S.N. Visser

Deep Learning Autoplanning In Photon Beam Radiotherapy For Head And Neck Cancer

Deep Learning Autoplanning In Photon Beam Radiotherapy For Head and Neck Cancer

By

S.N. Visser

in partial fulfilment of the requirements for the degree of

Master of Science
in BioMedical Engineering

at the Delft University of Technology,
to be defended publicly on Thursday July 20, 2023 at 14:00 PM.

Supervisor:	Dr. ir. D. Lathouwers	
Thesis committee:	Dr. ir. D. Lathouwers,	TU Delft
	Dr. ir. Z. Perko,	TU Delft
	Dr. ir. F. Dankers,	LUMC

TU Delft

Deep Learning Autoplanning in Photon Beam Radiotherapy for Head and Neck Cancer

Simone Visser
Delft University of Technology
Leiden University Medical Center

July 14, 2023

Abstract

Introduction: Head and neck cancer (HNC) is a common and diverse group of tumors located in the region from the nasopharynx down to the upper part of the esophagus. Radiotherapy plays a crucial role in HNC treatment. In this thesis the particular focus is on external photon beam radiotherapy. The research is conducted in collaboration with the radiotherapy department of the Leiden University Medical Center (LUMC).
Theoretical background: The quality of a radiation treatment plan significantly affects patient outcomes in radiotherapy. Excessive radiation to organs at risk (OARs) can lead to complications, while insufficient dose to the tumor may increase the risk of recurrence. Automating the treatment planning process has gained attention in recent years, aiming to improve plan consistency, quality, and planning time. This study focuses on the optimization of treatment planning using RayStation's deep learning autoplanning (DLAP) for patients with HNC.
Method: The study consists of three patient cohorts. The first and largest cohort includes 43 oropharynx patients. The second cohort includes eleven hypopharynx patients and eight larynx patients. The third cohort consists of nine unilateral oropharynx patients. Dosimetric analysis and normal tissue complication probability (NTCP) analysis are performed for both the clinical and DLAP plans in all cohorts. Dosimetric analysis uses parameters from dose volume histograms (DVH), while the NTCP analysis follows the Dutch National Indication Protocol for Proton Therapy for HNC.
Results: An initial sub-study determines the optimal tuning for the DLAP model, which is not only used in this study but also chosen for clinical implementation at the LUMC. In the following comparison study, all patient cohorts demonstrate higher Planning Target Volume (PTV) coverage in the DLAP compared to the clinical plans. The OAR dosimetric parameters show varying results, with DLAP generally demonstrating similar or better sparing of the OARs in oropharynx, hypopharynx, and larynx tumors. However, DLAP show a significantly higher dose in the brain stem core for larynx patients. Furthermore, an increased dose is observed in the mandible across all patient cohorts. Unilateral oropharynx patients treated with DLAP show a significant increase in dose to several OARs, particularly the contra-lateral salivary glands and swallowing muscles.
The NTCP analysis does not reveal notable improvements or worsening across all patient cohorts.
Conclusion: DLAP demonstrates promising results for oropharynx patients, raising the question if further improvements in PTV coverage to achieve lower doses in the OARs are necessary. Although the patient cohorts for hypopharynx and larynx are small, the study's findings indicate the potential for generating adequate treatment plans in these HNC regions. Furthermore, the results also highlight the need for further investigation in unilateral oropharynx cases, as DLAP did not sufficiently spare the contra-lateral side. The divergence in treatment technique suggests waiting for a specifically trained and designed DLAP for unilateral oropharynx patients.

Key words: Photon Beam Radiotherapy, Head and Neck Cancer (HNC), Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Automated Treatment Planning

Acknowledgement

This master thesis would not have been possible without the support and supervision of dr. ir. F. Dankers (LUMC) and dr. ir. D. Lathouwers (TU Delft). In this acknowledgement I would like to express my gratitude to both of them for their time, help and feedback on this thesis. Additionally I would like to thank the colleagues, especially the medical physicist and my fellow-students, of the radiotherapy department of LUMC for all the help and resources to make the study possible.

Contents

1 Introduction

Head and neck cancer (HNC) is a common and diverse group of tumors located in the region from the nasopharynx down to the upper part of the esophagus. Figure 1 provides an anatomical illustration of the different regions that can be affected by HNC, including the oral cavity (mouth), salivary glands, nasal cavity, paranasal sinuses, pharynx (throat) and larynx (voice box). The pharynx is subdivided in the nasopharynx, oropharynx and hypopharynx. Brain tumors are not considered part of HNC.
In the Netherlands, the incidence of HNC reached over 3000 new cases in 2021 [1]. Among Dutch men, HNC ranks as the eighth most common type of cancer, while among Dutch women, it is the ninth most common. The primary risk factors associated with HNC are tobacco smoking and alcohol consumption, which contribute significantly to tumors affecting the oral cavity, pharynx, and larynx [1, 2]. Furthermore, the importance of human papillomavirus (HPV) as a significant risk factor for oropharynx tumors is increasingly recognized [2, 3, 4]. Certain types of HNC, such as lip tumors, can be detected in early stages, allowing for an early intervention. However, tumors in the nasopharynx, oropharynx, and hypopharynx regions are typically identified on a later stage, with late detection rates of 71%, 67%, and 79%, respectively [1]. Symptoms associated with HNC vary depending on the tumor location. The most common symptoms are a lump in the neck region, a persistent sore throat, difficulties or pain while swallowing, continuing hoarseness and non-healing ulcers in the head and neck region [5]. Diagnosing HNC requires thorough medical evaluation by physical examination, medical imaging, tissue biopsy, etc.

The treatment of HNC involves a range of options, including surgery, radiotherapy, chemotherapy, immunotherapy, or a combination thereof [4, 6]. The choice of treatment depends on various factors, such as the tumor's location, stage, as well as the patient's overall health. With surgery the primary tumor and possible metastases are removed. In chemotherapy drugs are used to kill the cancer cells from inside the patient's body, while immunotherapy boost the patient's immune system to recognize and fight cancer cells [7]. Radiotherapy uses harmful ionizing radiation to target and damage the DNA of cancer cells. The most frequently used treatment modalities for HNC are surgery and radiotherapy [6, 8]. Where in this thesis the main focus is on the later treatment modality, in particular external beam photon radiotherapy (EBRT).
This thesis is conducted on behalf of and in collaboration with the radiotherapy department of the Leiden University Medical Center (LUMC).

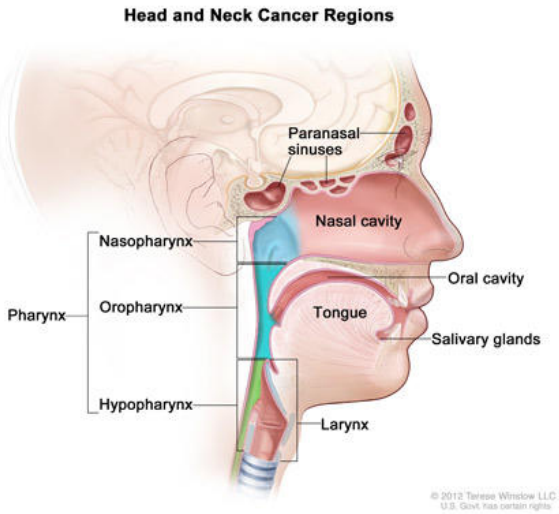


Figure 1: Anatomy of head and neck: HNC regions [9].

2 Theoretical background

2.1 Radiotherapy

Radiotherapy, a widely used treatment for cancer, uses high-energy radiation to selectively eliminate cancer cells while preserving healthy tissue. In EBRT a photon radiation beam is generated externally and directed through the patient. Cancer cells are sensitive to ionizing radiation and radiotherapy uses that sensitivity. When exposed to ionizing radiation the DNA within cancer cells is damaged, primarily through the breakage of DNA strands. This DNA damage disrupts the cell's ability to divide and multiply, ultimately leading to cell death [7]. EBRT, the most common form of radiotherapy, uses a linear accelerator to generate high-energy photons for treatment [4]. The radiation is delivered from outside the patient's body, needing accurate localization of the tumor for minimizing radiation exposure to healthy tissue. To achieve this, a precise initial localization of the tumor is essential, typically accomplished through a Computer Tomography (CT) scan. The CT scan provides a detailed image that allows for the segmentation of the tumor (target) and surrounding organs at risk (OAR). In HNC the primary tumor and the elective glands are segmented as targets. The elective glands are segmented as target because they are known for their tendency to metastasize, but they do receive a lower dose than the primary tumor. Further details regarding treatment delivery and the two different dose levels will be provided later on in this section.

Due to the complex anatomy of the head and neck region, the targets are surrounded by numerous OARs, including the brain, brain stem, swallowing muscles, lower jawbone, esophagus, oral cavity, eyes, spinal cord, etc. [4, 10, 11]. All these different OARs need to be segmented individually.
To spare the OARs as much as possible, accurate patient positioning during treatment is crucial. The patient position during treatment must align precisely with the position during the initial CT scan. To achieve this, the patient's head and neck are immobilized using a custom-made mask. This patient-specific mask is marked to indicate the isocenter, which corresponds to the position used during the CT scan and is necessary for treatment planning and delivery.

The segmentation of the targets and the OARs provides the input for the next step in the radiotherapy workflow, known as treatment planning. Treatment planning is a crucial and complex step in the radiotherapy workflow. Radiation therapy technicians (RTT) use specialized treatment planning software and algorithms to create a personalized radiation treatment plan that aligns with the unique anatomy of the patient. However, in the case of HNC, treatment planning poses particular challenges due to the complex anatomy in this region. Skilled and experienced RTTs are required for this labor-intensive task, which can take up to half a day or even longer for a single patient with HNC. Manual treatment planning by RTTs also has the drawback of resulting in inconsistent treatment plans [12]. It is important to note that there is no single definitive solution in treatment planning, leading to variations in treatment plans both between different radiotherapy institutions and even within the same radiotherapy institution.

The quality and effectiveness of a treatment plan significantly impacts the outcome of a patient's cancer treatment. Excessive radiation dose to OARs can cause temporary or permanent side effects or increase the risk of developing secondary malignancies [13]. However, inadequate radiation dose to the tumor may increase the risk of recurrence [6]. Therefore, treatment planning is aimed to achieve an optimal balance between delivering an effective radiation dose to the tumor and minimizing the dose to surrounding healthy tissues. This is crucial for maximizing treatment efficacy and minimizing potential complications.

In recent years, radiotherapy have introduced sophisticated treatment delivery methods such as Intensity-Modulated Radiation Therapy (IMRT) and Volumetric Modulated Arc Therapy (VMAT) [8, 14]. These techniques have significantly improved the ability to spare OARs during the radiation treatment. IMRT is a highly precise technique that uses multiple radiation beams with varying intensities to deliver dose to the tumor. By modulating the intensity of each beam, IMRT allows for a patient specific delivery of radiation doses within the treatment area. This customization shapes the radiation to the form of the tumor, enhancing minimization of radiation exposure to surrounding OARs [15].

VMAT, an advanced form of IMRT, involves the continuous rotation of the radiation delivery system around the patient during treatment. This rotational movement, combined with dynamic modulation of radiation beam intensity, enables the delivery of highly precise and efficient dose distributions. VMAT enhanced a shorter treatment time and improves dose conformity [16]. For the purpose of this thesis, the focus is solely on VMAT, since this is the chosen treatment delivery technique for the treatment plans that will be investigated.

In treatment planning it is the goal to establish optimal patient specific beam parameters, which means there needs to be enough dose in the planning target volume (PTV) and minimal dose in the OARs. The conventional approach to developing a treatment plan involves using a class solution, which consists of a predefined set of optimization objectives and weights for both the tumor and the OARs, specific to a particular tumor site. Inverse treatment planning generates a treatment plan using multi-objective optimization techniques to adjust treatment plan parameters. This optimization process takes into account clinical goals and criteria for targeting the tumor regions and OARs, resulting in a deliverable treatment plan. The specific clinical goals for HNC, as used in the LUMC, for both the target structures and OARs will be elaborated upon later in this thesis in section 3.3.2.

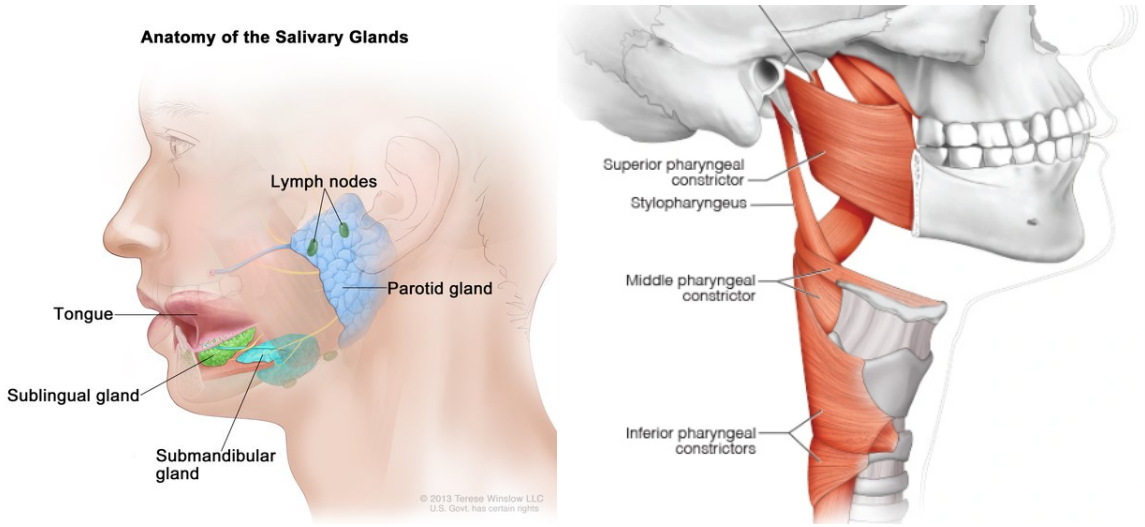
Treatment planning is conducted based on the initial CT scan, also called the planning CT scan, which is calibrated to convert Hounsfield units to electron densities for accurate dose calculation. Patients are typically scanned in their final treatment position to ensure precision. For HNC patient this means that the planning CT scan is made while the patient is wearing the patient specific immobilization mask. After treatment planning, a thorough evaluation of the treatment plan is performed to ensure that all the clinical goals are met. If necessary, further optimization can be achieved by the RTT through adjusting the objective function weights and adding additional objectives or contours. This iterative trial-and-error process continues until both the RTT and radiation oncologist are satisfied with the final treatment plan. Thus, a lot of time and experience is required here to achieve an optimal treatment plan.

Once the treatment plan is approved by both the radiation oncologist and the medical physicist, it can be initiated and delivered to the patient. However, rather than administering the entire radiation dose in a single session, the treatment is divided into fractions over multiple sessions.
In the case of HNC, it is common to make us of two different dose levels for the primary tumor and the elective glands. The primary tumor typically receives a total dose of 70 Gy (Gray), while the elective glands receive a total dose of 54.25 Gy. The delivery of the two dose levels in different target structures at the same time is called a simultaneously integrated boost (SIB) [17]. The treatment is delivered in daily fractions of 2 Gy per fraction, with a total of 35 fractions. This fractionation scheme allows for effective tumor control while minimizing the risk of complications [18]. Due to the fractionated approach, the overall duration of the treatment extends over a period of up to seven weeks. Patients typically undergo daily treatment sessions from Monday to Friday, with weekends serving as rest days. This extended treatment time frame allows healthy tissues to recover between fractions and helps to minimize the impact of radiation on normal healthy cells, while still delivering an adequate dose to the tumor cells.

2.2 Complications in Radiotherapy

Fractionated dose delivery plays a critical role in minimizing radiation-induced complications by allowing OARs to recover between treatment fractions. The tolerance of OARs to radiation varies depending on the tissue type and structure. OARs can be categorized as serial or parallel organs, which influences their radioresistance and potential complications [19]. Serial structured OARs are composed of functional subunits (FSUs), and irradiating a high dose to a single FSU can result in complications. In contrast, parallel OARs experience complications when multiple FSUs are affected, exceeding a critical threshold. Consequently, parallel organs can tolerate a higher dose to a small volume, while this is not feasible for serial organs. Therefore, considering the maximum dose (D_{max}) for serial organs and the mean dose (D_{mean}) for parallel organs is crucial. Examples of parallel organs include the lungs, parotids and other salivary glands, while the spinal cord and esophagus are examples of serial organs

[19, 20, 21]. In the context of HNC radiotherapy, the presence of multiple serial and parallel OARs highlights the importance of optimizing the dose distribution to minimize complications.



(a) Salivary glands [22].

(b) Pharyngeal constrictor muscles [23]

Figure 2: Anatomy of head and neck: salivary glands (2a) and pharyngeal constrictor muscles (2b)

Within HNC radiotherapy, two major complications often arise: xerostomia (dry mouth) and dysphagia (swallowing difficulties) [21, 24, 25]. Xerostomia occurs due to radiation-induced damage to the salivary glands, either partially or entirely. The anatomical location of the salivary glands can be seen in figure 2a. Dysphagia, on the other hand, results from radiation-induced damage to the swallowing muscles, e.g. the pharyngeal constrictor muscles (PCM). The anatomical location of the PCM superior, middle and inferior can be seen in figure 2b.

Although xerostomia may initially seem like a mild side effect, it can have severe consequences. When the salivary glands fail to produce an adequate amount of saliva, it can become extremely challenging for patients to eat and swallow solid foods. This condition could also affects their ability to speak properly. Dependency on tube feeding and the potential loss of speech can significantly diminish a patient's quality of life [26].
Similarly, dysphagia causes difficulties in swallowing and eating. In mild cases, patients may need to take sips of water to swallow food. The presence of dysphagia increases the risk of choking on food and may lead to frequent bouts of coughing. In more severe cases, individuals may become entirely dependent on tube feeding. Additionally, swallowing itself can be very painful due to dysphagia [26].

Both xerostomia and dysphagia can have a severe impact on a patient's ability to eat, swallow, and communicate properly. These complications can result in the need for long-term assistance with nutrition and a decrease in overall quality of life. Therefore, precise treatment planning and delivery is crucial to minimize radiation damage to critical structures within the HNC region, thereby reducing the possible side effect and the impact on patient's quality of life during and after radiotherapy treatment.

2.3 Automated Treatment Planning

In recent years, automating parts of the radiotherapy workflow gained significant attention in the field of radiotherapy research. For example, automated segmentation has undergone extensive research and has, in several radiotherapy institutions, replaced manual segmentation for certain commonly encountered tumor locations [27].

Another field that is under extensive research is automating the treatment planning step of the radio-

therapy workflow. This topic has attracted research interest within the radiotherapy community for years. Various algorithms are being developed, offering a promising approach to enhance plan consistency, quality and limit planning time. Recently, the use of artificial intelligence (AI) emerged in this field of research as well. Its primary focus is to explore and create techniques capable of mimicking human intelligence, like visual recognition and problem-solving [27].
Separate to this thesis a literature study is conducted on different types of autoplanning available in HNC radiotherapy. This literature study is included in the appendix (section 7.1) of this thesis.

Various software solutions are available to facilitate automated treatment planning. RayStation, developed by RaySearch Laboratories, is a widely used treatment planning system (TPS) and is part of the standard practice in numerous radiotherapy institutes worldwide. RayStation incorporates a deep learning model to automatically generate treatment plans for specific tumor sites [28]. The model is trained with prior treatment plans and clinical outcomes to provide intelligent guidance and automate various aspects of the treatment planning process. Through the integration of deep learning, the system can learn patterns and correlations, allowing for more efficient and accurate treatment plan generation. With the automation of certain treatment planning tasks, the model can reduce the burden on RTTs freeing up their time for other responsibilities. Additionally, the deep learning model is expected to enhance the quality of treatment plans by incorporating knowledge from a range of clinical cases, leading to more personalized and precise treatment strategies. In this study, the focus will center on automated treatment planning by RayStation's deep learning autoplanning (DLAP) in patients with HNC, specially trained for oropharynx patients. By using DLAP the treatment planning time is expected to be reduced tremendously since generating a treatment plan is done in approximately fifteen minutes.

