR			AUC		
	Mean	Med	Std	Mean	Med	Std	Mean	Med	Std	Mean	Med	Std
6-feature model	0.817	**0.821**	0.016	0.778	**0.788**	0.029	0.856	**0.853**	0.014	0.844	**0.843**	0.005
3-feature model	0.726	**0.731**	0.009	0.672	**0.667**	0.011	0.779	**0.779**	0.015	0.777	**0.776**	0.005

Regarding the robustness analysis, the results for the intra-class correlation coefficient (ICC (1,1)) obtained for each selected radiomic feature with the respective 95% confidence intervals are shown in table 11.

Table 11. ICC (1,1) and respective 95% confidence intervals (CI) for the 7 radiomic features.

Feature	ICC (1,1)	95% CI
R1	0.602	[0.523, 0.690]
R2	0.985	[0.980, 0.990]
R3	0.785	[0.725, 0.842]
R4	0.858	[0.813, 0.898]
R5	0.954	[0.937, 0.968]
R6	0.775	[0.713, 0.834]
R7	0.887	[0.850, 0.920]

5.4. Discussion

In this study, three separate models based on SVM were built with the purpose of predicting the need for ART in H&N patients prior to the beginning of treatment: 1) considering only pre-treatment clinical data from the patient; 2) considering only radiomic features extracted from pre-treatment CT images; 3) using a combination of features from 1 and 2.

The idea of applying radiomics to predict the need for ART is very recent, with only two published papers on the topic so far, to the best of our knowledge [72,73]. These studies were described in detail in section 3.4.1.7 of the State of the art.

As can be seen from the statistical analysis (table 4), the planned minimum dose to the GTV is significantly lower in the ART group. However, when comparing the ratio between the Dmin to the GTV and the prescribed dose to the PTV-T no differences were found between the two groups, which indicates that the PTV-T underdosage is not significantly larger in the ART group than it is in the control group.

Furthermore, statistically significant differences were found for the initial GTV volume and surface area between the two groups, with the ART group presenting on average an 82% higher initial volume when compared to the control group. These results are in line with several published studies that identify higher initial volumes to be associated with greater anatomical changes in the course of treatment [42,48,69,70]. Nevertheless, the initial GTV volume was not identified as a predictive factor in any of the models, which suggests that it would not add additional predictive information to the interactions between the selected variables.

Regarding the first predictive model, the features selected by the greedy-search algorithm using only pre-treatment semantic features were age, initial patient weight, the prescribed dose to the ADNs, T stage, the planned minimum dose to the GTV and whether the patient had marked drinking habits or not using the SVM model. These features are clinically relevant and in conformance with previously published results. Different studies have shown patient age to be predictive of both tumour [70] and parotid [49] shrinkage, as well as reported tumour staging to be significantly correlated with anatomical/dosimetric changes [48,69]. Looking at appendix 2 it can also be seen that the features Drinking History, T stage, Dmin GTV, age and initial weight were also selected for all three of the other tested classifiers, which confirms its ability to correctly discriminate between the need or not of ART.

As is shown in table 6, using the six selected variables in an SVM classifier and performing 30 repetitions of LOO cross-validation we were able to obtain a median accuracy of 80.6%, TPR of 77.8%, TNR of 81.9% and AUC of 80.3%. These results demonstrate that it is possible to obtain a quite reliable prediction for the need of ART using only pre-treatment semantic variables. Although the results for the 30 splits in 80/20 are comparable, the LOO cross-validation results should be a better representation of the real performance of the classifier, since the study cohort is relatively small and it is important to use as much as possible the available samples for training. Furthermore, using an 80/20 split could make the results too biased towards the training set, which is seen in the higher standard deviations of those results. As the final goal is to classify new samples using the whole of the cohort for training, the LOO cross-validation results are a good representation of the classifier performance.

Moving to the radiomics model, the best predictive performance was obtained for six selected features from the GTV volume in the d-CT. The better results obtained using the d-CT when compared to the p-CT are explained by the use of contrast enhancement in the d-CT, which results in greater detail in the tumour area, reflecting on the predictive power of the extracted radiomic features.

As the selected radiomic variables are highly dependent on the imaging modality, the cancer site and classification end-point, and this is the first study to assess CT radiomics for prediction of ART, more studies of this nature with different cohorts are needed to assess the consistency of the selected features.

The results obtained with LOO cross-validation for six features presented a median accuracy of 77.6%, a median TPR of 75.8%, a median TNR of 82.4% and finally a median AUC of 79.9%, all with standard deviations ranging from 0.4% to 1.5%.

It is very interesting to note that only three radiomic features are needed, namely the three meta-features, to reach very similar results: a median accuracy of 76.1%, TPR of 72.7%, TNR of 79.4% and AUC of 76.5% with standard deviations lower than 0.3%. This observation is not valid considering semantic features, where it is clear that much more information is lost when the total number of features is reduced. It seems reasonable to conclude that the radiomic model could have a higher ability of generalization considering new data, since the use of fewer features decreases the possibility of overfitting. Nevertheless, the results obtained using just radiomics are in general weaker than the ones considering just the semantic features.

A mixed model was also built using the information obtained from the feature selection procedure based on both the semantic features only and the radiomic features only. The feature selection algorithm showed the best performance for a set of six features, including four semantic and two radiomic features. The selected semantic features were also selected for the first model, confirming their predictive power. As for the radiomic features, one of them was the meta-feature including the grey level non uniformity normalised obtained from both the 2D grey level size and distance zone matrixes, which had also been selected in the radiomic-only model and which was the last standing feature in the radiomic greedy search, indicating its high predictive capacity. The other selected radiomic feature was also a clustered feature comprising the zone-size non-uniformity normalized and the small-zone emphasis from the GLSZM 3D, which on the contrary was not selected for the radiomic-model alone. This mixed model presented the best results out of the three models, indicating that a mixture of both semantic and radiomic variables is the most likely to be able to predict whether a given patient will need

ART during treatment. Nonetheless, the model built using only the three last standing variables, and the two radiomic features could not match the performance of the 3-feature model using only radiomic features.

One of the potential drawbacks of using a radiomics approach is that some features' values have shown to be affected by factors such as patient positioning, image acquisition parameters and segmentation [64,65]. In this way, it could be difficult to achieve a set of reproducible features, which in turn are able to generate generalizable models, if a robustness analysis is not included in the radiomic studies.

In this study, a robustness analysis was performed by applying different image perturbations, namely noise addition, rotations, volume adaptations and contour randomizations, to mimic the effects of the above mentioned factors on the calculated features, as proposed by Zwanenburg et al. (2019) [91]. From the results of this analysis, 6 out of the 7 radiomic features included in the models presented an ICC higher than 0.75, indicating a good reproducibility across the perturbed images for each patient.

The main limitation of this study is the fact that the sample cohort has a small number of patients and they all come from the same health institution. It would be interesting to assess the predictive capability of these models on unseen patients coming from different clinics. Despite this fact, the number of selected variables was six, which is a relatively low number for the size of the dataset, respecting the rule of thumb of having 10-15 samples per feature [80].

To the best of our knowledge this is the first study to use radiomic features from CT images to build a model to predict the need for ART in H&N patients. Furthermore, we compared the radiomic model with both a non-radiomic model built only with semantic features and a mixed model using a combination of radiomic and non-radiomic features

These results are very promising and in conformity with the ones reported by Ramella et al. [72] and Yu et al. [73], suggesting the ability of using radiomics and pre-treatment factors to predict the need for ART in a clinical H&N cancer dataset.

5.5. Conclusion

The presented study successfully demonstrated the ability to predict the need for ART in H&N patients from both pre-treatment semantic features and radiomic features extracted from the GTV of a contrast-enhanced CT scan acquired prior to the beginning of treatment. The best classification results were obtained considering 6 features, of which 4 were semantic and 2 radiomic features. A comparable result was achieved considering

only 3 radiomic features. A robustness analysis showed that 6 out of the 7 radiomic features that were used in the models presented good robustness across perturbed images. These models could be extremely valuable in order to optimize the ART workflow and the clinic's resources, as well as the patient outcome. Future work is needed to validate the proposed models using an independent patient cohort preferably from a different institution to incorporate them into clinical practice.

Chapter 6

Global conclusions and future work

The purpose of this dissertation was to address adaptive radiotherapy in H&N cancer patients, starting from the dosimetric impact of replanning in the course of the treatment and moving to a priori predicting which patients would require ART, in order to optimize the clinical workflow and benefit to the patient.

In the first part of the dissertation, it was shown that one adaptive replan translates into statistically significant differences in the coverage of the target volumes as well as the dose to OAR when compared to the scenario where no replanning occurs. Seventeen OARs were analysed and the vast majority of them presented higher mean/maximum doses in the non-ART scenarios, with the spinal cord showing the most marked differences in the maximum dose.

Furthermore, these dosimetric results translated into an overall loss of plan quality without ART, as it was shown by a significant increase in the global plan score obtained using the SPIDERplan tool.

These results support the idea that adaptation is required in some H&N cancer patients, and that it makes a significant difference in dosimetric terms both for targets and OAR. Nevertheless, further work is needed to investigate the long-term effects of ART in head and neck patients in terms of tumour recurrence, quality of life and overall survival.

In the second part of the dissertation, the ability to predict the need for ART was successfully demonstrated using pre-treatment semantic features and radiomic features extracted from the GTV of a contrast-enhanced CT scan acquired prior to the beginning of treatment.

The best classification results were an accuracy of 0.821 and AUC of 0.843, obtained considering 6 features, of which 4 were semantic and 2 radiomic features. A comparable result was achieved considering only 3 radiomic features, suggesting that radiomic features present higher generalizability than the semantic ones. Furthermore, the selected features showed an overall high robustness across perturbed images.

These models could be extremely valuable in order to optimize the ART workflow and the clinic's resources, as well as the patient outcome, but future work is needed to validate the proposed models using an independent patient cohort preferably from a different institution.

Blank Page

References

[1] Yan, D., Vicini, F., Wong, J., & Martinez, A. Adaptive radiation therapy. Physics in medicine and biology 1997, 42(1), 123-132.

[2] Borras JM, Lievens Y, Barton M, Corral J, Ferlay J, Bray F, Grau C. How many new cancer patients in Europe will require radiotherapy by 2025? An ESTRO-HERO analysis. Radiother Oncol 2016; 119:5–11.

[3] Miranda N, Portugal C. Programa Nacional para as Doenças Oncológicas - Doenças Oncológicas em Números 2015. Dgs 2016:5–65.

[4] Brock KK. Adaptive Radiotherapy: Moving Into the Future. Semin Radiat Oncol 2019; 29:181–184.

[5] Castelli J, Simon A, Lafond C, Perichon N, Rigaud B, Chajon E, De Bari B, Ozsahin M, Bourhis J, de Crevoisier R. Adaptive radiotherapy for head and neck cancer. Acta Oncol (Madr) 2018; 57:1284–1292.

[6] Giaccia AJ, Willson S, Shaw R, Dougherty B, Rivera B, Panetta A, Druding S. Radiobiology for the Radiologist. 7th ed. 2012. Chapters 5 and 19.

[7] Podgorsak EB. Radiation Oncology Physics: A Handbook for Teachers and Students. Vienna: International Atomic Energy Agency, 2005. Chapters 7 and 14.

[8] MC Lopes. Um século de terapia com radiação. Gazeta de Física 2007.

[9] Landberg T, Chavaudra J, Dobbs J, Hanks G, Johansson K-A, Möller T, Purdy J. Report 50. J Int Comm Radiat Units Meas 1993; os26:NP-NP.