The DLAP module in RayStation is designed to automate the treatment planning process and generete a treatment plan as good as an experienced RTT. The oropharynx DLAP model is trained by a patient dataset from the Princess Margaret Cancer Center in Toronto, Canada [28]. The patient dataset contains the CT-images, structures and dose distribution of hundred previously treated bilateral oropharynx patients. The treatment parameters of the training data are shown in table 1.

Table 1: DLAP treatment plan parameters of training data [28, 29]

Training Data	
Number of plans	100
Origin	Princess Margaret Cancer center
Training completion date	18 September 2021
Treatment position	Head First Supine
Modality	Photons
Energy	6 MV
Prescribed dose	70/56 Gy (SIB)
Dose per fraction	2/1.6 Gy
Type of plans	All clinically approved, peer reviewed and used for delivery

The model overview and modol protocol are included in the appendix in section 7.2 and 7.3. After training DLAP with the data, the DLAP is validated with ten other patients by RaySearch. The validation report (2022) of the DLAP oropharynx model is included in the appendix of this thesis, in section 7.4.

As input DLAP requires the CT image of the patient and segmentation of regions of interest (ROI). Thirty ROIs (both target and OAR) are involved in autoplanning and twenty-two of them are required for DLAP to generate a treatment plan. A list of the required and optional ROIs are included in the appendix, section 7.5.
The output of the DLAP is a predicted 3D dose distribution. The predicted dose is generated by a convolutional neural network, specifically a U-Net. This prediction is not directly useful in practice, for clinical application a plan and a deliverable dose is needed. A radiotherapy plan consists of beam segments, gantry angles, and monitor units, from which a deliverable dose can be calculated. DLAP

creates a plan and deliverable dose by mimicking the predicted dose in a way that it is deliverable to a real patient. The predicted dose from the DLAP is optimized by a preset of clinical goals for the targets and OARs [28]. This is an iterative process where the optimization algorithm tries to reach the clinical goals as much as possible. The predicted dose is mimicked by minimizing the difference between predicted dose and deliverable doses of intermediate plans. Additionally, the optimizer tries to minimize a set of dosimetric objectives for the targets and the OARs.
The goal of DLAP stays the same as in manual planning: getting a sufficient amount of dose in the target, while sparing the OARs as much as possible.

In figure 3 an example patient is shown to gain more insight in the difference between the deliverable dose (left) and the predicted dose (right). This patient is bilaterally treated since both sides of the elective glands are receiving the low dose level. In this example of a bilateral irradiated oropharynx patient it is visible that the predicted dose is not deliverable, i.e. the holes in the dose distribution are physically unrealistic. Nevertheless, the holes are explainable since that is where this patient's oral cavity (see figure 1) and submandible glands (see figure 2a) are situated. The predicted dose tries to spare the OARs as much as possible. Therefore, it is the goal to obtain a deliverable dose as close to the predicted dose as possible.

Figure 3: Transversal view of a bilateral oropharynx patient: deliverable dose (left) vs. predicted dose (right). The color bar on the right side shows that the low dose level (DL1) in the elective glands (5425 cGy) is indicated with the color green and the high dose level (DL2) in the primary tumor (7000 cGy) is indicated with yellow.

RayStation provides a DLAP for oropharynx patient with the same training data for every radiotherapy institution. In practice there is a variation in treatment planning between different radiotherapy institutions. There are (inter)national guidelines, but particular preferences may differ among institutions. These differences can be due to personal preferences of the radiation oncologist or because of availability in (technical) equipment, imaging modalities and treatment machinery. Thus, RayStation provides the ability to have an institution specific DLAP model by tuning the initial model to the specific preferences. In the LUMC such institution specific model is available, which is created in collaboration with the technical support of RaySearch. RaySearch tunes the model with a few treatment plans of patients previous treated at the LUMC. When the tuning is done, the tuning is validated by the involved medical staff of the LUMC by evaluating the DLAP plans of thirteen patients according to their standards and protocols. If the DLAP plans do not meet the requirements, the validation report is submitted to the technical support of RaySearch and they start a new tuning round. When the final tuning is conducted and approved by the medical staff, the DLAP can be used for automatically creating radiation treatment plans.

For additional information on DLAP, the white paper conducted by RaySearch in collaboration with the University Medical Center Groningen (UMCG) is included in the appendix, specifically in section 7.6.

3 Method

The research question addressed in this master's thesis is focused on investigating how RayStation's DLAP can enhance the quality, consistency, and efficiency of treatment planning for photon beam radiotherapy in patients with HNC. To study the research question, a plan comparison study is conducted at the radiotherapy department of the LUMC. The study uses resources provided by the LUMC, such as TPS RayStation, patient data, and treatment planning protocols. To ensure privacy and confidentiality, all patient information and the clinical plans are anonymized prior to generating DLAP plans. The DLAP plans are exclusively generated using the DLAP model developed by RaySearch.

3.1 Patient Cohort

This thesis investigates DLAP in radiotherapy treatment through the analysis of oropharynx patients previously treated at the LUMC. The cohort of anonymized patients is chosen based on tumor and treatment specific inclusion and exclusion criteria, which are listed in table 2.

Table 2: Inclusion and exclusion criteria for patient selection

Inclusion criteria	**Exclusion criteria**
Oropharynx (cohort 1)	No radiotherapy planning available and administered
Hypopharynx and larynx (cohort 2)	Patient referred to proton center
Unilateral (cohort 3)	Missing (too many) required ROIs
Two dose levels (54.25 Gy & 70 Gy)	Other dose level(s)
35 fractions	Other fractionation scheme
TPS: RayStation	Deprecated TPS
	Other tumor sites (nasopharynx, mouth, tongue)

The first cohort of anonymized patients contains 43 oropharynx patients. To ensure a comprehensive treatment plan comparison as desired in this study, it is crucial to include a plan that was clinically accepted for treatment. Consequently, patients who did not receive treatment or were referred to a proton center are excluded from this study, since they solely have a planning CT scan available. Additionally, patients are excluded from the study if a significant number of required ROIs are missing from the structure data set. As said before, there are required and optional ROIs associated with DLAP (see table with associated ROIs in the appendix, section 7.5). The accurate segmentation of certain structures cannot be done without the expertise of a radiation oncologist or an RTT. Since specific structures are mandatory for DLAP, it becomes impossible to generate a treatment plan for these patients without additional segmentation and therefore are not included. An example is that the DLAP needs the ROI larynx as structure to create a treatment plan. In clinical setting, when patients are treated for a larynx tumor it is possible that the tumor covers a large part of the larynx. In that case, the radiation oncologist could decide that it is not necessary to segment the larynx as OAR, since it will receive most of the dose anyway and therefore will not meet the clinical goal for the ROI larynx. In clinical setting this does not matter for manually planning a treatment plan, but DLAP cannot generate a plan when one of the required ROIs is missing.

Due to a transition in TPS at the LUMC a few years ago, only patients with available clinical plans in RayStation are included. Consequently, only patients treated from 2021 until 2023 are involved in this study. It is possible to include older patients from the previous TPS (Pinnacle) since the planning strategy and class solution are similar, but it is not identical. To reduce the chances that including the Pinnacle patients potentially lead to bias, only RayStation patients are concidered in this study.

3.1.1 Expanding Patient Cohort

After the comparison between the clinical plan and DLAP generated plans for patients with oropharynx tumors, the patient cohort is expanded to investigate the impact of DLAP on hypopharynx, larynx, and unilateral oropharynx patients. Figure 1 displays the location of the new proposed tumor sites. The inclusion of these cohorts results in the addition of eleven, eight, and nine patients, respectively. Hypopharynx and larynx tumors are located inferior to the oropharynx, with OARs that are relatively similar to those in oropharynx tumors. On the contrary, nasopharynx patients are excluded from this study since nasopharynx tumors, located superior to oropharynx tumors, are surrounded by OARs that are not considered in the DLAP. For instance, although the optic nerve is not among the required ROIs for DLAP, it is important to spare this nerve during treatment.
In HNC radiotherapy treatment, there are cases where the target tissue is irradiated unilaterally, rather than bilaterally. This means that the elective glands are irradiated only on the side where the primary tumor is located. The DLAP used in this study is primarily trained on treatment plans that involve bilateral irradiation. Therefore, the expansion of the patient cohort with unilateral patients aims to investigate the behavior of DLAP when applied to this specific patient group and to explore the obtained outcomes. Similarly, why this study includes patients with hypopharynx and larynx tumors, even though DLAP is not specifically trained on these tumor sites. This expansion is justified by the proximity of hypopharynx and larynx tumors to oropharynx tumors, as well as the presence of similar OARs in these tumor regions. The aim is to gain insights into the performance and effectiveness of DLAP in these specific clinical scenarios. Although the model is used outside the inclusion criteria of the training data, it is possible that the model still performs well for hypopharynx, larynx and unilateral cases due to the similarity in tumor and OAR regions. This similarity could potentially result in significant benefits for these target areas when using DLAP.

3.2 Treatment Planning by RayStation Deep Learning Autoplanning

All DLAP plans are generated in RayStation (version 10b) based on the same CT images and nearly all the same ROIs as the clinical plan. The clinical and DLAP plans have two prescribed dose levels of 70 Gy and 54.25 Gy, delivered in 35 fractions. The VMAT plans are generated using two full 360° arcs. The treatment machine is an Electa Agility. The DLAP plans are generated with the DLAP tool in RayStation and are not optimized further by an RTT. The DLAP model that is used is version: RSL-Oropharynx-7000-SIB (3.0.0). The model was validated for the DAHANCA 2020 protocol on a Elekta Versa machine using 6MV, setup with 2 full arcs with 2-degree gantry spacing [29]. The validation report, model protocol and model overview is included in the appendix.
Clinical plans that are referred to throughout this thesis were clinically accepted and used in the actual treatment of the patient.

The completion of the DLAP process is followed by a review of the generated DLAP plan. Although this evaluation lacks the scientific value of a radiation oncologist or clinical physicist, it gains insight and serves as an initial assessment of irregularities within the DLAP plans. For example, inclusion of tissue equivalent build-up material (WEM). WEM is used to place the dose build-up area in this material, whenever the skin is to be included in the target area and should receive the prescribed dose. If it is part of the target area in the clinical plan, then it is also applied in the DLAP. The primary objective of this reviewing is to identify areas for discussion with experts in order to enhance understanding of the underlying processes and results.

3.3 Data Analysis Methods

Providing an answer to the research question indicates the need for analysing methods to compare DLAP plans with clinical plans. To validate if DLAP improves the quality and consistency of radiotherapy treatment planning, the DLAP and clinical plan are compared on dosimetric parameters and normal tissue complication probability (NTCP) outcomes. The dosimetric analysis, parameters and the NTCP analysis are elaborated on in the following subsections.

3.3.1 Dosimetric Analysis

Dosimetric analysis involves evaluating the distribution of radiation dose in both the target tissue and OARs. To compare the dosimetry between the clinical plan and the DLAP, information regarding target coverage and dose distribution in OARs is obtained from the dose volume histograms (DVH). DVHs provide a graphical representation of the dose received within a ROI. An example DVH is illustrated in figure 4. The DVH parameters evaluated in this study are identical to the parameters used in clinical practice to evaluate treatment plans.

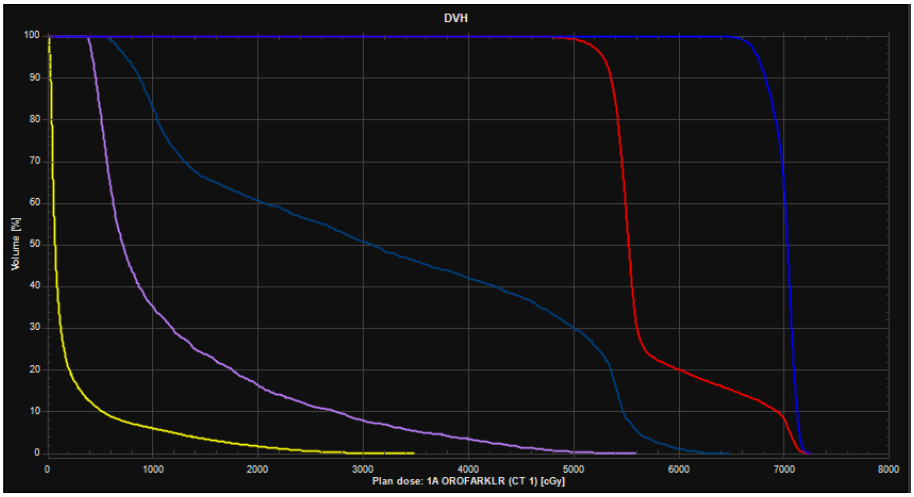


Figure 4: Example of a DVH displaying the brain (yellow), larynx (purple), PCM superior (dark blue), PTV DL1 (red), and PTV DL2 (blue). This DVH is obtained from the clinical plan of one of the anonymized unilateral oropharynx patients.

In this example, the DVH presents the dose volume relationship of two target structures (represented by red and blue curves) and three OAR structures (represented by yellow, purple, and dark blue curves). The DVH for target structures shows that as much volume as possible receives the desired dose, resulting in a straight line at the top of the DVH plot. Conversely, for the OARs, the goal is to minimize the volume receiving high doses, which is reflected by the lower curves on the DVH plot. To maintain clarity, only a few structures are displayed in the DVH graph of figure 4.

DVHs provide information about the dose distribution within the target structures and the mean and maximum dose values for the OARs. Throughout this thesis, target coverage and OAR dose are presented using boxplots generated by the computational software Matlab (version: 2021a). The data to obtain the boxplots is extracted from the DVHs. The boxplots visualize the median, spread, and significant differences, calculated using the Wilcoxon signed-rank test, providing a comprehensive representation of the dosimetric comparison between the clinical plan and the DLAP plan.

3.3.2 Clinical Goals of Treatment Planning

For the dosimetric analysis the plans are compared on dosimetric parameters and according to the clinical goals. The clinical goals that are used for the plans are shown in table 3, 4 and 5, obtained from the LUMC protocol for treatment planning in HNC. Since HNC patients are prescribed with two dose levels, there will be a separation in PTV of DL1 (54.25 Gy) and DL2 (70 Gy). The dose coverage of PTV DL1 refers to the percentage of target tissue that receives the prescribed dose of radiation, i.e. V95% $\geq$ 98% means the volume receiving 95% of the prescribed dose should be equal to or exceed 98% of the total target volume. The second constraint for PTV DL1 is that the mean dose should receive $\leq$ 102% of the prescribed dose. For PTV DL2 the same constraints are applicable, as can be seen in table 3, and an additional constraint that the dose received by 0.03cc of the volume within PTV DL2 should not exceed 107% of the prescribed dose. The naming convention in table 3 will correspond with the boxplots in the results part of this thesis, section 4.

Table 3: Clinical goals for target ROIs for HNC in LUMC

ROI	Item goal	Constraint
PTV DL1	$V_{95\%}$	$\geq 98\%$
	Mean dose	$\leq 102\%$
PTV DL2	$V_{95\%}$	$\geq 98\%$
	$D_{0.03cc}$	$\leq 107\%$
	Mean dose	$\leq 102\%$

Prior to addressing the clinical goals of OARs depicted in table 4, it is essential to comprehend their relevance. Each OAR will be clarified in sequence, following the same order as in table 4 (top to bottom) which will be the same order as the x-axis of the upcomming boxplots (from left to right).

Table 4: Clinical goals for OARs for HNC in LUMC

OAR	Item goal	Goal
Parotids ipsi-lateral	Mean dose	≤ 28 Gy
Parotids contra-lateral	Mean dose	≤ 17 Gy
Parotids contra-lateral*	Mean dose	≤ 5 Gy
Submandible glands	Mean dose	≤ 35 Gy
Constrictor muscles	Mean dose	≤ 40 Gy
Glottic area	Mean dose	≤ 40 Gy
Larynx	Mean dose	≤ 40 Gy
Cricopharyngeus	Mean dose	≤ 40 Gy
Oral cavity	Mean dose	≤ 28 Gy
Mandible	$D_{2\%}$	≤ 50 Gy
Mandible-PTV	$D_{2\%}$	≤ 40 Gy

The parotids, parotid glands, are major salivary glands and are bilateral located posterior to the mandible and anterior to the ear [30]. The parotid glands are referred to in the boxplot as *parotid ipsi* (ipsi-lateral) and *parotid contra* (contra-lateral). Ipsi-lateral indicates the parotid is on the same side as the primary tumor, while contra-lateral indicates the opposite side. Therefore, the dose constraints are different. The ipsi-lateral side is allowed to receive more dose since this is in close proximity to the primary tumor compared to the contra-lateral side. The contra-lateral parotid indicated with a star (*) in table 4 refers to the dose constraint of the contra-lateral parotid gland in the context of unilateral treatment.
Similar to the parotid glands the submandibular glands have an ipsi-lateral and contra-lateral gland, indicated as *SMG ipsi* and *SMG contra* in the boxplot. The submandibular glands are located inferior to the mandible (lower jawbone) and are an import saliva producing OAR. The anatomical location of the parotids and the submandible glands are displayed in figure 2a.
Subsequently, the PCMs are subdivided in three separate OARs; *PCM inf* (inferior), *PCM med* (medius/middle) and *PCM sup* (superior). For all three PCMs the same dose constraint applies. The anatomical location is displayed in figure 2b.

The *glottic area* (displayed in figure 1) is the middle part of the larynx and this is where the vocal cords are located. The part of the larynx inferior to the glottic area is indicated as *larynx SG* (sub glottis) and consists of cartilage rings. This OAR plays a crucial role in regulating air flow while breathing and speaking. The *cricopharyngeus* is located inferior to the larynx and is the circular muscle at the beginning of the esophagus. The cricopharyngeus regulates food and liquids entering the esophagus that passed through the pharynx. The *oral cavity* is the mouth and, finally, the *mandible* (as stated before) is the lower jawbone. The mandible is represented by two ROIs in table 4 and in the boxplots. One for the $D_{2\%}$, meaning that the dose received by 2% of the mandible volume should be kept below 50 Gy. This type of constraint is used to ensure that the dose delivered does not exceed a certain threshold. The other ROI is where the PTV structure is subtracted from the mandible structure, indicated as *mandible-PTV*. The later ROI should be kept below 40 Gy.

Apart from the previous named OARs the treatment plans are also evaluated on OARs that are depicted in table 5.

Table 5: Clinical goals for OARs for HNC in LUMC

OAR	Item goal	Goal
Brain	$D_{2\%}$	≤ 70 Gy
	$D_{0.03cc}$	≤ 65 Gy
brain stem core	$D_{0.03cc}$	≤ 54 Gy
Spinal cord	$D_{0.03cc}$	≤ 50 Gy
Spinal cord+3mm	$D_{0.03cc}$	≤ 52 Gy
Cochlea ipsi-lateral	Mean dose	≤ 45 Gy
Cochlea contra-lateral	Mean dose	≤ 45 Gy
Lenses	$D_{0.03cc}$	≤ 6 Gy

In table 5 the OARs related to the nervous system are shown. The brain has two criteria, one is the *Brain* $D_{2\%}$ and the other is *Brain* $D_{0.03cc}$. The former means that the dose received by 2% of the brain volume should be kept below 70 Gy and the later means that that the dose received by a volume of 0.03 cc within the brain should be kept below 65 Gy. For the brain stem core and spinal cord similar type of dose constraint are applicable. The brain stem core is the middle part of the brain stem. The brain stem is the part of the brain that connects the brain to the spinal cord. The spinal cord is the bundle of nerves that continues from the brain through the spine to rest of the human body [30]. The spinal cord + 3 mm means that the ROI spinal cord is expanded with 3 mm margin. This expanded ROI can receive up to 2 Gy more that the original ROI.
The cochlea is a part of the inner ear. Similar to other ipsi- and contra-lateral structures it depends on which side the primary tumor is situated. Nevertheless, the mean dose for both ipsi- and contra-lateral is the same. Finally the lenses of the eye are set to a maximum dose of 6 Gy.

3.3.3 NTCP Analysis

In radiotherapy the aim is to spare the OARs as much as possible to limit negative side effects. NTCP models are designed to predict the chance of radiation induced side effects [26, 31]. For HNC both xerostomia and dysphagia can be predicted with NTCP models for a grade ≥ 2 and ≥ 3 toxicities. Clinically the NTCP models are used to predict if a patient benefits from proton therapy instead of photon therapy by lowering predicted side effect probabilities. The models are developed by the Dutch National Indication Protocol for Proton Therapy (in Dutch: Landelijk Indidactie Protocol Protonentherapie) [26]. Patients might be referred to a proton therapy center based on the outcome of the NTCP models. When the gain in NTCP for xerostomia or dysphagia grade ≥ 2 toxicities is $\geq 10\%$ or the gain for grade ≥ 3 toxicities is $\geq 5\%$ the patient is applicable for proton therapy. Stated by the Dutch National Indication protocol, xerostomia grade ≥ 2 toxicities indicated that the patient suffers from mild symptoms that cause difficulties eating, e.g. the patient needs a glass water or other lubricant while eating. Xerostomia grade ≥ 3 toxicities indicate that eating is not possible and the patient is depended on tube feeding.
Dysphagia grade ≥ 2 toxicities indicate that the patient cannot longer eat solid food, e.g. eating is only possible when the food is pureed or very soft. Dysphagia grade ≥ 3 toxicities indicate that the patient can only consume liquids or is depended on tube feeding.

In this study the same NTCP models are used to compare NTCP outcome of the clinical plan with the DLAP to indicated if DLAP improves or worsen the NTCP. This models are based on patient, tumor and OAR specific parameters and is given by

$$NTCP = (1 + e^{-S})^{-1} \tag{1}$$

where S is different for xerostomia grade ≥ 2 and ≥ 3 toxicities and for dysphagia grade ≥ 2 grade ≥ 3 toxicities. The equations for S also depends on the NTCP model that is being used. If a patient is

primarily treated with radiotherapy (primary setting) the NTCP model differs from when the patient had surgery before radiotherapy (postoperative setting). Since the aim is to investigate the difference in NTCP outcome between the DLAP plan and the clinical only one NTCP model is used, namely the model for the primary setting. Meaning that the assumption is made that all patient were primary treated with radiotherapy and did not undergo surgery. Furthermore, the calculation assumes a baseline score. The baseline score is based on the severity of xerostomia and dysphagia before treatment. For xerostomia this can be none, little and severe. And for dysphagia the baseline can be grade 0-1, grade 2 or grade 3+. For all patients the baseline is set to 'none' for xerostomia and 'grade 0-1' for dysphagia. These baselines values correspond in both cases with 0.

This way all patients are set to the same settings and parameters to enhance a comparison between the NTCP outcomes of the two different treatment plans. The actual NTCP outcome of the plans could differ because the patient and tumor specific parameters are lost in anonymization. However, this is not relevant for the research purpose.
The value of S for xerostomia grade ≥ 2 toxicities in primary setting corresponds to

$$\begin{aligned} S = -2.2951 + 0.0996 * (\sqrt{D_{mean}\text{Parotid ipsi-lateral}} + \sqrt{D_{mean}\text{Parotid contra-lateral}}) \\ +0.0182 * (D_{mean} \text{ both submandible glands}) \\ +\text{baseline score} \end{aligned} \tag{2}$$

The value of S for xerostomia grade ≥ 3 toxicities in primary setting corresponds to

$$\begin{aligned} S = -3.7286 + 0.0855 * (\sqrt{D_{mean}\text{Parotid ipsi-lateral}} + \sqrt{D_{mean}\text{Parotid contra-lateral}}) \\ +0.0156 * (D_{mean} \text{ both submandible glands}) \\ +\text{baseline score} \end{aligned} \tag{3}$$

NTCP models for dysphagia have an additional parameter, namely the primary tumor location. The values to the corresponding tumor location for dysphagia grade ≥ 2 and grade ≥ 3 toxicities in primary setting are shown in table 6. For the NTCP calculations of treatment plans of bilateral oropharynx, unilateral oropharynx and hypopharynx tumors the tumor location is set to pharynx. For the larynx patients the tumor location is set to larynx.

Table 6: NTCP parameter: tumor location, for dysphagia grade ≥ 2 and grade ≥ 3 toxicities in primary setting

Tumor location	**Dysphagia grade** ≥ 2	**Dysphagia grade** ≥ 3
Oral cavity	0.0000	0.0000
Pharynx	-0.6281	0.0387
Larynx	-0.7711	-0.5303

The value of S for dysphagia grade ≥ 2 toxicities in primary setting corresponds to

$$\begin{aligned} S = -4.0536 + 0.0300 * D_{mean}\text{Oral cavity} \\ +0.0236 * (D_{mean}\text{PCM superior}) \\ +0.0095 * (D_{mean}\text{PCM middle}) \\ +0.0133 * (D_{mean}\text{PCM inferior}) \\ +\text{baseline score} \\ +\text{tumor location} \end{aligned} \tag{4}$$

The value of S for dysphagia grade ≥ 3 toxicities in primary setting corresponds to

$$\begin{aligned} S = -7.6174 + 0.0259 * D_{mean}\text{Oral cavity} \\ +0.0203 * (D_{mean}\text{PCM superior}) \\ +0.0303 * (D_{mean}\text{PCM middle}) \\ +0.0341 * (D_{mean}\text{PCM inferior}) \\ +\text{baseline score} \\ +\text{tumor location} \end{aligned} \tag{5}$$

For each treatment plan, all four NTCP models are employed. Consequently, there will be four NTCP outcomes for both the clinical plan and the DLAP plan for every patient. The NTCP outcome is presented as a percentage, and the primary focus lies in determining the difference in percentage between the clinical plan and the DLAP plan. The question is whether the NTCP improves or worsens with the use of DLAP.

4 Results

In this study, the oropharynx model of RayStation's DLAP software is used to generate automated treatment plans. The institution specific DLAP, which was developed by RaySearch in collaboration with the LUMC, undergoes tuning rounds to determine the optimal settings. Before comparing the clinical plans with the DLAP plans, the selection of the optimal tuning round is addressed in a sub-study, which covers the initial part of the results. Following the sub-study, the three patient cohorts will be individually addressed and analyzed.

4.1 Sub-study: Tuning Analysis

In the LUMC a dedicated team of medical and physical experts is involved is evaluating the performance of the DLAP tunings. Evaluation of the fifth tuning revealed that this tuning was very close to desirable, but a problem occurred in PTV coverage. At the level of the swallowing muscles the PTV coverage was sub-optimal since the sparing of the swallowing muscles caused a local underdosage in the PTV. RaySearch tried to solve this problem in the DLAP model and this resulted in a few more tuning rounds that followed.
When RaySearch introduces a new tuning, it involves modifying the model settings file of the DLAP model. Specifically, prediction settings or the weights assigned to ROI objectives in the settings file are adjusted to achieve the desired outcome. Another part of the model settings file that might undergoes changes is the mimick settings, leading to a different dose mimicking approach. By changing the parameters, the intention is to improve the treatment plan's ability to address the underdosage problem in the PTV. The changes that were made in the model settings file are relatable to what can be observed in the deliverable dose of the DLAP plans.
The underdosage problem in the PTV did not get better in tuning six and seven. Tuning eight followed and when comparing the model settings file of tuning five with tuning eight, it shows an increase in PTV objectives. Furthermore, it shows a doubling of the weight of one of the PCMs. All though tuning eight seemed to be close to clinically acceptable, there was still a slight underdosage noticeable in the PTV around to the swallowing muscles. Therefore, a work around was designed by the project team of the LUMC.

To resolve the underdosage, the PTV is locally expanded with 2 mm in medial direction at the level of the swallowing muscles. To create this local expansion an additional ROI is created by an intersection of the swallowing muscles and the PTV DL1 to segment the overlapping region between the two ROIs. In this overlapping region the underdosage is occurring. This is shown in figure 5 by the green colored segmentation.

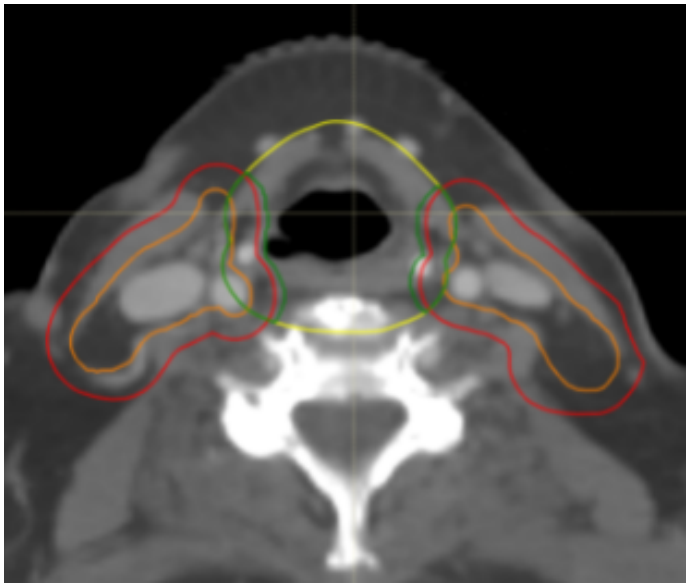
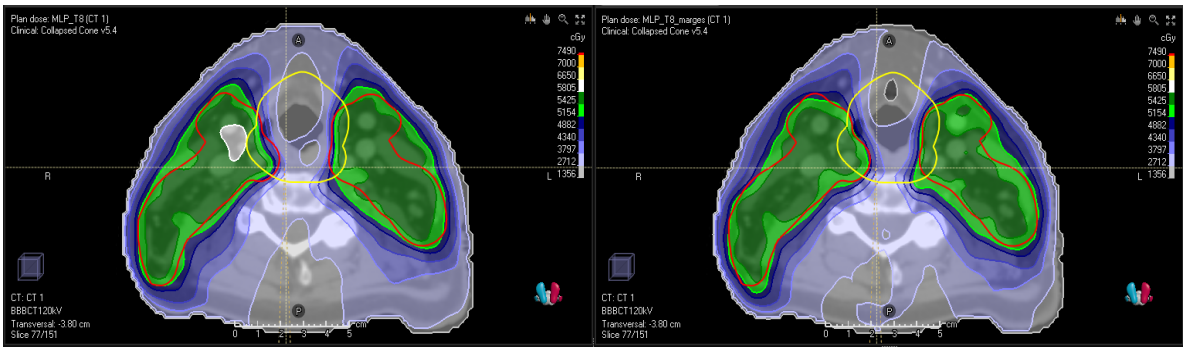


Figure 5: Transversal view of bilateral oropharynx patient. Clinical target volume of the low dose level (CTV DL1) is indicated with orange, PTV DL1 with red, the combined swallowing muscles with yellow and the overlapping region with green.

In figure 5 CTV DL1 (orange) is expanded with 5 mm all around the structure to form PTV DL1 (red). The individual PCMs, cricopharyngeus, larynx, glottic area and esophagus are all combined into one structure (yellow). When creating an intersection of this combined swallowing muscles structure an the PTV DL1, the overlapping region is left. Only the overlapping region is then expanded with 2 mm in medial direction. Following the medial expansion, the green colored overlapping region is combined with the original PTV DL1 to form a new PTV DL1. The new PTV DL1 is used as input for the DLAP. The old PTV DL1 is used for comparison and analyzing purposes, because the old PTV DL1 corresponds to the structure used in the clinical plan.
The effect of the PTV expansion on the treatment plan is displayed in figure 6. The aim is to fully cover PTV DL1 with green colored dose. Preferably, the light green dose nicely follows the contour of the PTV DL1. In 6a at the intersection of the swallowing muscles with the PTV DL1 an underdosage occurs in medial direction.



(a) Tuning eight (b) Tuning eight with local PTV DL1 expansion

Figure 6: Transveral view of bilateral oropharynx patient. DLAP plan of the same patient with tuning eight with and without local PTV DL1 expansion. The red segmentation indicates PTV DL1 and the yellow segmentation indicates the combined swallowing muscles and associated structures.

When comparing figure 6a with figure 6b is it visible that the underdosage at the intersection is resolved. In figure 6b the green colored dose now fully covers PTV DL1 and there is no underdosage any more.

To show the difference in dose distribution in the DLAP with and without local PTV DL1 expansion more clearly, the DLAP plans are subtracted from one another in RayStation. This leads to a difference in dose distribution that is shown in figure 7.

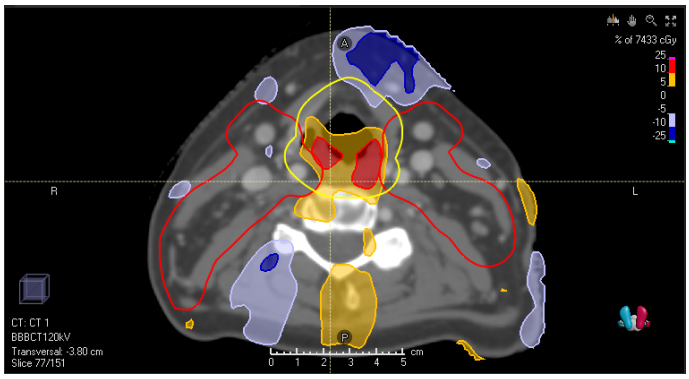


Figure 7: Transveral view of bilateral oropharynx patient. Difference in dose distribution when DLAP tuning eight with and without local PTV DL1 expansion are subtracted.

The dose difference is indicated as a percentage difference with respect to the maximum dose (7410 cGy), meaning that in this case the orange and red color means that the dose in the DLAP plan of tuning eight with local PTV DL1 expansion receives 5% and 10% more dose in that area, respectively. Figure 7 makes it clearly visible that at the intersection of the swallowing muscles and the PTV DL1 a higher dose is received when using the local PTV DL1 expansion. A similar trend in observable for tuning five with the local PTV DL1 expansion.

By creating this local PTV DL1 expansion, the underdosage is resolved in both tuning five and eight. The outstanding question now is: which tuning offers a better performance - tuning five or eight? To address this question a small sub-study is conducted to determine the optimal tuning of the DLAP that will be used for the remainder of the thesis.

4.1.1 Sub-study: Dosimetric Analysis

Comparing DLAP tuning five and eight to their corresponding clinical plans is done by dosimetric, NTCP and visual analysis of thirteen oropharynx patients. First, all thirteen patients needed the local PTV DL1 expansion. This is manually done in RayStation by creating the new ROI, as explained before. Subsequently, for all thirteen patients new DLAP plans are created for both tuning five and eight. This resulted in generating a total of 26 new treatment plans. Finally, evaluation of tuning five and eight is conducted by comparing the new DLAP plans to each other and to the clinical plan.

The first comparison is done by evaluating the DVH parameters from the clinical plan, the DLAP tuning five and the DLAP tuning eight. The results are graphically depicted in boxplots that are shown in figure 8, 9 and 10.

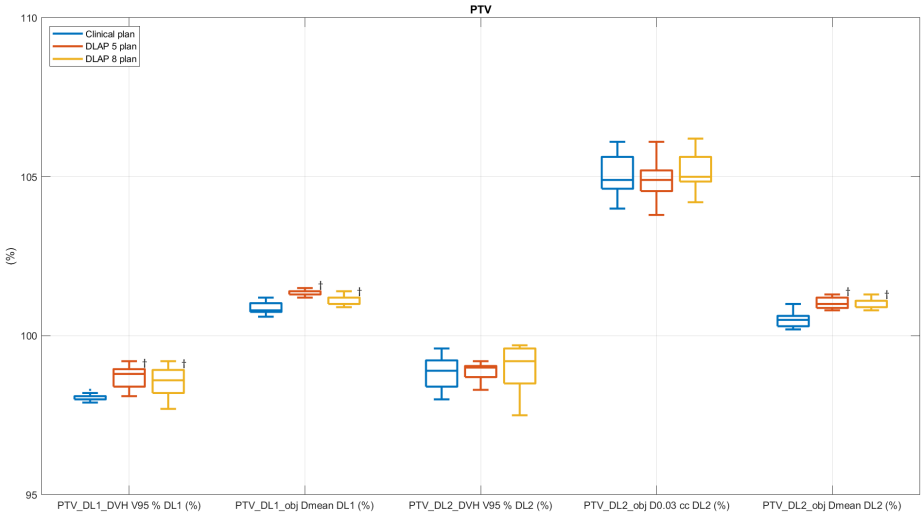


Figure 8: Dosimetric parameters of the targets in the clinical plan (blue), DLAP tuning five (orange) and DLAP tuning eight (yellow).

Figure 8 shows that target coverage is significantly higher (indicated with a cross) in DLAP tuning five and eight compared to the clinical plan for PTV DL1 $V_{95\%}$, PTV DL1 D_{mean} and PTV DL2 D_{mean}. This is explainable because in clinical setting the RTTs strive for a coverage of 98% and if the coverage is $> 98\%$ the RTTs try to lower the coverage so that the surrounding OARs will also receive a lower dose. This is observable through the small spread, i.e. height of the box, in the clinical PTV DL1 $V_{95\%}$ in figure 8. In the DLAP this is not the case and that is why the coverage in the DLAP plans is higher than in the clinical plans.

The DVH parameters of the OARs are shown in figure 9, to evaluate what the higher coverage and the DLAP in general does to the OARs.

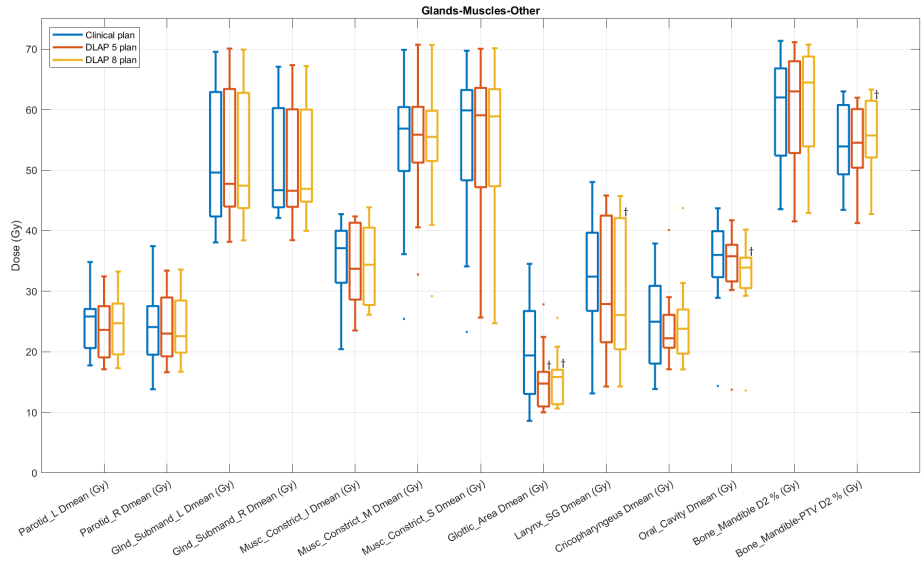


Figure 9: Dosimetric parameters of the OARs in the clinical plan (blue), DLAP tuning five (orange) and DLAP tuning eight (yellow).

The boxplot in figure 9 shows the dosimetric parameters of the OARs and show a significantly better sparing of the glottic area in DLAP tuning five and eight and a significantly better sparing of the oral cavity in DLAP tuning eight compared to the clinical plan. In general, both tunings of DLAP demonstrate improved or comparable sparing of OARs compared to clinical plans. However, it is noteworthy that DLAP tuning five and eight show an increase in the dosimetric parameter for the mandible. For DLAP tuning eight this is even a significant increase. It is important to consider that the mandible is not a structure as critical as others, and in clinical practice it is not always segmented or prioritized for dose optimization.

Figure 10 illustrates additional OARs that are considered for evaluation.