[10] Landberg T, Chavaudra J, Dobbs J, Gerard J-P, Hanks G, Horiot J-C, Johansson K-A, Möller T, Purdy J, Suntharalingam N, Svensson H. Report 62. J Int Comm Radiat Units Meas 1999; os32:NP-NP.

[11] Breedveld S, Craft D, van Haveren R, Heijmen B. Multi-criteria optimization and decision-making in radiotherapy. Eur J Oper Res 2019; 277:1–19.

[12] Small W, Tarbell NJ, Yao M. Clinical Radiation Oncology. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2017. Pages 130 and 296.

[13] Garibaldi C, Jereczek-Fossa BA, Marvaso G, Dicuonzo S, Rojas DP, Cattani F, Starzyńska A, Ciardo D, Surgo A, Leonardi MC, Ricotti R. Recent advances in radiation oncology. Ecancermedicalscience 2017; 11.

[14] Connell PP, Hellman S. Advances in radiotherapy and implications for the next century: A historical perspective. Cancer Res 2009; 69:383–392.

[15] Elith C, Dempsey SE, Findlay N, Warren-Forward HM. An introduction to the intensity-modulated radiation therapy (IMRT) techniques, tomotherapy, and VMAT. J Med Imaging Radiat Sci 2011; 42:37–43.

[16] IAEA-TECDOC-1588. Transition from 2-D Radiotherapy to 3-D Conformal and Intensity Modulated Radiotherapy. 2008.

[17] Garibaldi C, Jereczek-Fossa BA, Marvaso G, Dicuonzo S, Rojas DP, Cattani F, Starzyńska A, Ciardo D, Surgo A, Leonardi MC, Ricotti R. Recent advances in radiation oncology. Ecancermedicalscience 2017; 11:1–19.

[18] Court LE, Balter P, Mohan R. Principles of IMRT. Intensity-Modulated Radiation Therapy: Clinical Evidence and Techniques. Springer Japan; 2015:15–42.

[19] Bénézery K. New techniques in Radiation Therapy for head and neck cancer. https://www.excemed.org/resources/new-techniques-radiation-therapy-head-and-neck-cancer Accessed at 27 April 2020.

[20] Jin J-Y, Wen N, Ren L, Glide-Hurst C, Chetty IJ. Advances in Treatment Techniques. Cancer J 2011; 17:166–176.

[21] Welsh JS, Patel RR, Ritter MA, Harari PM, Mackie TR, Mehta MP. Helical tomotherapy: An innovative technology and approach to radiation therapy. Technol Cancer Res Treat 2002; 1:311–316.

[22] New Linac Suite - Logan Construction. https://www.loganconstruction.co.uk/new-linac-suite/. Accessed 3 April 2020.

[23] Two of a Kind: VMAT Versus Tomotherapy | Imaging Technology News. https://www.itnonline.com/article/two-kind-vmat-versus-tomotherapy. Accessed 3 April 2020.

[24] Yan D, Lockman D, Martinez A, Wong J, Brabbins D, Vicini F, Liang J, Kestin L. Computed tomography guided management of interfractional patient variation. Semin Radiat Oncol 2005; 15:168–179.

[25] Green OL, Henke LE, Hugo GD. Practical Clinical Workflows for Online and Offline Adaptive Radiation Therapy. Semin Radiat Oncol 2019; 29:219–227.

[26] Schwartz DL, Dong L. Adaptive radiation therapy for head and neck cancer-can an old goal evolve into a new standard? J Oncol 2011; 2011:690595.

[27] Kibrom AZ, Knight KA. Adaptive radiation therapy for bladder cancer: A review of adaptive techniques used in clinical practice. J Med Radiat Sci 2015; 62:277–285.

[28] Keall P, Poulsen P, Booth JT. See, Think, and Act: Real-Time Adaptive Radiotherapy. Semin Radiat Oncol 2019; 29:228–235.

[29] Jia X, Tian Z, Xi Y, Jiang SB, Wang G. New concept on an integrated interior magnetic resonance imaging and medical linear accelerator system for radiation therapy. J Med Imaging 2017; 4:015004.

[30] Mutic S, Dempsey JF. The ViewRay System: Magnetic Resonance-Guided and Controlled Radiotherapy. Semin Radiat Oncol 2014; 24:196–199.

[31] Wu Q, Chi Y, Chen PY, Krauss DJ, Yan D, Martinez A. Adaptive Replanning Strategies Accounting for Shrinkage in Head and Neck IMRT. Int J Radiat Oncol Biol Phys 2009; 75:924

932.

[32] Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, Dicker DJ, Chimed-Orchir O, Dandona R, Dandona L, Fleming T, Forouzanfar MH, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study Global Burden of Disease Cancer Collaboration. JAMA Oncol 2017; 3:524–548.

[33] Yeh S-A. Radiotherapy for Head and Neck Cancer. Semin Plast Surg 2010; 24:127–136.

[34] Mendenhall WM, Mancuso AA, Amdur RJ, Stringer SP, Villaret DB, Cassisi NJ. Squamous cell carcinoma metastatic to the neck from an unknown head and neck primary site. Am J Otolaryngol - Head Neck Med Surg 2001; 22:261–267.

[35] Tobias JS. Current Issues in Cancer: Cancer of the head and neck. BMJ 1994; 308:961.

[36] Chin D, Boyle GM, Porceddu S, Theile DR, Parsons PG, Coman WB. Head and neck cancer: Past, present and future. Expert Rev Anticancer Ther 2006; 6:1111–1118.