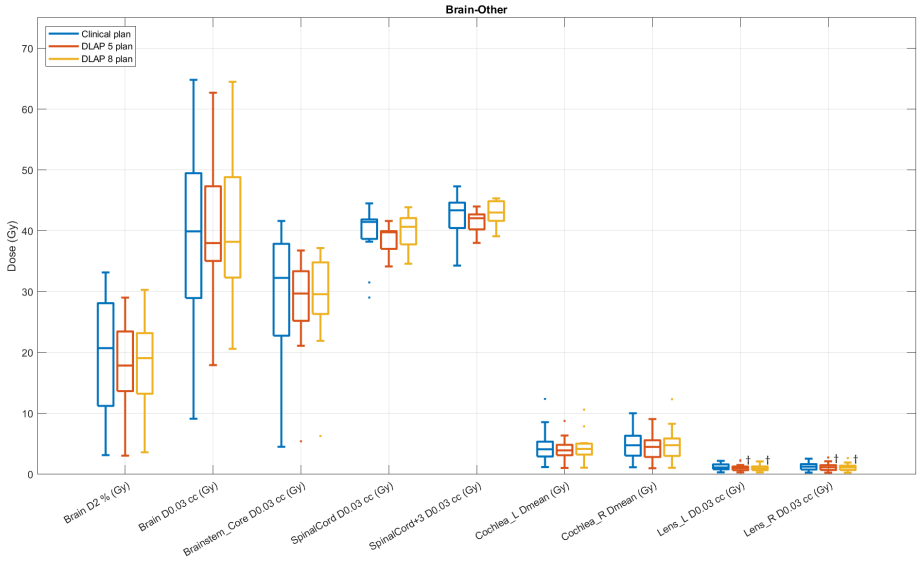


Figure 10: Dosimetric parameters of the OARs in the clinical plan (blue), DLAP tuning five (orange) and DLAP tuning eight (yellow).

The boxplot analysis of OARs related to the nervous system, as depicted in figure 10, only reveals a significant difference in the dose administered to the lenses. However, no significant differences are observed in the other OARs. Nevertheless, it is worth noting that there is less variability in the brain and brain stem core for both DLAP tunings compared to the clinical plan. Additionally, DLAP tuning five demonstrates a lower median dose in the brain and spinal cord regions compared to tuning eight and the clinical plan. Overall, the dosimetric parameters for these OARs exhibit similarity across all three treatment plans.

4.1.2 Sub-study: NTCP Analysis

The NTCP models for dysphagia and xerostomia are applied to all thirteen patients in this sub-study. Each NTCP model generates two percentages, one corresponding to grade ≥ 2 toxicities and the other to grade ≥ 3 toxicities. The outcomes of these NTCP models are graphically presented in bar graphs, as shown in figures 11 and 12.

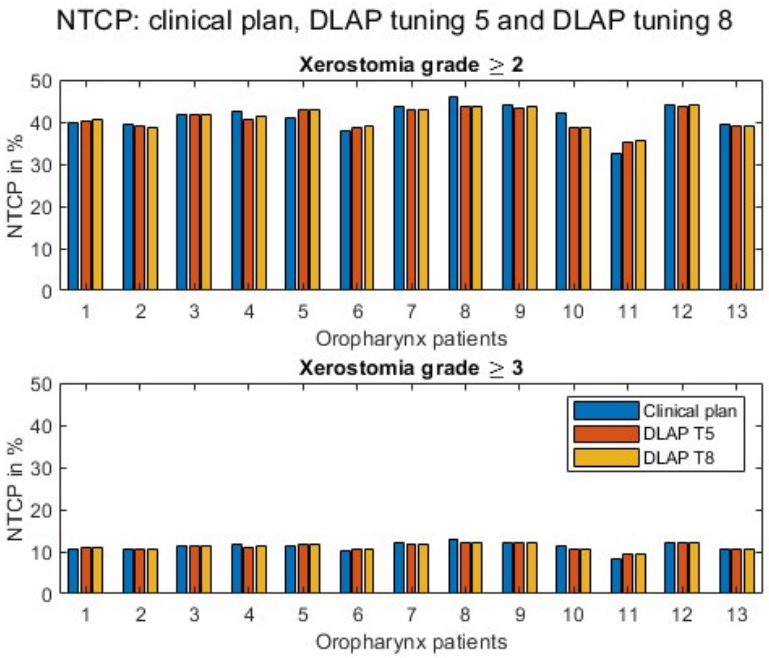


Figure 11: NTCP outcome for xerostomia grade ≥ 2 and 3 toxicities, clinical plan (blue), DLAP tuning five (orange) and DLAP tuning eight (yellow).

First the NTCP models for xerostomia are evaluated, as shown in figure 11. Here, both DLAP tunings demonstrated a slight decrease in terms of xerostomia compared to the clinical plan. This decrease is present in eight out of the thirteen patient for xerostomie grade ≥ 2 toxicities. However, the difference between DLAP tuning five and DLAP tuning eight is very small. For xerostomia the preference leans towards DLAP tuning five with an average decrease of 0.2% for grade ≥ 2 toxicities and an average decrease of 0.1% for grade ≥ 3 toxicities. In clinical practice, an improvement of more than 10% in the NTCP outcome for grade ≥ 2 toxicities is considered beneficial for the patient. In the case of xerostomia, the observed improvements are relatively small. Consequently, the NTCP outcomes suggest no significant increase or decrease between the two DLAP tunings or between the DLAP tunings and the clinical plan. However, it is not possible to draw a binding conclusion due to the limited patient group size that is used in this NTCP analysis.

In the following bargraph, displayed in figure 12, the NTCP outcome for dysphagia is graphically shown.

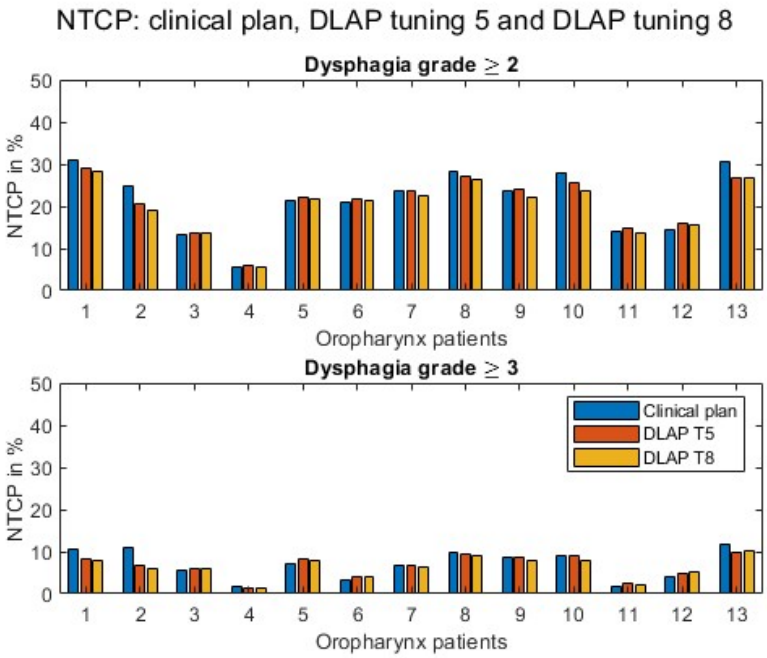


Figure 12: NTCP outcome for dysphagia grade ≥ 2 and 3 toxicities, clinical plan (blue), DLAP tuning five (orange) and DLAP tuning eight (yellow).

Similar to the xerostomia bargraph there is a decrease in NTCP noticeable in figure 12 in both DLAP tunings compared to the clinical plan. But in this case, for dysphagia the preference leans towards tuning eight compared to DLAP tuning five with an average decrease of 0.8% for grade ≥ 2 toxicities and 0.3% for grade ≥ 3 toxicities. Particularly for dysphagia grade ≥ 3 toxicities, a notable difference in NTCP was observed when comparing both DLAP tunings to the clinical plan. Three patients show an improvement of more than 2% compared to the clinical plan, with one outlier displaying a difference of 6%. With this outlier the threshold of 5% difference is exceeded, indicating a significant improvement of treatment outcome for this patient in terms of dysphagia grade ≥ 3 toxicities.
Due to the limited size of the patient cohort, the NTCP analysis will be continued using the larger cohort of oropharynx patients. This additional analysis aims to determine if the outlier noticed in figure 12 is due to patient-specific parameters or if there is a trend emerging from the DLAP NTCP outcomes compared to the clinical data.

Concluding this sub-study, DLAP with tuning eight is selected for further investigation in this study, because the target coverage, sparing of OARs and the NTCP outcome were preferable compared to tuning five. Furthermore, radiation oncologist and expert RTTs involved in the clinical introduction of HNC DLAP visually analysed all treatment plans and preferred tuning eight over tuning five and the clinical plan in eight out of thirteen cases.
Tuning eight, including the local PTV DL1 expansion, has also been selected for clinical implementation in the LUMC.

4.2 Patient Cohort 1: Oropharynx

The first results of the comparison study include a cohort of 43 patients with oropharynx tumors who are previously treated at the LUMC. The prescribed dose and fractionation scheme is consistent for all patients, meaning that the low dose level (PTV DL1) is 54.25 Gy, the high dose level (PTV DL2) is 70 Gy and the treatment is delivered in 35 fractions. All patients are irradiated bilaterally.
The DLAP oropharynx model is trained with data from patients with this tumor location, dose levels and fractionation scheme (see table 1) and is therefor expected to generate adequate treatment plans.

4.2.1 Dosimetric Analysis

In the first boxplot, displayed in figure 13, the dosimetric parameters of the target volumes in the clinical and DLAP plan are depicted. Here is a significant difference noticeable in PTV coverage of PTV DL1 $V_{95\%}$, PTV DL1 D_{mean} and PTV DL2 D_{mean}.

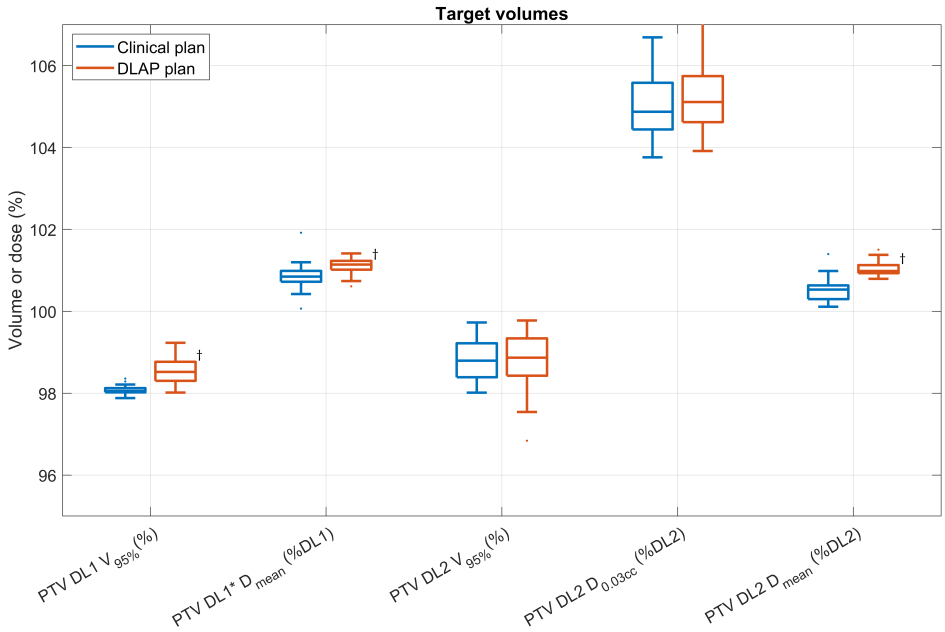


Figure 13: Dosimetric parameters of target structures from bilateral oropharynx patients from the clinical plan (blue) and DLAP plan (orange).

As previously explained in the sub-study, the DLAP achieves a significantly higher PTV coverage compared to the clinical plan. This trend is explainable by the fact that in clinical setting the RTT's steer the plan to a coverage of 98%. Meaning that if the coverage is higher than 98% the RTT's try to lower the coverage. The idea behind this is that a coverage $> 98\%$ can be decreased to spare other organs better. In the DLAP this is not the case and the DLAP tends to generate a plan with more dose in the target structures. Thus, the DLAP generates *hotter* treatment plans.

The dosimetric parameters of the clinical plans and the DLAP plans of the OARs associated with HNC are shown in the following boxplots, displayed in figure 14 and 15

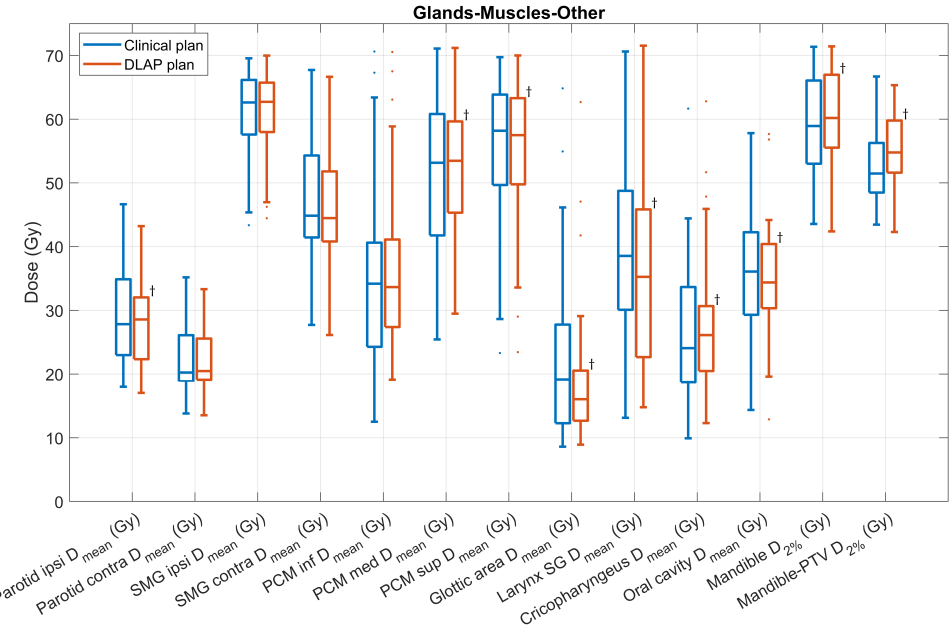


Figure 14: Dosimetric parameters of OARs (glands, muscles and others) from bilateral oropharynx patients from the clinical plan (blue) and DLAP plan (orange).

In regard to the OARs including the salivary glands, swallowing muscles, and other important structures, the DLAP treatment plan demonstrated overall similarity to the clinical plan. However, notable differences were observed in specific areas. There is a significant decrease of dose in the glottic area, larynx, and oral cavity in the DLAP compared to the clinical plan, potentially reducing the risk of complications in these regions. Conversely, higher median doses were observed in the cricopharyngeus and mandible. These variations imply that the DLAP prioritized certain areas differently. Higher dose in the cricopharyngeus could be explained by the purposely enlarged PTV DL1 at this structure since it is considered to be part of the combined swallowing muscles structure. Similar to the findings in the sub-study, the mandible exhibits a significantly higher dose in the DLAP plan. However, this outcome is of less concern for this specific ROI.

The following boxplot shows the dosimetric parameters of the OARs related to the nervous system. These are displayed in figure 15.

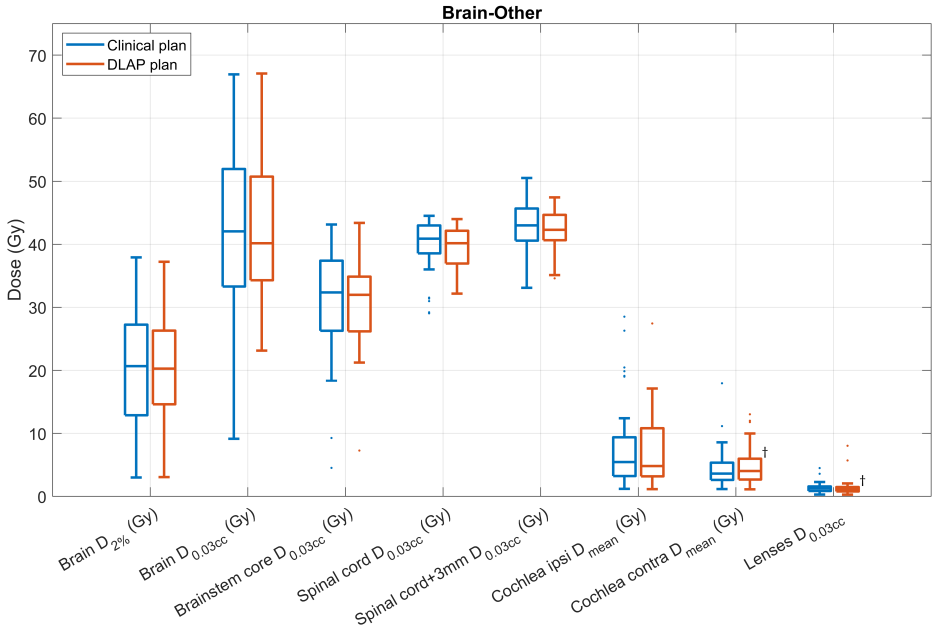


Figure 15: Dosimetric parameters of OARs (related to nervous system) from bilateral oropharynx patients from the clinical plan (blue) and DLAP plan (orange).

The comparison between the clinical plan and the DLAP plan revealed similar results for the nervous system related OARs. The doses delivered to the OARs in both plans were comparable, this finding suggests that the DLAP successfully achieved the clinical goals for these OARs, as shown in table 5. However, a significant difference is shown in the contra-lateral cochlea and the lenses, but do not exceed the dose constraints.

4.2.2 NTCP Analysis

The outcomes of the NTCP models are calculated for each plan individually. The outcome of each model results in a NTCP value (percentage) for the clinical plan and DLAP. The values are subtracted to calculate the difference (Δ NTCP) between NTCP of the clinical plan and the DLAP. The Δ NTCP for each patient and each NTCP model is plotted in figure 16 and 17. In these plots a positive value means the DLAP has a better NTCP outcome and a negative value means the clinical plan has a better NTCP outcome.

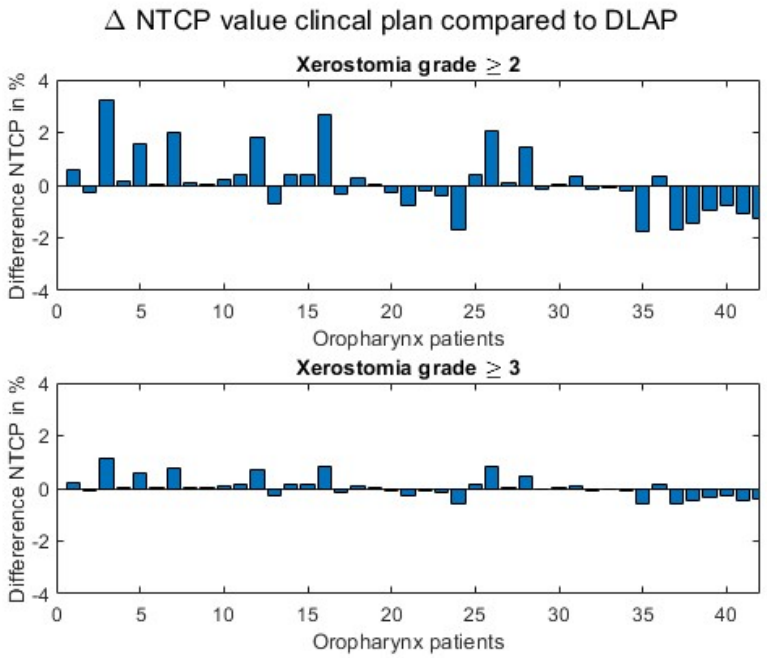


Figure 16: NTCP outcome for xerostomia grade ≥ 2 and 3 toxicities for bilateral oropharynx patients.

In clinical setting an improvement of 10% and 5% in NTCP outcome for grade ≥ 2 toxicities and grade ≥ 3 toxicities, respectively, indicates a sufficient gain for the patient. For comparison of the clinical plan with the DLAP plan the same threshold values are used. As can been seen in the plots of figure 16, there are no cases exceeding the threshold values. In fact, for xerostomia the maximum Δ NTCP for grade ≥ 2 toxicities is 3.23%.
The NTCP outcome favors one treatment plan over the other. For xerostomia the NTCP outcomes favors the DLAP plan over the clinical plan in 53% of the patients.
In figure 16 it is visible that this preference in some patients is so small, that for these patients the preference is negligible, especially for grade ≥ 3 toxicities. The NTCP analysis for xerostomia indicates a slight improvement in patient outcome when using DLAP, although this improvement is not exceeding the threshold values and therefore not decisive.

Similar findings hold for dysphagia grade ≥ 2 toxicities and grade ≥ 3 toxicities; there are no values exceeding the threshold values, as can be seen in figure 17.

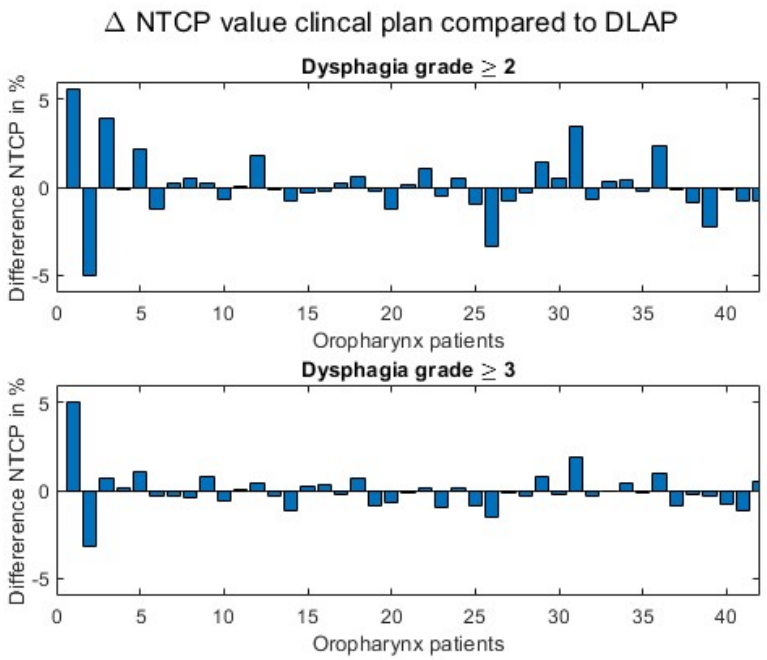
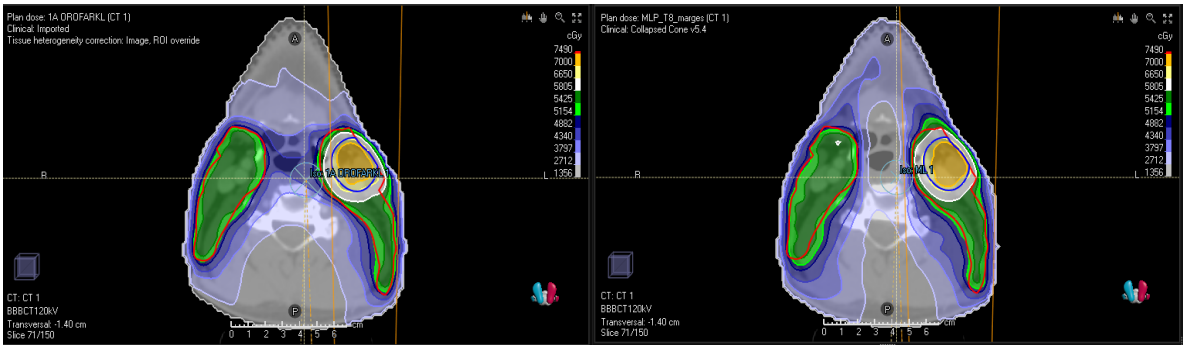


Figure 17: NTCP outcome for dysphagia grade ≥ 2 and 3 toxicities for bilateral oropharynx patients.

However, in figure 17 the first two patients show a relatively large NTCP difference between the clinical plan and the DLAP plan compared to other patients. To gain insight, the clinical and DLAP plans of this particular anonymized patients are further visually evaluated and compared.

In figure 18 the same transversal slice out of the clinical and DLAP plan are shown. For this patient the DLAP plan is preferred over the clinical plan based on NTCP outcome, and almost exceeds the NTCP threshold value of 5% for dysphagia grade ≥ 3 toxicities. In figure 18 the low dose level (PTV DL1) is indicated in a green color which corresponds to a dose level of 54.25 Gy. The high dose level (PTV DL2) is indicated with a yellow color which corresponds to a dose level of 70 Gy. This is where the primary oropharynx tumor is located. The segmentation of these structures are also visualized, with PTV DL1 in red and PTV DL2 in blue.



(a) Clinical (b) DLAP

Figure 18: Transversal view of bilateral oropharynx patient, where decrease in NTCP for dysphagia ≥ 2 and 3 toxicities is noticed

In figure 18a the clinical plan shows a continuous blue colored dose distribution between both sides of PTV DL1. This connection is where the PCM inferior is located, which indicated a dose level of approximately 38 to 49 Gy in these structures. In the DLAP plan, displayed in figure 18b, the PCM inferior is better spared due to the axial dose *gab* between the irradiated elective glands. The dose

distribution around the PCM inferior is now visualized with a white and light blue color, indication approximately 13.5 to 27 Gy. To represent the difference more clearly the clinical plan and DLAP plan are subtracted from each other and this is shown in figure 19. This is the same patient and transversal slice as shown in figure 18.

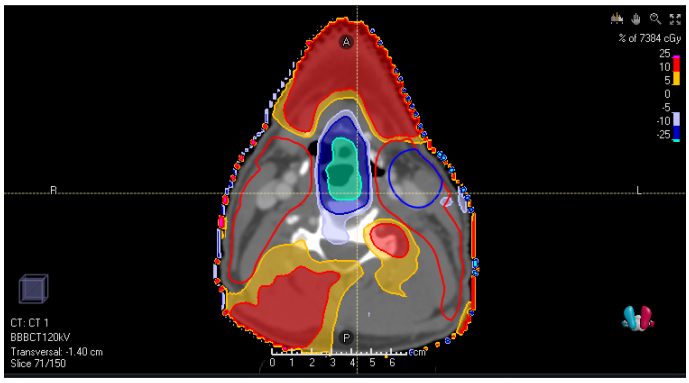
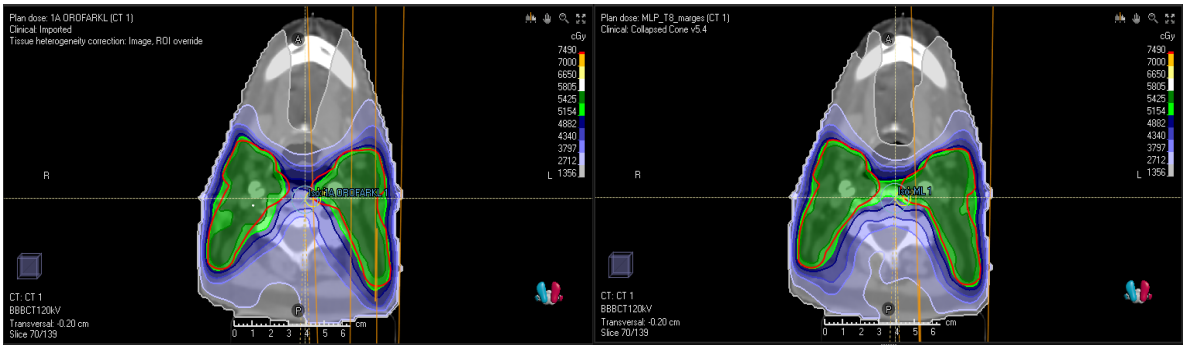


Figure 19: Transversal view of bilateral oropharynx patient where difference in dose distribution is visualized. Clinical plan is subtracted from the DLAP plan.

Here, again the PTV DL1 is indicated with a red colored segmentation and PTV DL2 is indicated with a blue colored segmentation. The dose difference is indicated as a percentage difference, meaning that in this case the red color means that the DLAP plan is 10% hotter than the clinical plan. The orange color means a difference of 5% more dose in the DLAP plan. On the contrary, the blue colors means that this part is *colder* in the DLAP plan than in the clinical plan. The turquoise color indicated a dose decrease of 25% compared to the clinical plan. The dark blue and light blue color show a decrease of 10% and 5% respectively. This is in line with what happened with the NTCP, since the PCM inferior receives less dose and the NTCP for dysphagia shows a decrease compared to the clinical plan. Furthermore, this also shows a indication of the significant increase in the mandible as previously seen in figure 14, since the red area indicates a higher dose in this structure compared to the clinical plan. This phenomenon is appearing in more DLAP plans.

The second patient in the NTCP plot (figure 17) shows a negative NTCP value, indicating that the clinical plan is preferred over the DLAP plan. In figure 20 the clinical and DLAP plan of this bilateral oropharynx patient is shown. Here the opposite effect of the previous patient can be observed. This figure shows again twice the same transversal slice of a patient. Anatomically this slice goes right through the PCM medius.



(a) Clinical (b) DLAP

Figure 20: Transversal view of bilateral oropharynx patient, where an increase in NTCP dysphagia $\geq$ 2 and 3 toxicities is noticed

Here, the low dose level in the elective glands is again indicated with green. In the DLAP (figure 20b)

this green dose distribution continues between the two parts of PTV DL1, right through the PCM medial. The primary tumor and thereby PTV DL2, are not visible in this slice. In the DLAP there are three subsequent slices where this continuous low dose level is causing more dose in the PCM medius, contributing to a increase in NTCP compared to the clinical plan.

It is important to notice that NTCP of both patient discussed here did not exceed the threshold value for dysphagia in grade ≥ 2 and ≥ 3 toxicities. The aim of the more in depth visual comparison is to gain insight in the difference of dose distribution in clinical and DLAP plans.

4.3 Patient Cohort 2: Hypopharynx and Larynx

The second patient cohort is divided into two tumor sites, namely hypopharynx tumors and larynx tumors. These two cohorts consist of eleven and eight patients, respectively. The prescribed dose and fractionation scheme remains the same as for the oropharynx patients. All patients from both tumor sites are bilateral irradiated in treatment.
The hypopharynx and larynx are located inferior to the oropharynx. The ROIs, target and OARS, are approximately similar to oropharynx patients. Hence, it is expected that DLAP generates acceptable treatment plans even though the DLAP is not trained on these tumor sites.

4.3.1 Dosimetric Analysis

The dosimetric analysis of hypopharynx and larynx patients is combined in one subsection of this thesis. Nevertheless, the parameters conducted from the DVHs are graphically represented in separate boxplots since the larynx tumors and hypopharynx tumors are different subgroups in HNC.
While generating the DLAP plans for larynx patient the problem occurred that the cochlea (inner ear) was not segmented in the clinical plan of a few patients. The cochlea is one of the mandatory ROIs to let the DLAP generate a treatment plan, which means that the cochlea had to be segmented. For the segmentation of this small ROI the autosegmentation module in RayStation is used, this model utilizes deep learning to automatically segment ROIs [32].

Fist of all the dosimetric parameteres in the target structures of hypopharynx patients are shown in figure 21 and for larynx patients in figure 22.

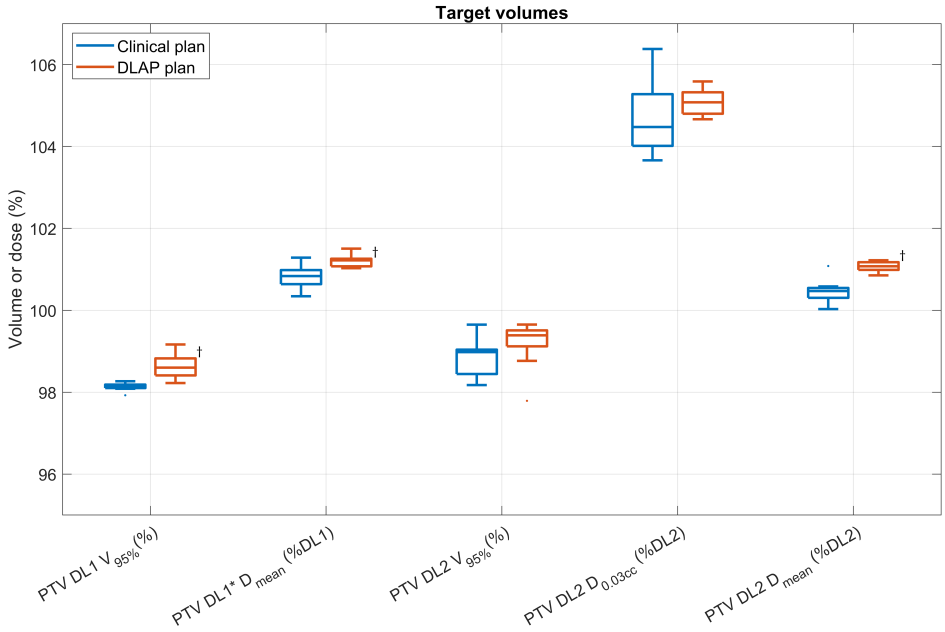


Figure 21: Dosimetric parameters of target structures from bilateral hypopharynx patients from the clinical plan (blue) and DLAP plan (orange).

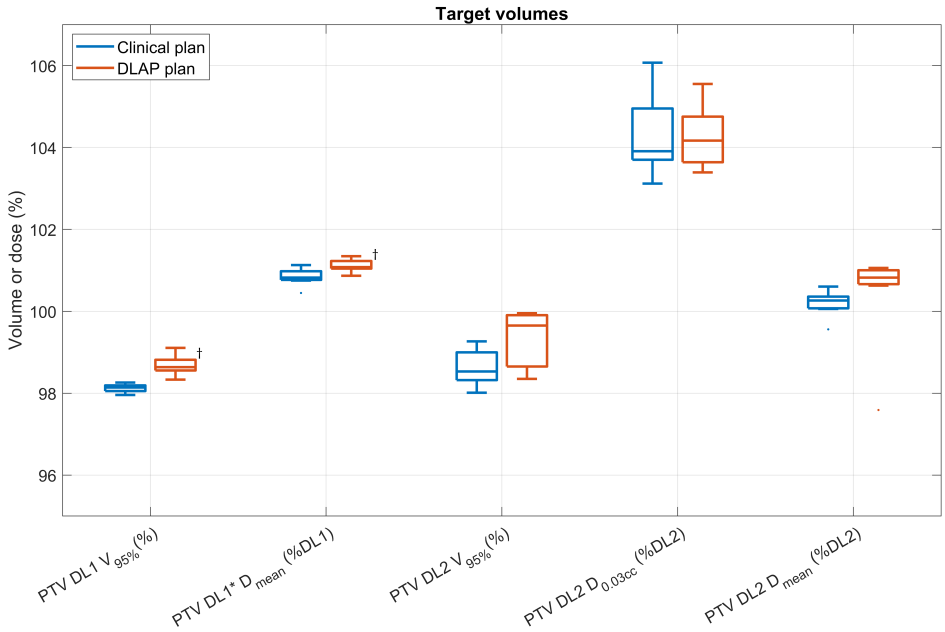


Figure 22: Dosimetric parameters of target structures from bilateral larynx patients from the clinical plan (blue) and DLAP plan (orange).

Comparison between the clinical plan and the DLAP for patients with tumors in the hypopharynx and larynx revealed significantly higher coverage of the PTV in both cases. The hypopharynx DLAP plans show a significantly higher PTV coverage compared to the clinical plan in PTV DL1 $V_{95\%}$, PTV DL1 D_{mean} and PTV DL2 D_{mean}. Similarly, in the larynx patients, the DLAP demonstrates a higher median PTV coverage for PTV DL1 and PTV DL2, although the difference is only statistically significant in PTV DL1 $V_{95\%}$ and PTV DL1 D_{mean}. These findings indicate that the trend of a higher coverage

in the DLAP, that was previously observed in the sub-study and in oropharynx patients, continues for hypopharynx and larynx patients as well.

The boxplots in figure 23 and 24 show the dosimetric parameters of the OARs in the case of hypopharynx and larynx tumors.

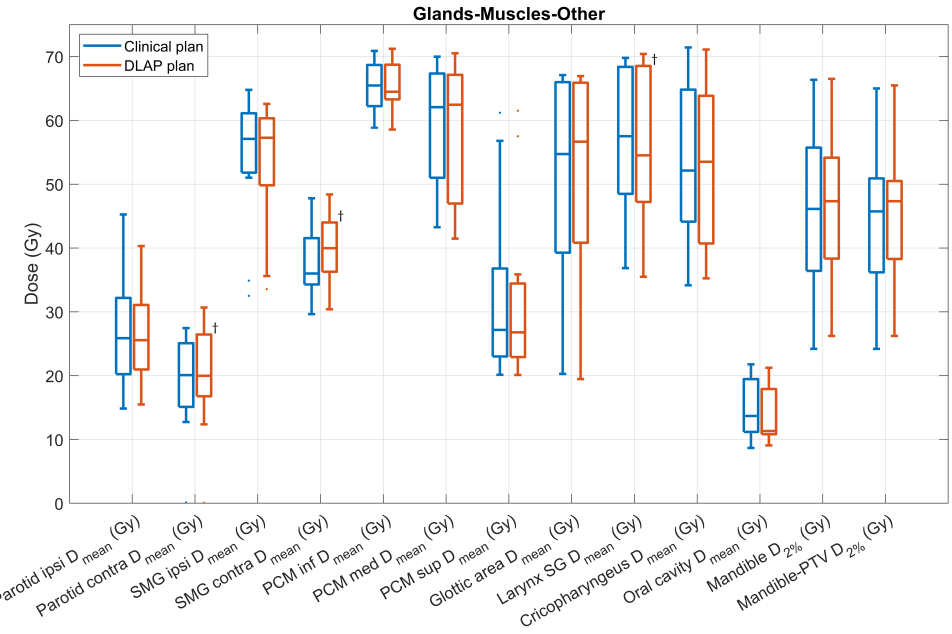


Figure 23: Dosimetric parameters of OARs (glands, muscles and others) from bilateral hypopharynx patients from the clinical plan (blue) and DLAP plan (orange).

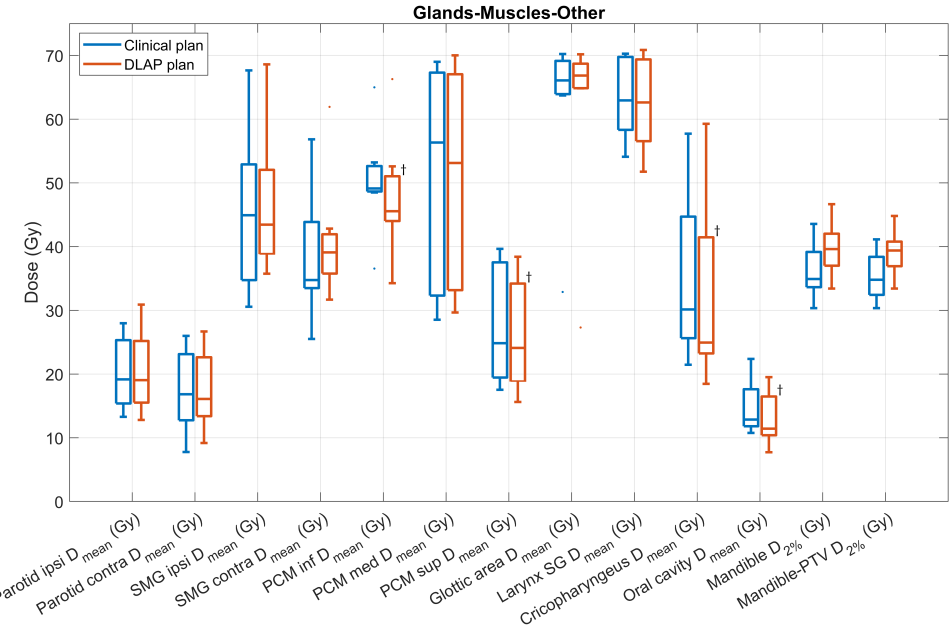
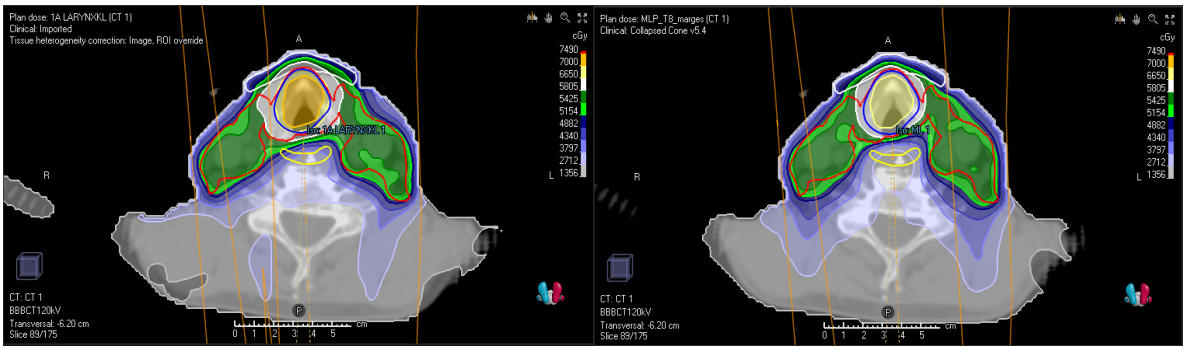


Figure 24: Dosimetric parameters of OARs (glands, muscles and others) from bilateral larynx patients from the clinical plan (blue) and DLAP plan (orange).