[37] XiaoShen Wang, Avraham Eisbruch, IMRT for head and neck cancer: reducing xerostomia and dysphagia, Journal of Radiation Research, Volume 57, Issue S1, August 2016, Pages i69–i75,

[38] Ghosh-Laskar S, Yathiraj PH, Dutta D, Rangarajan V, Purandare N, Gupta T, Budrukkar A, Murthy V, Kannan S, Agarwal JP. Prospective randomized controlled trial to compare 3-dimensional conformal radiotherapy to intensity-modulated radiotherapy in head and neck squamous cell carcinoma: Long-term results. Head and Neck, vol. 38. John Wiley and Sons Inc.; 2016:E1481–E1487.

[39] Pow EHN, Kwong DLW, McMillan AS, Wong MCM, Sham JST, Leung LHT, Leung WK. Xerostomia and quality of life after intensity-modulated radiotherapy vs. conventional radiotherapy for early-stage nasopharyngeal carcinoma: Initial report on a randomized controlled clinical trial. Int J Radiat Oncol Biol Phys 2006; 66:981–991.

[40] Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, Miles EA, Miah AB, Newbold K, Tanay MA, Adab F, Jefferies SJ, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): A phase 3 multicentre randomised controlled trial. Lancet Oncol 2011; 12:127–136.

[41] Castadot P, Lee JA, Geets X, Grégoire V. Adaptive Radiotherapy of Head and Neck Cancer. Semin Radiat Oncol 2010; 20:84–93.

[42] Barker JL, Garden AS, Ang KK, O'Daniel JC, Wang H, Court LE, Morrison WH, Rosenthal DI, Chao KSC, Tucker SL, Mohan R, Dong L. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. Int J Radiat Oncol Biol Phys 2004; 59:960–970.

[43] Geets X, Tomsej M, Lee JA, Duprez T, Coche E, Cosnard G, Lonneux M, Grégoire V. Adaptive biological image-guided IMRT with anatomic and functional imaging in pharyngo-laryngeal tumors: Impact on target volume delineation and dose distribution using helical tomotherapy. Radiother Oncol 2007; 85:105–115.

[44] Bhide SA, Davies M, Burke K, McNair HA, Hansen V, Barbachano Y, El-Hariry IA, Newbold K, Harrington KJ, Nutting CM. Weekly Volume and Dosimetric Changes During Chemoradiotherapy With Intensity-Modulated Radiation Therapy for Head and Neck Cancer: A Prospective Observational Study. Int J Radiat Oncol Biol Phys 2010; 76:1360–1368.

[45] Mnejja W, Daoud H, Fourati N, Sahnoun T, Siala W, Farhat L, Daoud J. Dosimetric impact on changes in target volumes during intensity-modulated radiotherapy for nasopharyngeal carcinoma. Reports Pract Oncol Radiother 2020; 25:41–45.

[46] Castadot P, Geets X, Lee JA, Grégoire V. Adaptive functional image-guided IMRT in pharyngo-laryngeal squamous cell carcinoma: Is the gain in dose distribution worth the effort? Radiother Oncol 2011; 101:343–350.

[47] O'Daniel JC, Garden AS, Schwartz DL, Wang H, Ang KK, Ahamad A, Rosenthal DI, Morrison WH, Asper JA, Zhang L, Tung SM, Mohan R, et al. Parotid Gland Dose in Intensity-Modulated Radiotherapy for Head and Neck Cancer: Is What You Plan What You Get? Int J Radiat Oncol Biol Phys 2007; 69:1290–1296.

[48] Brouwer CL, Steenbakkers RJHM, Langendijk JA, Sijtsema NM. Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? Radiotherapy & Oncology. Volume 115, Issue 3, P285-294, 2015.

[49] Sanguineti G, Ricchetti F, Thomas O, Wu B, McNutt T. Pattern and predictors of volumetric change of parotid glands during intensity modulated radiotherapy. Br J Radiol 2013; 86:20130363.

[50] Schwartz DL, Garden AS, Shah SJ, Chronowski G, Sejpal S, Rosenthal DI, Chen Y, Zhang Y, Zhang L, Wong PF, Garcia JA, Kian Ang K, et al. Adaptive radiotherapy for head and neck cancer - Dosimetric results from a prospective clinical trial. Radiother Oncol 2013; 106:80–84.

[51] Schwartz DL, Garden AS, Thomas J, Chen Y, Zhang Y, Lewin J, Chambers MS, Dong L. Adaptive radiotherapy for head-and-neck cancer: Initial clinical outcomes from a prospective trial. Int J Radiat Oncol Biol Phys 2012; 83:986–993.

[52] Zhao L, Wan Q, Zhou Y, Deng X, Xie C, Wu S. The role of replanning in fractionated intensity modulated radiotherapy for nasopharyngeal carcinoma. Radiother Oncol 2011; 98:23–27.

[53] Surucu M, Shah KK, Roeske JC, Choi M, Small W, Emami B. Adaptive Radiotherapy for Head and Neck Cancer: Implications for Clinical and Dosimetry Outcomes. Technol Cancer Res Treat 2017; 16:218–223.

[54] Zhang P, Simon A, Rigaud B, Castelli J, Ospina Arango JD, Nassef M, Henry O, Zhu J, Haigron P, Li B, Shu H, De Crevoisier R. Optimal adaptive IMRT strategy to spare the parotid glands in oropharyngeal cancer. Radiother Oncol 2016; 120:41–47.

[55] Yang H, Hu W, Wang W, Chen P, Ding W, Luo W. Replanning during intensity modulated radiation therapy improved quality of life in patients with nasopharyngeal carcinoma. Int J Radiat Oncol Biol Phys 2013; 85:e47–e54.

[56] Nishimura Y. Adaptive radiation therapy in intensity- modulated radiation therapy for head and neck cancer. Intensity-Modulated Radiat Ther Clin Evid Tech 2015:111–130.

[57] Ahn PH, Chen CC, Ahn AI, Hong L, Scripes PG, Shen J, Lee CC, Miller E, Kalnicki S, Garg MK. Adaptive planning in intensity-modulated radiation therapy for head and neck cancers: Single-institution experience and clinical implications. Int J Radiat Oncol Biol Phys 2011; 80:677–685.

[58] Hansen EK, Bucci MK, Quivey JM, Weinberg V, Xia P. Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer. Int J Radiat Oncol Biol Phys 2006; 64:355–362.

[59] Capelle L, Mackenzie M, Field C, Parliament M, Ghosh S, Scrimger R. Adaptive Radiotherapy Using Helical Tomotherapy for Head and Neck Cancer in Definitive and Postoperative Settings: Initial Results. Clin Oncol 2012; 24:208–215.

[60] Ting KM. Confusion Matrix. Encyclopedia of Machine Learning and Data Mining. Springer US; 2017:260–260.

[61] Marques JP. Pattern Recognition 2001. Chapter 2, section 2.6, chapter 5 section 5.10.

[62] Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology 2016; 278:563–577.

[63] Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and generalizability in radiomics modeling: Possible strategies in radiologic and statistical perspectives. Korean J Radiol 2019; 20:1124–1137.

[64] Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. Int J Radiat Oncol Biol Phys 2018; 102:1143–1158.

[65] Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJWL, Dekker A, Fenstermacher D, Goldgof DB, Hall LO, et al. Radiomics: The process and the challenges. Magn Reson Imaging 2012; 30:1234–1248.

[66] Shultz TR, Fahlman SE, Craw S, Andritsos P, Tsaparas P, Silva R, Drummond C, Ling CX, Sheng VS, Drummond C, Lanzi PL, Gama J, et al. Cross-Validation. Encyclopedia of Machine Learning. Springer US; 2011:249–249.

[67] Webb GI, Sammut C, Perlich C, Horváth T, Wrobel S, Korb KB, Noble WS, Leslie C, Lagoudakis MG, Quadrianto N, Buntine WL, Quadrianto N, et al. Leave-One-Out Cross-Validation. Encyclopedia of Machine Learning. Springer US; 2011:600–601.

[68] Brown E, Owen R, Harden F, Mengersen K, Oestreich K, Houghton W, Poulsen M, Harris S, Lin C, Porceddu S. Predicting the need for adaptive radiotherapy in head and neck cancer. Radiother Oncol 2015; 116:57–63.

[69] Brouwer CL, Steenbakkers RJHM, van der Schaaf A, Sopacua CTC, van Dijk L V., Kierkels RGJ, Bijl HP, Burgerhof JGM, Langendijk JA, Sijtsema NM. Selection of head and neck cancer patients for adaptive radiotherapy to decrease xerostomia. Radiother Oncol 2016; 120:36–40.

[70] Surucu M, Shah KK, Mescioglu I, Roeske JC, Small W, Choi M, Emami B. Decision Trees Predicting Tumor Shrinkage for Head and Neck Cancer: Implications for Adaptive Radiotherapy. Technol Cancer Res Treat 2016; 15:139–145.

[71] Castelli J, Simon A, Rigaud B, Lafond C, Chajon E, Ospina JD, Haigron P, Laguerre B, Loubière AR, Benezery K, de Crevoisier R. A Nomogram to predict parotid gland overdose in head and neck IMRT. Radiat Oncol 2016; 11:1–11.

[72] Ramella S, Fiore M, Greco C, Cordelli E, Sicilia R, Merone M, Molfese E, Miele M, Cornacchione P, Ippolito E, Iannello G, D'Angelillo RM, et al. A radiomic approach for adaptive radiotherapy in non-small cell lung cancer patients. PLoS One 2018; 13:1–14.

[73] Yu TT, Lam SK, To LH, Tse KY, Cheng NY, Fan YN, Lo CL, Or KW, Chan ML, Hui KC, Chan FC, Hui WM, et al. Pretreatment Prediction of Adaptive Radiation Therapy Eligibility Using MRI-Based Radiomics for Advanced Nasopharyngeal Carcinoma Patients. Front Oncol 2019; 9:1–10.

[74] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RGPM, Granton P, Zegers CML, Gillies R, Boellard R, Dekker A, Aerts HJWL. Radiomics: Extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012; 48:441–446.

[75] Midya A, Chakraborty J, Gönen M, Do RKG, Simpson AL. Influence of CT acquisition and reconstruction parameters on radiomic feature reproducibility. J Med Imaging 2018; 5:1.

[76] Franco P, Arcadipane F, Trino E, Gallio E, Martini S, Iorio GC, Piva C, Moretto F, Ruo Redda MG, Verna R, Tseroni V, Bona C, et al. Variability of clinical target volume delineation for rectal cancer patients planned for neoadjuvant radiotherapy with the aid of the platform Anatom-e. Clin Transl Radiat Oncol 2018; 11:33–39.

[77] Piotrowski T, Gintowt K, Jodda A, Ryczkowski A, Bandyk W, Bąk B, Adamczyk M, Skorska M, Kazmierska J, Malicki J. Impact of the intra- And inter-observer variability in the delineation of parotid glands on the dose calculation during head and neck helical tomotherapy. Technol Cancer Res Treat 2015; 14:467–474.

[78] Vinod SK, Min M, Jameson MG, Holloway LC. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. J Med Imaging Radiat Oncol 2016; 60:393–406.

[79] Zwanenburg, A.; Leger, S.; Vallières, M.; Löck, S. Image Biomarker Standardisation Initiative. Image biomarker standardisation initiative. arXiv, 2016; arXiv:1612.07003.

[80] Yip SS, JWL Aerts H. Applications and limitations of radiomics. Physics in Medicine & Biology 2016. 61(13): 150-66.

[81] Hatt M, Tixier F, Pierce L, Kinahan PE, Cheze Le Rest C, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present… any future? HHS Public Access. Eur J Nucl Med Mol Imaging 2017; 44:151–165.

[82] Amadasun M, King R. Texural Features Corresponding to Texural Properties. IEEE Trans Syst Man Cybern 1989; 19:1264–1274.