In the cohort of hypopharynx patients, the comparison between the clinical plan and the DLAP plan revealed not many notable differences in the doses delivered to the OARs. DLAP plans do show higher doses in contra-lateral OARs, which is statistically significant. However, there is a significantly lower dose in the larynx and a lower median dose in the oral cavity. Further, the dose distribution in the OARs is very similar to the clinical plans. This suggests that the DLAP is able to generate treatment plans for hypopharynx patients that spare the OARs similarly to the clinical plan while the PTV coverage is significantly higher.

In the cohort of larynx patients, DLAP plans showed significantly lower doses in the PCM inferior, PCM superior, cricopharyngeus and the oral cavity. Surprising, since the PCM inferior and cricopharyngeus are situated in close proximity to the larynx, see figure 1 and 2b.
An interesting trend might be emerging in the DLAP plans for larynx patients when investigating the potential cause of the significant decrease in the D_{mean} in the PCM inferior. In one patient the D_{mean} of the PCM inferior in DLAP is almost 6 Gy lower than in the clinical plan. This particular patient is shown in figure 25. The yellow segmented ROI is the PCM inferior.



(a) Clinical (b) DLAP

Figure 25: Transversal view of bilateral larynx patient, where PCM inferior (yellow) is better spared in the DLAP compared to the clinical plan. Note that WEM is used in treatment planning of this patient, indicated with a white colored segmentation.

In figure 25b DLAP shows a dose distribution where the PCM inferior is avoided better than in the clinical plan (figure 25a). This particular dose distribution causes the decrease of average dose in the PCM inferior for this patient, it seems that the DLAP is capable of better sparing the PCM inferior than in the clinical plan. This trend of avoiding the PCM inferior, even though this structure is situated close to PTV DL1 and DL2, is similar in other larynx patients from this patient cohort.
The rest of the OARs depicted in the boxplots for the larynx patients are quite similar to the clinical plan.

For the hypopharynx and larynx patient the boxplots for the OARs related to the nervous system are shown in figure 26 and 27.

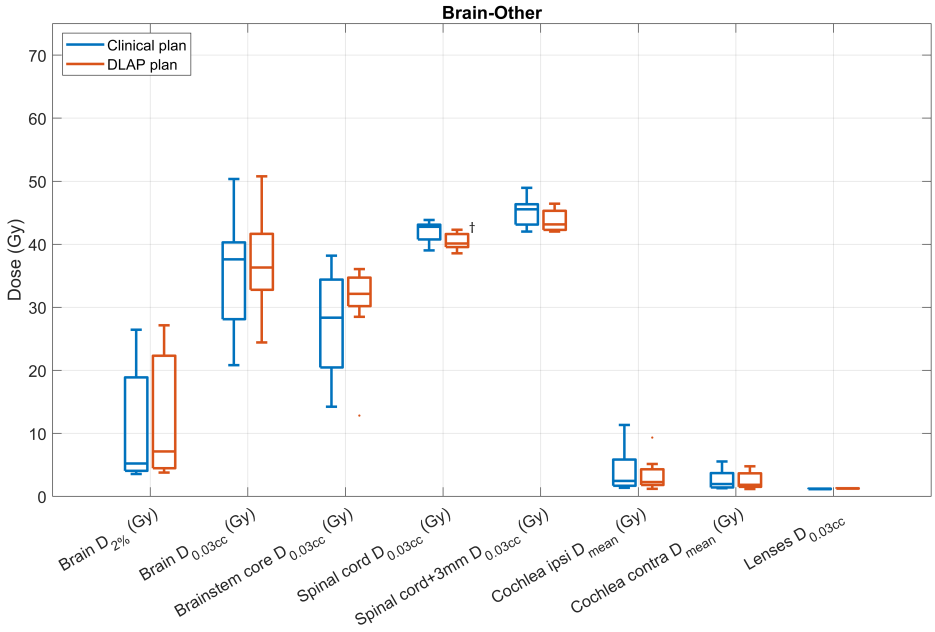


Figure 26: Dosimetric parameters of OARs (related to nervous system) from bilateral hypopharynx patients from the clinical plan (blue) and DLAP plan (orange).

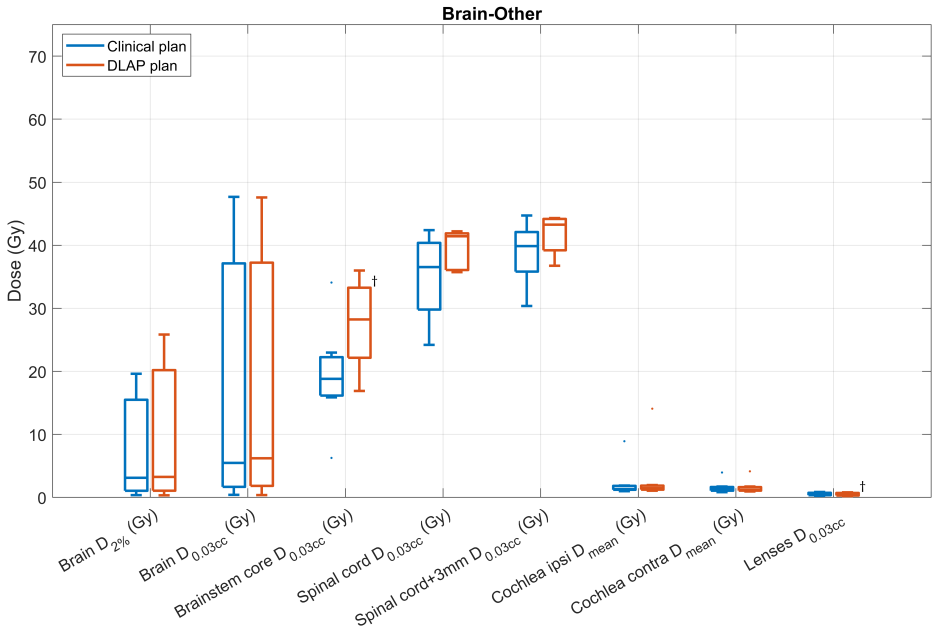
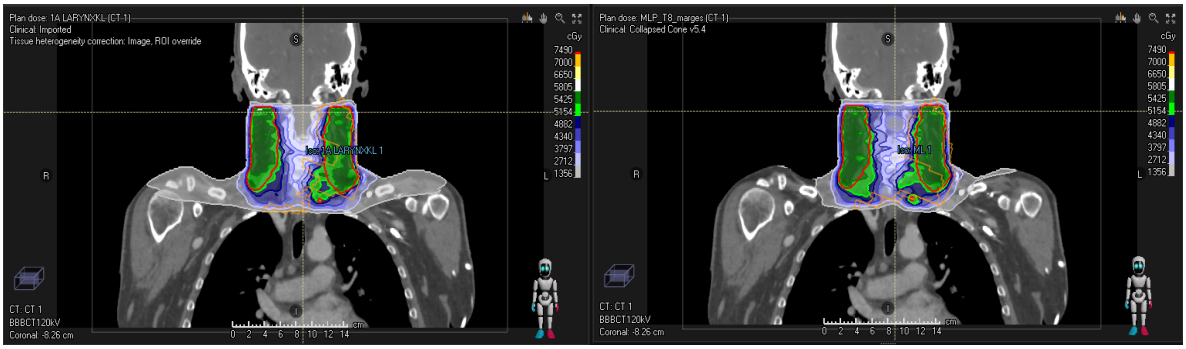


Figure 27: Dosimetric parameters of OARs (related to nervous system) from bilateral larynx patients from the clinical plan (blue) and DLAP plan (orange).

For hypopharynx patients the only significant difference is noticed in the spinal cord. The spinal cord received significantly less dose, while the other OARs are similar to the clinical plan.
In larynx patients, there is a significant higher doses observed in the brain stem core. To find the cause of this significant difference the DVH parameters were further investigated. This lead to a increase in dose in the brain stem core in multiple patients. The coronal view of the clinical plan and DLAP plan of a larynx patient is shown in figure 28.



(a) Clinical (b) DLAP

Figure 28: Coronal view of bilateral larynx patient, where brain stem core receives more dose in the DLAP compared to the clinical plan.

In figure 28 both clinical and DLAP plan are visible in a coronal plane to have a better view on the PTV DL1 and brain stem core simultaneously. PTV DL1 is indicated in red, the primary tumor is not visible in this slice. In both clinical and DLAP plan the green lobe outside the PTV DL1 is due to the location of the primary tumor (PTV DL2).
Figure 28b shows that the DLAP distributes a higher dose (darker blue) closer to the brain stem core which is at the base of the brain. This trend is noticeable in more of the larynx patient, causing the significant increase of dose in the DLAP compared to the clinical plan. To represent this difference also graphically the DVH is shown in figure 29.

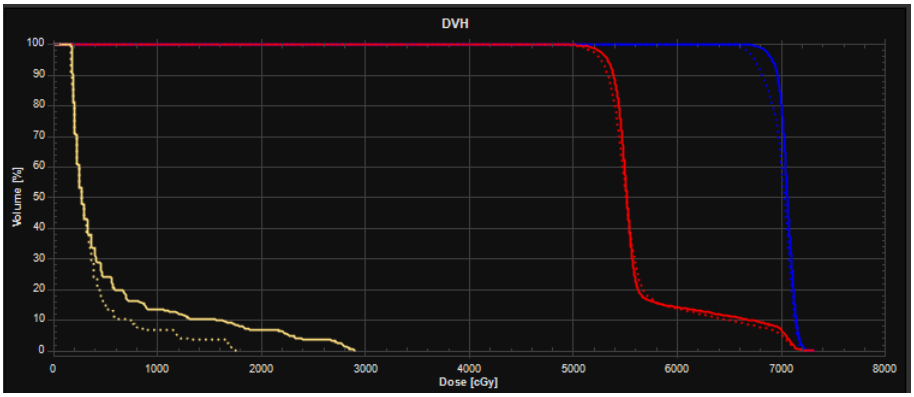


Figure 29: DVH of larynx patient from figure 28. PTV DL1 is red, PTV DL2 is blue and brain stem core is yellow. The dashed line indicated the clinical plan and the continuous line the DLAP.

In the DVH of this patient only the structures in question are shown, meaning the PTV DL1 (red), the PTV DL2 (blue) and the brain stem core (yellow). The dashed line is the clinical plan and the continuous line is the DLAP. In this DVH it is visible that PTV DL1 and PTV DL2 both contribute to more dose in the target structures, complimenting the overall trend of a higher PTV coverage in the DLAP. Undesirably, in this case the same happens in the ROI brain stem core. The continuous line, the DLAP plan, indicated that there is more dose distributed in the brain stem core than in the clinical plan. Important to notice is that even though there is an increase, the dose in the brain stem core is not exceeding the clinical goal of $D_{0.03cc} \leq 54$ Gy (table 5).

The dosimetric analysis indicated that the DLAP generated plans that meet the clinical goals not only for oropharynx patients, but also for hypopharynx and larynx patients. But, important to mention, due to the small size of the patients cohort it is hard to draw a binding conclusion from the data as provided in the boxplots, DVH and other figures in this subsection. Patient specific differences might contribute significantly to the representation in the boxplots due to the small cohort size.

4.3.2 NTCP Analysis

In order to evaluate the impact of DLAP on patient outcome, the NTCP models for xerostomia and dysphagia grades ≥ 2 toxicities and grades ≥ 3 toxicities are once again used for all hypopharynx and larynx patients. A comparison is made by examining the difference in outcomes per model per patient. Given the relatively small patient cohort for both tumor sites compared to the oropharynx cohort, the absolute average difference between the plans was calculated using the formula:

$$\Delta NTCP = \frac{1}{N} \sum |NTCP_{clinical} - NTCP_{DLAP}| \tag{6}$$

Here, N represents the number of patients. The absolute difference between the NTCP values of the clinical plan and the DLAP is considered, as the values could be either positive or negative, depending on which plan exhibited a more favorable NTCP outcome. By taking the absolute difference, the calculation of the average difference accounts for both positive and negative deviations between the two plans.

Table 7: NTCP outcome for xerostomia and dysphagia grade ≥ 2 and 3 toxicities for bilateral hypopharynx and larynx patients.

	Xerostomia grade ≥ 2	**Xerostomia grade ≥ 3**	**Dysphagia grade ≥ 2**	**Dysphagia grade ≥ 3**
Hypopharynx	$\Delta 0.97\%$	$\Delta 0.32\%$	$\Delta 0.55\%$	$\Delta 0.21\%$
Larynx	$\Delta 0.68\%$	$\Delta 0.21\%$	$\Delta 0.62\%$	$\Delta 0.56\%$

Analysis of the NTCP values for hypopharynx and larynx patients, as shown in table 7, reveals that the difference in NTCP between the DLAP and clinical plan is minimal. With a threshold of 10% for grades ≥ 2 toxicities and 5% for grades ≥ 3 toxicities considered as a criterion for selecting an alternative plan, none of the individual cases exceeds this value. Therefore, the observed differences in NTCP between the DLAP and clinical plan can be considered insignificant, suggesting that both plans yield comparable NTCP outcomes for patients with hypopharynx and larynx tumors.

Nevertheless, it is interesting to investigate some of the differences in NTCP outcome. For example the treatment plans of a hypopharynx patient that is shown in figure 30. What is interesting about this particular patient is that the NTCP outcome indicated a 2.35% increase in xerostomia grade ≥ 2 toxicities in the DLAP compared to the clinical plan, which was the largest difference of all eleven hypopharynx patients.

(a) Clinical (b) DLAP

Figure 30: Transversal view of bilateral hypopharynx patient.

While visually evaluating both treatment plans in figure 30 it is noticeable that the dose in the DLAP on the patient's contra-lateral side is more equally distributed within the boundaries of the PTV DL1 than in the clinical plan. This means that the contra-lateral submandible gland therefore receivers more dose, hence the increase in NTCP for xerostomia. On the other hand, the low dose level now

stays within the medial boundaries causing a lower dose in the PCMs. The difference between the two treatment plans is also depicted in figure 31.

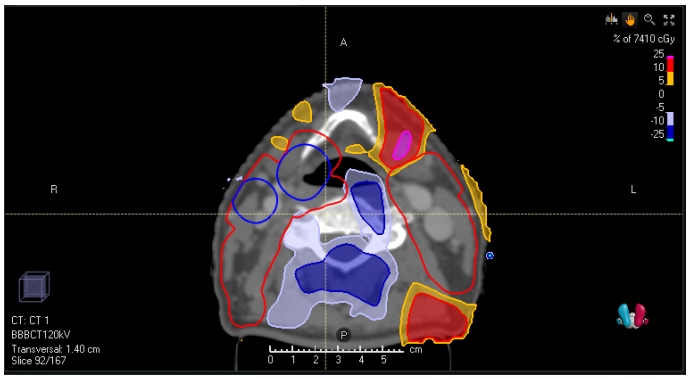


Figure 31: Transversal view of bilateral hypopharynx patient where difference in dose distribution is visualized. Clinical plan is subtracted from the DLAP plan.

The orange, red and pink *hot spot* on the left side of the patient right above PTV DL1 shows a relative increase with respect to the maximum dose of 5%, 10% and 25% respectively. The pink colored area that indicates the 25% increase is partly located inside the contra-lateral submandible gland. This might also contribute to the significant difference shown in the boxplot of this ROI, see figure 23. The ipsi-lateral submandible gland in this patient receives a high dose anyway since most of it is part of the PTV DL2.
The light blue and dark blue medial *cold spot* indicates a decrease in dose of 5% for the light blue part and 10% for the darker blue part. Due to this cold spot the PCMs receive less dose and this contributes in a decrease of NTCP for dysphagia grade ≥ 2 toxicities of 1.05% compared to the clinical plan.

This specific example demonstrates the importance of an accurate treatment plan. Modifying the dose distribution to minimize exposure to certain OARs can lead to an increased dose in other OARs. It highlights the need for thorough evaluation and review of treatment plans by experts to ensure optimal balancing of dose distribution and minimize potential risks to critical structures.

4.4 Patient Cohort 3: Unilateral Oropharynx

Previous patient cohorts focused on bilateral irradiation of oropharynx, hypopharynx and larynx patients. Another possibility in radiotherapy treatment of HNC is unilateral irradiation. In this case only the patient's ipsi-lateral side is irradiated. This third and final cohort consists of nine unilateral oropharynx patients. The prescribed dose and fractionation scheme remains the same as for bilateral oropharynx, hypopharynx and larynx patients.
DLAP is trained on bilateral oropharynx patients. Since only one side of the patient contains target ROIs, but the OARs are similar is it expected that DLAP generates a treatment plan that satisfies the treatment constraints but possibly delivers a higher dose in the contra-lateral side.

4.4.1 Dosimetric Analysis

The analysis starts once again with the dosimetric parameters in the target structures. This is similar to what is done with the bilateral oropharynx patient, but the PTV DL1 now only exists on the ipsi-lateral side. The dose constraints for target structures remain the same as for bilateral oropharynx patients. The dose distribution is graphically displayed in the boxplot in figure 32.

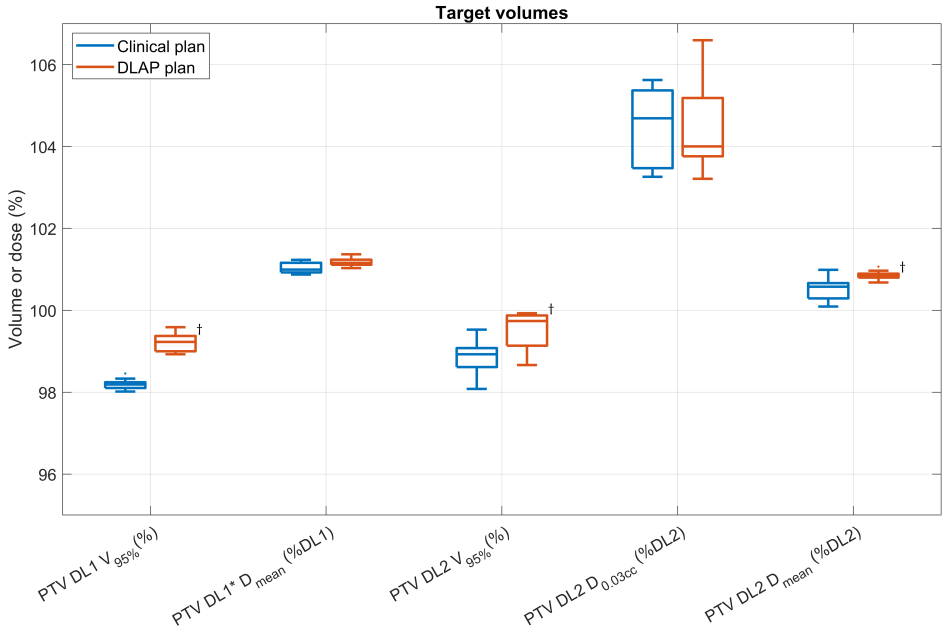


Figure 32: Dosimetric parameters of target structures from unilateral oropharynx patients from the clinical plan (blue) and DLAP plan (orange).