[83] Sun C, Wee WG. Neighboring gray level dependence matrix for texture classification. Comput Vision, Graph Image Process 1983; 23:341–352.

[84] Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. Clin Radiol 2004; 59:1061–1069.

[85] Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep 2015; 5:1–11.

[86] Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJWL. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. Front Oncol 2015; 5.

[87] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

[88] Grosan C, Abraham A. Machine Learning. vol. 17. 2011. Chapters 3 and 6.

[89] Kuhn M., Johnson K. (2013) Nonlinear Regression Models. In: Applied Predictive Modeling. Springer, New York, NY.

[90] Wong AJ, Kanwar A, Mohamed AS, Fuller CD. Radiomics in head and neck cancer: From exploration to application. Transl Cancer Res 2016; 5:371–382.

[91] Zwanenburg A, Leger S, Agolli L, Pilz K, Troost EGC, Richter C, Löck S. Assessing robustness of radiomic features by image perturbation. Sci Rep 2019; 9:1–10.

[92] Breedveld S, Storchi PRM, Voet PWJ, Heijmen BJM. ICycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans. Med Phys 2012; 39:951–963.

[93] Ventura T, Rocha H, da Costa Ferreira B, Dias J, Lopes M do C. Comparison of two beam angular optimization algorithms guided by automated multicriterial IMRT. Phys Medica 2019; 64:210–221.

[94] Ventura T, Lopes MC, Costa B, Khouri L. SPIDERplan : A tool to support decision-making in 2018; Reports of practical oncology and radiotherapy 21 2016; 508–516.

[95] Morgan HE, Sher DJ. Adaptive radiotherapy for head and neck cancer. Cancers Head Neck 2020; 5:1.

[96] Chitapanarux I, Chomprasert K, Nobnaop W, Wanwilairat S, Tharavichitkul E, Jakrabhandu S, Onchan W, Traisathit P, Van Gestel D, A dosimetric comparison of two-phase adaptive intensity-modulated radiotherapy for locally advanced nasopharyngeal cancer, Journal of Radiation Research, Volume 56, Issue 3, May 2015, Pages 529–538

[97] Buvat I, Green AJ, Frey E, Ljungberg M. Quantitative Nuclear Medicine Imaging. IAEA Tech Doc - Taskgr Rep 2009; Manuscrip.

[98] Leger S, Zwanenburg A, Pilz K, Zschaeck S, Zöphel K, Kotzerke J, Schreiber A, Zips D, Krause M, Baumann M, Troost EGC, Richter C, et al. CT imaging during treatment improves radiomic models for patients with locally advanced head and neck cancer. Radiother Oncol 2019; 130:10–17.

[99] Optimize an SVM Classifier Fit Using Bayesian Optimization - MATLAB & Simulink. https://www.mathworks.com/help/stats/optimize-an-svm-classifier-fit-using-bayesian-optimization.html. Accessed 5 May 2020.

[100] Rotate 3-D volumetric grayscale image - MATLAB imrotate3. https://www.mathworks.com/help/images/ref/imrotate3.html#description. Accessed 4 May 2020.

[101] Chang SG, Yu B, Vetterli M. Adaptive wavelet thresholding for image denoising and compression. IEEE Trans Image Process 2000; 9:1532–1546.

[102] Ikeda M, Makino R, Imai K, Matsumoto M, Hitomi R. A method for estimating noise variance of CT image. Comput Med Imaging Graph 2010; 34:642–650.

[103] GitHub - cerr/CERR: Matlab based platform for Radiological Research. https://github.com/cerr/CERR. Accessed 4 May 2020.

[104] 3-D superpixel oversegmentation of 3-D image - MATLAB superpixels3. https://www.mathworks.com/help/images/ref/superpixels3.html. Accessed 4 May 2020.

[105] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability.1. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979, 86:420–8. Psychol Bull 1979; 86:420–8.

[106] McGraw KO, Wong SP. Forming Inferences about Some Intraclass Correlation Coefficients. Psychol Methods 1996; 1:30–46.

[107] Cancer Staging - National Cancer Institute. https://www.cancer.gov/about-cancer/diagnosis-staging/staging. Accessed 28 April 2020.

[108] Definition of Karnofsky Performance Status - NCI Dictionary of Cancer Terms - National Cancer Institute. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/karnofsky-performance-status. Accessed 28 April 2020.

Blank Page

Appendix 1 – Pre-treatment feature statistical analysis

The values of the extracted semantic features for the control and ART groups, as well as the respective p-values from the statistical analysis are shown in table A1. The categorical features histology, T stage, N stage, chemotherapy, smoking habits, drinking habits and dose levels are represented by the number of patients which presents the different levels possible for that feature. The smoking habits were categorized in 3 levels: 0 for no smoking habits, 1 for former smoking habits, 2 for active smoking habits. The drinking habits were categorized in 4 levels: 1 for marked drinking habits, 2 for moderate drinking habits, 3 for no drinking habits and 4 for former drinking habits.

For the continuous features, the mean ± standard deviation is shown for each group.

Table 12. Results from the statistical analysis

Feature	Control	ART	p-value
T stage (1/2/3/4)	2/4/12/18	1/1/10/22	0.243
N stage (0/1/2/3/x)	4/5/25/1/1	4/1/26/5/0	0.242
Gender (M/F)	32/2	32/4	0.157
Age	57.28 ± 8.77	61.19 ± 10.24	0.102
Chemotherapy (yes)	32	26	0.157
Smoking Habits (0/1/2)	7/11/18/0	7/13/14	0.199
Drinking Habits (1/2/3/4)	6/10/3/17	14/3/8/11	0.213
Initial Weight (kg)	67.05 ± 14.14	62.39 ± 14.59	0.140
Regular Weight (kg)	71.56 ± 12.40	68.75 ± 13.52	0.368
Hight (m)	1.67 ± 0.07	1.65 ± 0.09	0.221
BMI (kg/m^2)	23.91 ± 4.33	23.08 ± 5.46	0.217
Dose levels (1/2/3)	2/15/17	5/15/16	0.238
Nº fractions	32.72 ± 1.16	31.69 ± 3.78	0.415
Time Gap (days) *[1]	14.19 ± 4.22	15.97 ± 6.14	0.142
Dpresc Tumour (Gy) *[2]	68.87 ± 4.54	67.09 ± 6.67	0.261
Dpresc ADNs (Gy) *[3]	52.14 ± 30.55	52.98 ± 29.50	0.889
Dpresc homo ln (Gy) *[4]	58.45 ± 2.49	51.94 ± 19.07	0.378
Dpresc contra ln (Gy) *[5]	55.97 ± 3.02	50.57 ± 18.64	0.587
Dmax GTV (Gy)	72.22 ± 5.36	70.66 ± 7.47	0.765
Dmin GTV (Gy)	**67.47 ± 4.44**	**64.78 ± 6.67**	**0.031**
Dmedian GTV (Gy)	69.31 ± 4.74	67.43 ± 6.74	0.147
Dmean GTV (Gy)	69.37 ± 4.74	67.40 ± 6.66	0.277
Dmax BM (Gy) *[6]	32.92 ± 5.04	32.87 ± 6.43	0.744

Feature	Control	ART	p-value
Dmax BS (Gy) *[7]	21.18 ± 12.74	24.98 ± 12.27	0.215
Dmedian PH (Gy) *[8]	28.28 ± 9.01	25.72 ± 9.66	0.267
Dmean PH (Gy) *[9]	30.52 ± 8.71	29.07 ± 10.14	0.879
Dmedian PC (Gy) *[10]	25.85 ± 9.58	23.63 ± 8.60	0.554
Dmean PC (Gy) *[11]	27.87 ± 8.75	27.34 ± 10.94	0.886
GTV vol (cm^3)	**47.10 ± 46.02**	**85.56 ± 69.95**	**0.006**
GTV sa (mm^2)	**4284.08 ± 3157.49**	**6367.05 ± 3778.02**	**0.007**
PH vol (cm^3) *[12]	23.69 ± 11.54	22.39 ± 19.23	0.076
PH sa (mm^2) *[13]	2130.12 ± 2193.31	1807.10 ± 777.54	0.325
PC vol (cm^3) *[14]	21.38 ± 8.92	19.96 ± 11.02	0.128
PC sa (mm^2) *[15]	1833.56 ± 612.93	1818.63 ± 810.47	0.569
Homolat GG vol (cm^3) *[16]	0.94 ± 0.94	0.88 ± 1.11	0.433
Contralat GG vol (cm^3) *[17]	0.61 ± 1.02	0.54 ± 0.84	0.843

*[1] Time gap: days between planning CT and 1st treatment fraction

*[2] Prescribed dose to the PTV

*[3] Prescribed dose to the adenopathies

*[4] Prescribed dose to the homolateral lymph nodes

*[5] Prescribed dose to the contralateral lymph nodes

*[6] Planned maximum dose to the bone marrow

*[7] Planned maximum dose to the brainstem

*[8] Planned median dose to the homolateral parotid

*[9] Planned mean dose to the homolateral parotid

*[10] Planned median dose to the contralateral parotid

*[11] Planned mean dose to the contralateral parotid

*[12] Homolateral parotid volume

*[13] Homolateral parotid surface area

*[14] Contralateral parotid volume

*[15] Contralateral parotid surface area

*[16] Volume of the intersection of the homolateral lymph nodes with the PTV

*[17] Volume of the intersection of the contralateral lymph nodes with the PTV

Appendix 2 – Feature Selection

Pre-treatment data

Table 13. Results from the feature selection algorithm considering only pre-treatment semantic features and using Bayes classifier. The results are shown for feature sets containing 1 to 10 features. The mean Accuracy, TPR and TNR are presented.

Bayes 30 reps 80/20 splits				Bayes LOO			
Vars	Mean Accuracy	Mean TPR	Mean TNR	Vars	Accuracy	TPR	TNR
[1:10]	0.723	0.627	0.821	[1:10]	0.722	0.611	0.833
[1:9]	0.728	0.715	0.765	[1:9]	0.722	0.611	0.833
[1:8]	0.715	0.666	0.806	[1:8]	0.722	0.611	0.833
[1:7]	0.695	0.617	0.787	[1:7]	0.722	0.611	0.833
[1:6]	0.723	0.629	0.807	[1:6]	0.694	0.583	0.806
[1:5]	0.715	0.672	0.799	[1:5]	0.694	0.528	0.861
[1:4]	0.692	0.582	0.839	[1:4]	0.708	0.611	0.806
[1:3]	0.710	0.567	0.865	[1:3]	0.694	0.583	0.806
[1:2]	0.641	0.447	0.930	[1:2]	0.667	0.556	0.778
1	0.610	0.366	0.922	1	0.625	0.361	0.889
Var nº	Name			Var nº	Name		
1	Nº Sessions			1	Nº Sessions		
2	Dmin GTV			2	GTV vol		
3	GTV vol			3	Time Gap		
4	Age			4	Dmin GTV		
5	PH sa			5	Dpresc contra LN		
6	Dmax GTV			6	PH sa		
7	Smoking			7	Age		
8	Drinking History			8	Drinking History		
9	Dmean PC			9	Tumour Presc		
10	Dmedian GTV			10	Initial Weight		

Table 14. Results from the feature selection algorithm considering only pre-treatment semantic features and using Decision Tress classifier. The results are shown for feature sets containing 1 to 10 features. The mean Accuracy, TPR and TNR are presented.