Comparing the clinical plan with the DLAP plan for unilateral oropharynx patients, it is observed that DLAP has a higher PTV coverage. The DLAP shows a significant increase in coverage in PTV DL1 $V_{95\%}$, PTV DL2 $V_{95\%}$ and PTV DL2 D_{mean}. This finding aligns with the trend seen in the DLAP for the bilateral patients, where the DLAP consistently show a higher PTV coverage compared to the clinical plan.

The dose distribution in the OARs for unilateral oropharynx patients is shown in figure 33.

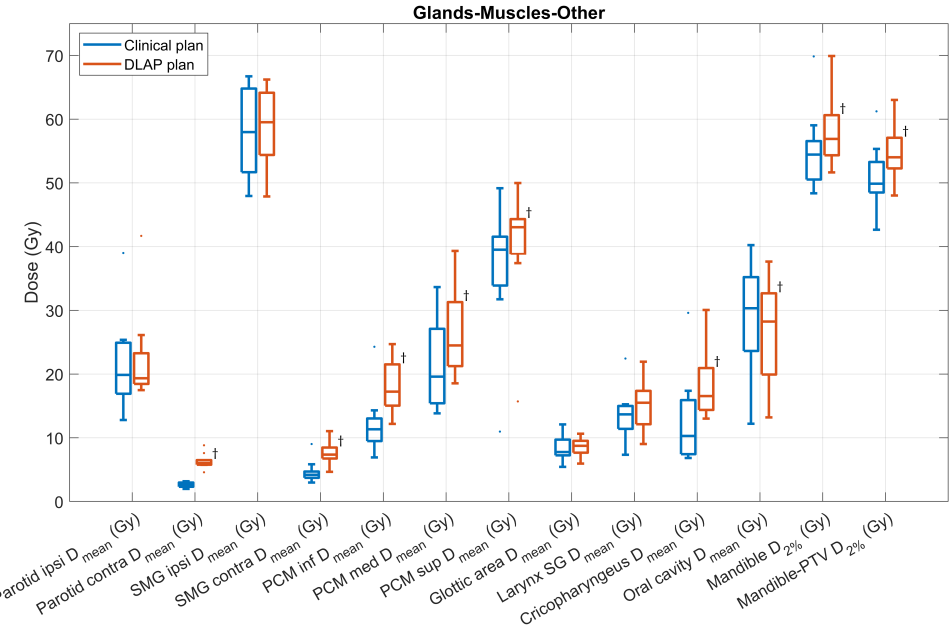


Figure 33: Dosimetric parameters of OARs (glands, muscles and others) from unilateral oropharynx patients from the clinical plan (blue) and DLAP plan (orange).

In terms of dose distribution in the OARs, the comparison between the clinical plan and the DLAP plan for unilateral oropharynx patients shows significant differences. The DLAP results in significantly higher doses in the contra-lateral OARs, while the doses in the ipsi-lateral OARs are similar between the two plans. In the DLAP the contra-lateral sides show a increase of multiple Gray compared to the clinical plan. Furthermore, the contra-lateral parotid exceeds the dose constraint of ≤ 5 Gy, as can be seen in table 4.
Additionally, in the DLAP there are significantly higher doses observed in all the ROIs associated with the swallowing muscles, namely the PCM inferior, PCM median, PCM superior and cricopharyngeus. The glottic area and larynx have a higher median dose, but there is no significant difference with the clinical plan. The higher dose in the swallowing muscles, glottic area and larynx can be a result of the extra dose distributed in these structures from the contra-lateral direction in the DLAP compared to the clinical plan. It is important to note once more that the DLAP used in this study was not trained on unilateral patients, and this is observable in the comparison, as the contra-lateral side receive significantly higher doses compared to the clinical plan. Another significant increase in dose is visible in the mandible structures. However, this is a trend that is observed before as well in the DLAP of bilateral patients.

A representative clinical and corresponding DLAP plan of one of the unilateral oropharynx patients is shown in figure 34.

(a) Clinical (b) DLAP

Figure 34: Transversal view of unilateral oropharynx patient.

Figure 34a shows the clinical treatment plan of an unilateral oropharynx patient. In this treatment plan the elective glands are only irradiated at the ipsi-lateral side. The PTV DL1 is segmented with red and PTV DL2 is segmented with blue. The clinical plan seems to realy spare the contra-lateral side, but as shown in figure 34b the DLAP tends to do this less. This phenomenon happens in all the nine patients, indicating that the DLAP does not work that well on unilateral patients as clinically. Hence, the significant increase in the various OARs. The difference between the clinical plan and DLAP of this patient is best visualized in figure 35 where the two plans are subtracted from each other.

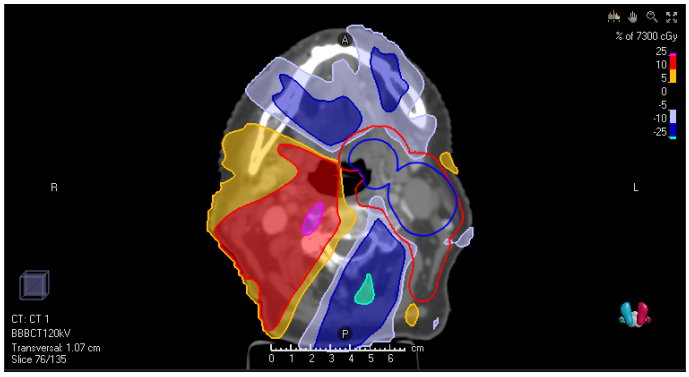


Figure 35: Transversal view of unilateral oropharynx patient where difference in dose distribution is visualized. Clinical plan is subtracted from the DLAP plan.

The big orange, red and pink hot spot in figure 35 shows clearly the effect on the contra-lateral side when using DLAP on unilateral oropharynx patients. Even though the OARs most of the time stay within the limits of the clinical goals, the higher dose on the contra-lateral side and other OARs is undesirable. The findings emphasize the influence of the change in treatment approach in DLAP. Furthermore, it shows the need for further improvement or training of the DLAP to better optimize unilateral radiation delivery for oropharynx patients.

Finally the OARs accosiated with the nervous system are shown in figure 36.

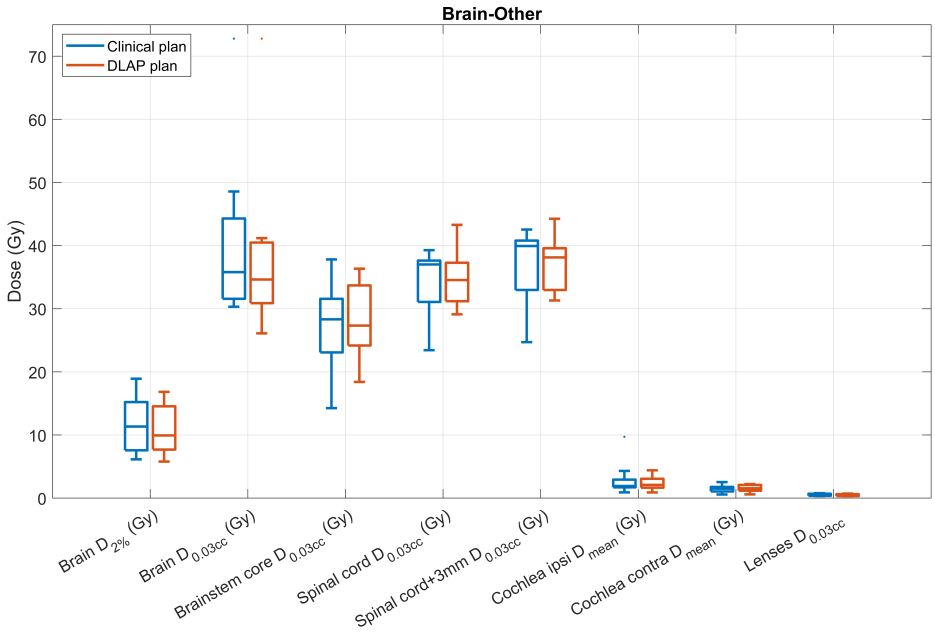


Figure 36: Dosimetric parameters of of OARs (related to nervous system) from unilateral oropharynx patients from the clinical plan (blue) and DLAP plan (orange).

Regarding the OARs in figure 36, there is no significant differences noticeable between the clinical plan and the DLAP for unilateral oropharynx patients. This indicates that the DLAP did not introduce any considerable increase or decrease in the radiation doses received by these OARs compared to the clinical plan.

4.4.2 NTCP Analysis

In the case of unilateral oropharynx patients, the absolute difference (6) between the NTCP of the DLAP and clinical plan is calculated accordingly to the NTCP models for xerostomia en dysphagia. The results are presented in table 8 for each model and grade of toxicity seperately.

Table 8: NTCP outcome for xerostomia and dysphagia grade ≥ 2 and 3 toxicities for unilater oropharynx.

	Xerostomia grade ≥ 2	**Xerostomia grade** ≥ 3	**Dysphagia grade** ≥ 2	**Dysphagia grade** ≥ 3
Unilateral	Δ1.46%	Δ0.40%	Δ0.53%	Δ0.17%

Specifically, when considering xerostomia grade ≥ 2, it is worth noting that only one case demonstrates a slightly better NTCP with the DLAP plan, showing a difference of +0.45%. Conversely, in the eight other patients, the NTCP values favor the clinical plan, with a maximum difference of -3.38%. However, it is important to highlight that none of these differences exceed the threshold value. Furthermore, despite the boxplot indicating that all the swallowing muscles (PCM) receive a significantly higher dose in the DLAP plan, the variance in NTCP for dysphagia is minimal.

5 Discussion

The importance of accurate treatment planning in the effectiveness and safety in delivering radiotherapy is hopefully clear. Consequently, treatment planning significantly contributes to the outcome of a patients cancer treatment and quality of life. This can be a challenging and certainly a time consuming task when done manually, especially in HNC. In this thesis the use of DLAP in HNC radiotherapy is validated to investigate if DLAP could enhance the quality, consistency and efficiency in radiotherapy treatment planning.

5.1 Treatment Planning

The DLAP that is proposed in the thesis does still need a work around, i.e. the extra local 2 mm margin of the PTV in medial direction at level of the swallowing muscles. The expansion can be automated and does not need manual intervention. Although the current clinical introduction of DLAP may involve the PTV expansion, the ultimate goal is to eliminate the need for such interventions and solve this underdosage within in the DLAP model. Therefore, it is essential for the LUMC, in collaboration with the technical support of RayStation, to focus on resolving this problem within the DLAP.

The OARs are located differently towards the primary tumor in every patient. Some OARs are located inside the high or low target dose (PTV DL1 or PTV DL2). This can cause a higher dose in the OAR than in other patients. In manual treatment planning it occurs that specific ROIs are not segmented due to tumor location, e.g. with larynx tumors when the primary tumor covers up most of the larynx or glottic area. To generate a DLAP plan the structures need to be segmented in spite of the primary tumor location for the DLAP to work. In a few patient the mandatory structures are segmented later on to generate a DLAP corresponding to the clinical plan of the patient. This also results in the OAR receiving much more dose compared to the same OAR in other patients. The patient specific differences explain the sometimes wide spread that is noticeable in the boxplots.

Treatment planning will obviously keeps needing the professional opinion from a radiation oncologist and clinical physicist before administrating the dose to a patient. Nevertheless, generating a treatment plan that meets the clinical goals much faster than manually can result in big advancements for radiotherapy treatment.

5.2 Data Analysis

The available patient data was limited due to factors such as manual anonymization, manual expansion of the PTV, and the availability of patients in RayStation. As a result, the patient cohort for oropharynx ultimately consisted of 43 patients, which, although not extensive, still provides some meaningful insights. With cautious interpretation, it can be suggested that DLAP generates qualitatively good treatment plans compared to the corresponding clinical plans.
However, it is important to note that the patient groups for hypopharynx, larynx, and unilateral oropharynx were relatively small. In the larynx patient cohort, a single patient significantly contributed to the observed increase in dose in the brainstem core. This outcome might be caused by the small size of the patient group. With a larger sample size, possible outliers could be better recognized and their impact could be more accurately evaluated in the graphical representation of the boxplots.

For the NTCP analysis assumptions were made about the baseline and treatment technique, i.e. zero baseline and primarily treated. The outcome of the NTCP is unlikely to be different when other assumption would have been made. but this is not tested. For this thesis the goal was to evaluate the NTCP of the clinical plan with the DLAP, justifying the choice to make the same assumption for every patient.
Additionally, the validation, accuracy and reliability of NTCP models in HNC are continuously being studied and refined [33, 34, 35]. Therefore, a conclusion solely based on the NTCP outcome might not be valid.
However, the found difference in NTCP outcome for both xerstomia and dyspagia for all tumor sides are very small. This indicates that the DLAP did not significantly improved of worsen the NTCP, the DLAP generated a plan that is comparable to clinical based on NTCP.
The NTCP analysis contributed to some insight in the differences that appear in the boxplots. The NTCP per patients displayed a nice and clear overview where to look for a certain significant difference when it has to do with an OAR that is involved in developing xerostomia or dysphagia. Furthermore, the described NTCP models are currently used in national guidelines to stratify patients between photon and proton radiotherapy and therefore the analysis could not be left out of this thesis.

5.3 Results

The patient plans that are displayed as example were carefully chosen to explain the differences that are visible in the boxplots or NTCP analysis, but are not all representative for the whole patient cohort. For example, the unilateral patients all do show a higher dose on the contra-lateral side as can be seen in figure 34 and 35. But the exact tumor location, segmentation and dose distribution differs from each individual patient. The same holds for the other examples that are given throughout the thesis.

DLAP demonstrates a consistent trend across all investigated tumor sites, showing significantly higher coverage of the PTV compared to the clinical plan. This indicates that the tumor region receives a higher radiation dose, resulting in a potentially *hotter* treatment area. The potential effects of the increased radiation dosage on patients needs to be considered. While most OARs did not show significant differences in terms of the average or maximum dose and the NTCP outcome was negligible in most patients, it is still important to proceed with caution and careful consideration. There is potentially room for improvement here. If coverage is lowered to imitate the coverage of the clinical plans, the OAR dose might reduce some more.
In clinical setting, DLAP plans with relatively high PTV coverage may require additional manual planning steps to bring the coverage closer to the minimum threshold of 98%. The development of an exact protocol will depend on the observations and outcomes in the first clinical patients plannend with DLAP.
An alternative approach to address the issue of higher coverage in treatment plans is through normalization. Normalizing the treatment plan involves adjusting the dose distribution so that the higher coverage is normalized to a lower level, while maintaining consistency with the rest of the dose distribution. In plan normalization the beam monitor units are scaled, and thus the dose distribution, such that $V_{95\%} = 98\%$ is prescribe for every patient.

However, it is important to note that in this thesis, the normalization of treatment plans was not investigated as it is not a standard procedure at the LUMC. While normalization may have the potential to address the issue of higher coverage in treatment plans, its clinical implementation does not align with the protocols followed at the LUMC. As a result, investigating the normalization of treatment plans fell outside the scope of this thesis. Nevertheless, the possibility of normalization remains an area of interest for future research and warrants further investigation to assess its effectiveness and potential benefits.

DLAP will be implemented for treatment planning in oropharynx patients in a clinical setting. To investigate the reduction in treatment planning time achieved compared to manual treatment planning, it is essential to evaluate the frequency of immediate acceptance of DLAP plans and the cases where additional manual planning is required. This will provide valuable insights into the efficiency of DLAP. By analyzing the acceptance rate and the need for additional manual planning, as well as quantifying the time spend on additional manual treatment planning, a evaluation of the efficiency of DLAP can be obtained.

Drawing a definitive conclusion from the dosimetric and NTCP data from the hypopharynx, larynx and unilateral patients is challenging due to the limited sizes of the patient groups. For hypopharynx and larynx tumors the DLAP shows promising results, even though the DLAP is not trained on these specific tumor sites. Due to the small number of patients in the study, caution must be taken when interpreting the results. Further research with a larger patient cohort is needed to validate and strengthen the conclusions drawn from the dosimetric and NTCP analysis form these patient cohorts.

The observation that the unilateral treatment approach results in significantly higher dose distribution in the contra-lateral side is concerning and not in line with the treatment objective of sparing that region. The primary intention of choosing a unilateral treatment strategy is to minimize radiation exposure and potential side effects to the contra-lateral side. This finding highlights the need for further investigation or optimization of the DLAP for unilateral oropharynx patients to ensure better sparing of the contra-lateral side during unilateral treatments.

6 Conclusion

In conclusion, this thesis demonstrates that DLAP, with the local expansion of the PTV around the swallowing muscles, can be effectively used for treatment planning in oropharynx patients. Oropharynx patients represent the largest subgroup of HNC patients treated at the LUMC. The consistency in PTV coverage of the clinical plans is better than in DLAP plans, due to intentionally steering to a coverage of 98%. Even though the PTV DL1 and PTV DL2 receives a higher dose in the DLAP, the OARs are still spared within the range of the clinical goals set by the LUMC. In some patients the OARs are even better spared compeared to the clinical plan.
The NTCP comparison between the clinical and the DLAP plan indicates that there is no higher chance of patients developing xerostomia or dysphagia with DLAP. In almost half of the patients there is even a lower chance of developing xerostomia or dysphagia after radiotherapy treatment with DLAP.

For hypopharynx and larynx tumors, which share similarities in target and OAR structures with oropharynx tumors but differ in their inferior location, the results indicate promising outcomes. However, further validation and a larger patient cohort are necessary before considering clinical implementation of the DLAP oropharynx model for these, not intended, tumor locations.

An area that requires additional research is the treatment of unilateral oropharynx patients using the proposed DLAP. Currently, the sparing of the contra-lateral side is not achieved equally good as the clinical plan. It is recommended to explore further research in this area or await the development of a specialized DLAP module from RaySearch designed specifically for unilateral radiation therapy in HNC.

Overall, this thesis highlights the potential benefits of DLAP for treatment planning in HNC, specifically for oropharynx patients. However, ongoing research and optimization efforts are crucial to address the challenges identified in hypopharynx, larynx, and unilateral cases, ensuring improved treatment outcomes and the preservation of critical structures in these patient as well.

References

[1] Integraal Kankercentrum Nederland, "Hoofd-hals kanker." https://iknl.nl/kankersoorten/hoofd-halskanker accessed: 22.03.2023.

[2] E. Tumban, "A current update on human papillomavirus-associated head and neck cancers," *Viruses*, vol. 11, no. 10, p. 922, 2019.

[3] C. De Martel, M. Plummer, J. Vignat, and S. Franceschi, "Worldwide burden of cancer attributable to hpv by site, country and hpv type," *International journal of cancer*, vol. 141, no. 4, pp. 664–670, 2017.

[4] L. Q. Chow, "Head and neck cancer," *New England Journal of Medicine*, vol. 382, no. 1, pp. 60–72, 2020.

[5] W. Pinkas, M. Jankowski, and W. Wierzba, "Awareness of head and neck cancers: A 2021 nationwide cross-sectional survey in poland," *Journal of Clinical Medicine*, vol. 11, no. 3, p. 538, 2022.

[6] D. Alterio, G. Marvaso, A. Ferrari, S. Volpe, R. Orecchia, and B. A. Jereczek-Fossa, "Modern radiotherapy for head and neck cancer," in *Seminars in oncology*, vol. 46, pp. 233–245, Elsevier, 2019.

[7] National Cancer Institute, "Types of cancer treatment." https://www.cancer.gov/about-cancer/treatment/types accessed: 12.07.2023.

[8] D. Nguyen, T. Long, X. Jia, W. Lu, X. Gu, Z. Iqbal, and S. Jiang, "A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning," *Scientific reports*, vol. 9, no. 1, p. 1076, 2019.

[9] National Cancer Institute, "Head and neck cancers fact sheet." https://www.cancer.gov/types/head-and-neck/head-neck-fact-sheet accessed: 21.03.2023.

[10] A. F. Osman and N. M. Tamam, "Attention-aware 3d u-net convolutional neural network for knowledge-based planning 3d dose distribution prediction of head-and-neck cancer," *Journal of Applied Clinical Medical Physics*, vol. 23, no. 7, p. e13630, 2022.