DT 30 reps 80/20 splits				DT LOO			
Vars	Mean Accuracy	Mean TPR	Mean TNR	Vars	Accuracy	TPR	TNR
[1:10]	0.667	0.629	0.753	[1:10]	0.736	0.750	0.722
[1:9]	0.721	0.720	0.710	[1:9]	0.722	0.694	0.750
[1:8]	0.685	0.661	0.708	[1:8]	0.708	0.694	0.722
[1:7]	0.672	0.653	0.711	[1:7]	0.722	0.694	0.750
[1:6]	0.659	0.705	0.674	[1:6]	0.736	0.694	0.778
[1:5]	0.623	0.568	0.727	[1:5]	0.722	0.667	0.778
[1:4]	0.600	0.604	0.674	[1:4]	0.722	0.667	0.778
[1:3]	0.646	0.591	0.761	[1:3]	0.722	0.667	0.778
[1:2]	0.564	0.548	0.644	[1:2]	0.653	0.722	0.583
1	0.695	0.665	0.731	1	0.681	0.611	0.750
Var nº	Name			Var nº	Name		
1	Drinking History			1	Drinking History		
2	Nº fractions			2	Age		
3	Dmean PC			3	Dmean PC		
4	Age			4	TimeGap1		
5	Nº Sessions			5	Dpresc contra LN		
6	Tumor Presc			6	GTV sa		
7	Dmin GTV			7	Dmax MED		
8	DmeanPH			8	PH vol		
9	**Dpresc contra LN**			9	Chemotherapy Protocol		
10	Dose levels			10	GTV vol		

Table 15. Results from the feature selection algorithm considering only pre-treatment semantic features and using Bayes classifier. The results are shown for feature sets containing 1 to 10 features. The mean Accuracy, TPR and TNR are presented.

MARS 30 reps 80/20 splits				MARS LOO			
Vars	Mean Accuracy	Mean TPR	Mean TNR	Vars	Accuracy	TPR	TNR
[1:10]	0.736	0.626	0.875	[1:10]	0.653	0.722	0.583
[1:9]	0.708	0.572	0.845	[1:9]	0.653	0.722	0.583
[1:8]	0.697	0.570	0.855	[1:8]	0.667	0.694	0.639
[1:7]	0.713	0.590	0.838	[1:7]	0.681	0.722	0.639
[1:6]	0.697	0.615	0.797	[1:6]	0.694	0.722	0.667
[1:5]	0.726	0.671	0.819	[1:5]	0.694	0.722	0.667
[1:4]	0.679	0.613	0.798	[1:4]	0.708	0.778	0.639
[1:3]	0.697	0.588	0.846	[1:3]	0.708	0.778	0.639
[1:2]	0.685	0.626	0.773	[1:2]	0.681	0.667	0.694
1	0.600	0.864	0.260	1	0.597	0.639	0.556
Var nº	Name			Var nº	Name		
1	DrinkingHistory_2			1	GTV sa		
2	DrinkingHistory_4			2	Dmedian PH		
3	Nstage_1			3	PH vol		
4	Doselevels_2			4	Smoking_0		
5	Doselevels_1			5	DmedianPC		
6	Smoking_0			6	Tstage_4		
7	Doselevels_3			7	DmeanPH		
8	nSessions_2			8	Homolat GG vol		
9	nSessions_1			9	Dmean PC		
10	Tstage_2			10	Dmin GTV		

Radiomics

Table 16. Results for the feature selection classifier using 30 80/20 splits for the radiomic features in the diagnostic and planning CT, considering the Bayes classifier. The results for the mean accuracy, TPR and TNR for the feature sets containing 1 to 10 features are presented.

Bayes	Diagnostic CT			Planning CT			
Vars	Mean Accuracy	Mean TPR	Mean TNR	Vars	Mean Accuracy	Mean TPR	Mean TNR
[1:10]	0.752	0.813	0.730	[1:10]	0.750	0.776	0.746
[1:9]	0.748	0.800	0.673	[1:9]	0.731	0.677	0.797
[1:8]	0.769	0.833	0.708	[1:8]	0.769	0.745	0.795
[1:7]	0.790	0.862	0.718	[1:7]	0.774	0.814	0.750
[1:6]	0.769	0.799	0.731	[1:6]	0.729	0.773	0.704
[1:5]	0.757	0.818	0.701	[1:5]	0.771	0.781	0.753
[1:4]	0.745	0.832	0.662	[1:4]	0.764	0.798	0.756
[1:3]	0.743	0.846	0.664	[1:3]	0.712	0.661	0.785
[1:2]	0.676	0.827	0.562	[1:2]	0.669	0.645	0.690
1	0.674	0.721	0.650	1	0.636	0.675	0.618

Table 17. Results for the feature selection classifier using 30 80/20 splits for the radiomic features in the diagnostic and planning CT, considering the Decision Trees classifier. The results for the mean accuracy, TPR and TNR for the feature sets containing 1 to 10 features are presented.

Decision Trees	Diagnostic CT			Planning CT			
Vars	Mean Accuracy	Mean TPR	Mean TNR	Vars	Mean Accuracy	Mean TPR	Mean TNR
[1:10]	0.702	0.755	0.705	[1:10]	0.702	0.737	0.673
[1:9]	0.726	0.704	0.719	[1:9]	0.740	0.785	0.696
[1:8]	0.702	0.636	0.760	[1:8]	0.702	0.749	0.669
[1:7]	0.760	0.750	0.760	[1:7]	0.736	0.774	0.701
[1:6]	0.760	0.743	0.759	[1:6]	0.717	0.689	0.737
[1:5]	0.736	0.733	0.762	[1:5]	0.717	0.741	0.706
[1:4]	0.745	0.687	0.814	[1:4]	0.738	0.690	0.778
[1:3]	0.755	0.702	0.829	[1:3]	0.662	0.652	0.674
[1:2]	0.671	0.620	0.750	[1:2]	0.638	0.576	0.679
1	0.698	0.678	0.689	1	0.662	0.622	0.731