[11] C. Wang, X. Zhu, J. C. Hong, and D. Zheng, "Artificial intelligence in radiotherapy treatment planning: present and future," *Technology in cancer research & treatment*, vol. 18, 2019.

[12] I. Hazell, K. Bzdusek, P. Kumar, C. R. Hansen, A. Bertelsen, J. G. Eriksen, J. Johansen, and C. Brink, "Automatic planning of head and neck treatment plans," *Journal of Applied Clinical Medical Physics*, vol. 17, pp. 272–282, 2016.

[13] S. P. Ng, C. Pollard III, M. Kamal, Z. Ayoub, A. S. Garden, H. Bahig, G. B. Gunn, S. J. Frank, H. D. Skinner, J. Phan, *et al.*, "Risk of second primary malignancies in head and neck cancer patients treated with definitive radiotherapy," *NPJ Precision Oncology*, vol. 3, no. 1, p. 22, 2019.

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

[15] J. ur Rehman, N. Ahmad, M. Khalid, Z. A. Gilani, I. Ullah, G. Nasar, M. M. Akhtar, M. N. Usmani, *et al.*, "Intensity modulated radiation therapy: A review of current practice and future outlooks," *Journal of radiation research and applied sciences*, vol. 11, no. 4, pp. 361–367, 2018.

[16] M. Teoh, C. Clark, K. Wood, S. Whitaker, and A. Nisbet, "Volumetric modulated arc therapy: a review of current literature and clinical use in practice," *The British journal of radiology*, vol. 84, no. 1007, pp. 967–996, 2011.

[17] E. Orlandi, M. Palazzi, E. Pignoli, C. Fallai, A. Giostra, and P. Olmi, "Radiobiological basis and clinical results of the simultaneous integrated boost (sib) in intensity modulated radiotherapy (imrt) for head and neck cancer: a review," *Critical reviews in oncology/hematology*, vol. 73, no. 2, pp. 111–125, 2010.

[18] D. Head and N. C. Group, "Radiotherapy guidelines 2020," *Clinical Practice Guideline*, 2020.

[19] W. Dörr, T. Herrmann, and K.-R. Trott, "Normal tissue tolerance," *Translational Cancer Research*, vol. 6, no. Suppl 5, 2017.

[20] A. Niemierko and M. Goitein, "Modeling of normal tissue response to radiation: the critical volume model," *International Journal of Radiation Oncology* Biology* Physics*, vol. 25, no. 1, pp. 135–145, 1993.

[21] P. G. Hawkins, A. S. Kadam, W. C. Jackson, and A. Eisbruch, "Organ-sparing in radiotherapy for head-and-neck cancer: Improving quality of life," vol. 28, no. 1, pp. 46–52, 2018.

[22] National Cancer Institute, "Head and neck cancers fact sheet." https://visualsonline.cancer.gov/details.cfm?imageid=9435.

[23] L. LoPresti, "Head and neck anatomy: Part ii - musculature," *DenalCare*, 2022.

[24] I. Beetz, C. Schilstra, A. Van Der Schaaf, E. R. Van Den Heuvel, P. Doornaert, P. Van Luijk, A. Vissink, B. F. van der Laan, C. R. Leemans, H. P. Bijl, *et al.*, "Ntcp models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: the role of dosimetric and clinical factors," *Radiotherapy and Oncology*, vol. 105, no. 1, pp. 101–106, 2012.

[25] H. P. van der Laan, H. P. Bijl, R. J. Steenbakkers, A. van der Schaaf, O. Chouvalova, J. G. Vemer-van den Hoek, A. Gawryszuk, B. F. van der Laan, S. F. Oosting, J. L. Roodenburg, *et al.*, "Acute symptoms during the course of head and neck radiotherapy or chemoradiation are strong predictors of late dysphagia," *Radiotherapy and Oncology*, vol. 115, no. 1, pp. 56–62, 2015.

[26] Landelijk Platform Protonentherapie and Landelijk Platform Radiotherapie Hoofd-halstumoren, "Landelijk indicatie protocol protonentherapie versie 2.2," 2019.

[27] C. Franzese, D. Dei, N. Lambri, M. A. Teriaca, M. Badalamenti, L. Crespi, S. Tomatis, D. Loiacono, P. Mancosu, and M. Scorsetti, "Enhancing radiotherapy workflow for head and neck cancer with artificial intelligence: A systematic review," *Journal of Personalized Medicine*, vol. 13, no. 6, p. 946, 2023.

[28] RaySearch Laboratories AB (publ), "Deep learning planning models," 2022.

[29] R. Laboratories, *Validation report for model RSL Oropharynx 7000 SIB (3.0)*. Stockholm, Sweden: RaySearch Laboratories, 2022.

[30] L. Grégoire, A. van Straaten-Huygen, and R. Trompert, *Anatomie en fysiologie van de mens*. Amersfoort: ThiemeMeulenhoff, 2014.

[31] G. Luxton, P. J. Keall, and C. R. King, "A new formula for normal tissue complication probability (ntcp) as a function of equivalent uniform dose (eud)," *Physics in Medicine & Biology*, vol. 53, no. 1, p. 23, 2007.

[32] RaySearch Laboratories AB (publ), "Deep-learning segmentation," 2020. https://www.raysearchlabs.com/siteassets/media/publications/white-papers/wp-pdfs/wp$_m l_d eeplearning_2$020.03.25.*pdf*.

[33] C. Hansen, J. Friborg, K. Jensen, E. Samsøe, L. Johnsen, R. Zukauskaite, C. Grau, C. Maare, J. Johansen, H. Primdahl, *et al.*, "Ntcp model validation method for dahanca patient selection of protons versus photons in head and neck cancer radiotherapy," *Acta Oncologica*, vol. 58, no. 10, pp. 1410–1415, 2019.

[34] M. Sharabiani, E. Clementel, N. Andratschke, and C. Hurkmans, "Generalizability assessment of head and neck cancer ntcp models based on the tripod criteria," *Radiotherapy and Oncology*, vol. 146, pp. 143–150, 2020.

[35] P. Blanchard, A. J. Wong, G. B. Gunn, A. S. Garden, A. S. Mohamed, D. I. Rosenthal, J. Crutison, R. Wu, X. Zhang, X. R. Zhu, *et al.*, "Toward a model-based patient selection strategy for proton therapy: external validation of photon-derived normal tissue complication probability models in a head and neck proton therapy cohort," *Radiotherapy and Oncology*, vol. 121, no. 3, pp. 381–386, 2016.

7 Appendix

7.1 Literature Study

Automated Treatment Planning in Photon Beam Radiotherapy for Head and Neck Cancer - Systematic Review

Simone Visser
Delft University of Technology
Leiden University Medical Center

April 17, 2023

Abstract

Introduction: Treatment planning in photon beam radiotherapy for head and neck cancer (HNC) is a time consuming task that requires experienced radiation treatment technologists (RTT) to make an adequate treatment plan. Automated treatment planning (autoplanning) in HNC is expected to improve the quality, consistency and efficiency of treatment planning in radiotherapy. This literature study examines the different types of autoplanning systems and compares automated treatment plans with clinical treatment plans.
Method: For this study, the literature search is limited to English written articles about autoplanning in HNC from the last ten years. This resulted in multiple articles about the working principles of autoplanning systems, and quantitative and qualitative comparison studies of autoplanning with manual treatment planning.
Results: Autoplanning is a much studied subject and is widely used in research setting. This study explains the working principles of autplanning systems, both self-made and commercially available. This includes atlas-based and knowledge-based autoplanning, Erasmus i-Cycle, Pinnacle Auto-Planning, RapidPlan by Varian and RayStation by RaySearch Laboratories.
Quantitative comparison studies conducted amongst different research groups and medical centers overall show a decrease in dose in the OARs and a similar or increased dose in the target. In all the studies the overall treatment planning time was significantly decreased with the use of autoplanning systems.
In qualitative comparison studies the radiation oncologists are asked to choose between the autoplanning treatment plan and the clinical treatment plan. In most cases the physician prefered the autoplanning treatment plan.
Conclusion: The results of the studies provided in this literature study show promising results and conclude that autoplannig has the potential to improve the quality, consistency and efficiency of radiotherapy treatment planning for HNC.
Key words: Automated Treatment Planning, Photon Beam Radiotherapy, Head and Neck Cancer

1 Introduction

Head and neck cancer (HNC) is a type of cancer that affects the head and neck region, including the oral cavity, larynx, and pharynx. The pharynx is subdivided in a high, middle and low part, the nasopharynx, oropharynx and hypopharynx, respectively. The HNC region has a complex anatomy with various important organs that needs to be preserved during cancer treatment [1, 2, 3, 4]. Radiotherapy is one of the main treatment options for this type of cancer. The workflow of radiotherapy starts with imaging the patient with a CT-scanner. The images are then used to precisely segment the patient's tumor and other (healthy) organs. The information from the imaging and segmentation is used to create a patient specific treatment plan. Treatment planning is a crucial step in the delivery of radiotherapy. The outcome of the treatment plan will directly affect the outcome of the patient's treatment [2, 3, 4]. The accuracy that is needed for making an adequate treatment plan, due to the complex anatomy and many important organs, makes this a time consuming and labor-intensive task. Manually planning a clinically acceptable treatment plan for a tumor in the HNC region can take up to half a day, sometimes even longer.
Generally, in treatment planning there is a wide solution space, i.e. there are many clinically ac-

ceptable plans possible and there is not one right solution [5]. This immediately raises the question: how does a radiation therapy technologist (RTT) know when the treatment plan is optimal? As mentioned before, in the HNC region the anatomy is complex and it is very likely that the target is close to an organ at risk (OAR). The goal of radiotherapy treatment is to deliver maximum dose of radiation to the tumor while minimizing dose to the surrounding OARs. By generating a treatment plan manually it causes variation among RTTs due to difference in planning skills and the limitation in time [6]. Furthermore, manual planning can cause variations due to institute specific protocols in radiotherapy, specific patient anatomy and preferences of the radiation oncologist [2, 3, 5, 7]. Automated treatment planning (autoplanning) is expected to minimize the variations between treatment plan, and to generate good treatment plans faster, which allows more time for the RTTs to spend on patient care or other important tasks.

Different models and algorithms have been developed to help the RTTs and improve the quality and consistency of treatment plans, e.g. atlas-based models, knowledge-based models and deep learning [2, 6]. These models do not only occur in the treatment planning, but in more parts of the radiotherapy workflow. For example, the segmentation of the tumor and OARs can also be done automatically by some of the available treatment planning system (TPS). Automated segmentation is outside the scope of this literature study.

The goal of this literature study is to investigate how autoplanning can improve radiotherapy and to provide an overview of the available techniques and models for autoplanning. The research question is how autoplanning will contribute to the quality, consistency and efficiency of photon beam radiotherapy treatment plans for patients with HNC.

2 Method

A comprehensive search of the literature was conducted using the PubMed and Google Scholar database. The literature study has been structured according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (2020).
The Leiden University Medical Center (LUMC) specifically requested a literature study on autoplanning in the HNC region due to the complexity of these treatment plans. Therefore; autoplanning in radiotherapy, specifically for HNC treatment plans, will be the focus of this literature study. Articles about autoplanning for other types of cancer are not included for quantitative and qualitative results of this study.

2.1 Literature Search

In PubMed the following search query was used: 'Automated' AND 'Planning' AND 'Radiotherapy' AND 'Head and Neck'. The search query was limited by articles from the last ten years. This has been chosen to prevent getting outdated results. Autoplanning is a relatively new development. Any autoplanning techniques older than ten years are most certainly no longer relevant to the current time. After limiting the articles on the past ten years and English written only, this query resulted in 188 results. The search was specified by removing papers about automated segmentation, online adaptive radiotherapy, proton therapy and Magnetic Resonance Imaging (MRI), which resulted in fifty remaining articles. Reading the abstract of the remaining articles led to taking out sixteen more articles that did not meet the specification criteria (results of the articles were e.g. about other cancer sites or were not obtained by autoplanning). Additionally, eight articles collected in another way were added to the selection. These articles were found on Google Scholar with the same search query, but did not emerge from the PubMed search. Finally, the brochures and white-papers of commercially available autoplanning system are consulted to explain how these systems work e.g. for Auto-Planning, RapidPlan and RayStation [8, 9, 10].

2.2 Structure

This study will first address the various autoplanning systems and explain their key concepts. Then the quantitative and qualitative results of comparing the autoplan with clinical plan[1] for HNC patients will be provided and interpreted. Subsequently, the results and the use of autoplanning in radiotherapy will be discussed. This literature study will be finalized with a conclusion.

[1] In this literature study the *clinical plan* always refers to the clinically accepted plan for a patient or the plan that has been used in radiotherapy treatment.

3 Results

There are many studies conducted on automated treatment planning. The goal is to overcome the stumbling points of manual treatment planning, e.g. to create treatment plans of high quality and consistency in a more efficient and quick way. In the last years autoplanning has evolved from relatively 'simple' atlas-based autoplanning to machine- and deep learning-based autoplanning [2].

3.1 Manual Treatment Planning

When a new patient comes in at the radiotherapy department, the workflow starts with imaging the patient. When the CT-scan of the patient is made, the segmentation of the target and OARs is done manually or automatically. After the segmentation the plan for treatment delivery can be formed. The current way of treatment delivery in radiotherapy is by IMRT or VMAT. Both techniques are intensity modulated and consist of many beam segments. In treatment planning it is the goal to establish optimal patient specific beam parameters, which means there needs to be enough dose in the planning target volume (PTV) and minimal dose in the OARs. The most common approach of making a good treatment plan is to work with a class solution. This is a standard list of optimization objectives and weights for the tumor and the OARs on a specific tumor site. The treatment plan is generated by inverse treatment planning, a multi-objective optimization of treatment plan parameters (i.e. beam settings), using radiation dose goals and criteria to target regions and OARs, resulting in a deliverable treatment plan. Creating a treatment plan is done by minimizing the cost function. Radiation treatment planning is performed on a radiotherapy planning CT. This CT is calibrated to allow conversion of the Hounsfield units to electron densities to enable accurate dose calculation, and patients can be scanned in their final treatment position. After wards, it will be checked whether all the clinical goals are met. The plan can be optimized even further by the RTT changing the objective function weights and add objectives or contours when necessary. This iterative trial-and-error process will continue until the RTT and radiation oncologist are satisfied. Thus, a lot of time and experience is required here.

3.2 Erasmus i-Cycle

i-Cycle is a treatment planning system developed at the Erasmus Medical Center in Rotterdam, The Netherlands. Erasmus i-Cycle generates optimized radiotherapy treatment plans, also applicable for HNC. i-Cycle is based on an iterative process that starts with a wish list of objective functions. The system then generates a treatment plan based on the multi-objective optimisation. The priority of the objectives is defined by the user and i-Cycle follows this prioritisation. Subsequently the plan is optimized and the output is a Pareto optimal plan [11, 12, 13]. A Pareto optimal plan is reached when there is no more improvement possible for any of the objectives, without it being at the expense of any other objective or constraint [11].

3.3 Atlas-based autoplanning

Atlas-based autoplanning involves using a pre-existing *atlas* of the HNC region to generate a treatment plan. The database, also known as the atlas, is created from a large data set of a previously treated patient population with similar cases, e.g. tumor size and location. This atlas will guide the treatment planning process by generating an initial treatment plan for a new patient based on plans of similar patients in the atlas. The new patient's CT-scan will be registered to the best matching reference patient's CT-scan from the atlas by deformable image registration, to generate a dose to the new patient. This initial plan is likely to be suboptimal, but is used as an input for the optimization step in the treatment planning workflow. Optimization can be done mathematically and/or manually, depending on the available TPS. After optimization the final treatment plan will have a deliverable dose and can be used for treatment after approval of the radiation oncologist [14, 15, 16].

3.4 Knowlegde-based autoplanning

Similar to other technological fields, the use of artificial intelligence (AI) is emerging in radiotherapy. In different steps of the radiotherapy workflow the use of AI is embraced, e.g. target and OAR segmentation and treatment planning [1, 3]. Machine learning is a subfield of AI that involves the development of algorithms that can learn from and make predictions based on data. Thus it is very important to provide the machine learning with good quality data. Ultimately, it is

also the limitation of machine learning when the training data contains missing data, duplicates, noise, etc. [17]. Machine learning algorithms are used in autoplanning in radiotherapy with the same goal: to improve the quality, consistency and efficiency of the treatment planning [18].
Unlike atlas-based autoplanning the system does not register a new patient to a pre-exciting treatment plan. Knowledge-based autoplanning in radiotherapy uses prior knowledge and clinical experience to guide the treatment planning process. This is done by including information about treatment techniques in a specific cancer site, target volumes and dose constraints. Through the existing data, the system is able to predict dose volume histograms (DVH), dose metrics or dose to individual voxels for new patients [13, 19]. Using knowledge-based autoplanning on a new patient means that the CT image with segmentation of the patient is compared to the knowledge-based system. In this system the treatment settings and outcome of similar patients is known and, by using machine learning, the system can predict an initial treatment plan for new patients. The initial plan needs to be optimized to meet the objectives and constraint for the patient's specific case. Similar to atlas-based autoplanning the optimization can be done manually and automatically depending on the TPS.

The difference between atlas-based and knowledge-based autoplanning is not quite clear described in the literature, it seems that the concepts of atlas- and knowledge-based autoplanning are used interchangeably. Y. Ge and Q.J. Wu (2019) devide knowledge-based autoplanning in two major categories: (1) case- or atlas-based methods and (2) statistical modeling and machine learning methods [20]. They explain the first method as finding similar patients based on similarity measures, e.g. clinical stage, tumor location, OVH values. Finding the similarity is done with and without machine learning. In the second method, statistical and machine learning approaches use the data of previous patients to create a prediction model. A new patient is used as input in the prediction model rather than finding a similar patient.
In addition, studies describe atlas-based autoplanning without the use of machine learning [15], while the same author one year later described atlas-based autoplanning based on machine learning [14]. This once more shows that the differences between the two autoplanning methods are not entirely clear described in the literature.

3.5 Deep Learning Algorithms

Deep learning is a subset of machine learning that utilizes multi-layer neural networks with hidden layers to extract features from the input dataset. The idea behind the algorithm is to mimic the working principle of the human brain. Meaning in this case that deep learning autoplanning can behave and create treatment plans like an experienced human RTT [2, 3, 18]. By learning from a big data set, the deep learning algorithm could predict patient specific outcomes of treatment plans and contributes in optimizing treatment plans.
A way to work with deep learning in treatment planning is to let the deep learning algorithm predict DVHs. The main idea behind this approach is to use a data set of previously treated patients and their corresponding DVHs to learn the relationship between the patient's imaging data and the dose distribution [21]. The DVHs show what amount of radiation will be received in the target and OARs. DVHs are an important tool in radiotherapy that are used in evaluation and optimization of treatment plans. In autoplanning deep learning based DVHs assists inverse planning by automatically creating constraints that guide the treatment planning process [22].
Deep learning algorithms could also be used in the prediction of dose distribution or treatment response for new patients. Commercial available autoplanning system from RayStation uses deep learning dose prediction, this will be presented in the following subsection of this study.

3.6 Commercial Autoplanning Systems

Most literature concerns in-house self-made algorithms that are used as autoplanning systems. Currently, there are commercially available systems as well. In the upcoming part the autoplanning systems of various companies that focus on radiotherapy are explained, namely Auto-Planning from Philips, RapidPlan from Varian and Raystation's deep learning autoplanning from RaySearch Laboratories.

3.6.1 Auto-Planning

Pinnacle Evolution (Philips Radiation Oncology Systems) is a commercially available TPS. In this TPS there is an Auto-Planning (AP) module that simplyfies and accelerates the inverse planning process with algorithms. The goal of AP is sum-

marized in the following key aspects: improving quality, consistency and efficiency [9, 23]. The automated planning workflow is a three step process after imaging the patient and segmenting the tumor and OARs:

1. *Select a treatment technique:* During initial setup the user defines optimization goals (constraints and objectives) for different treatment techniques. This results in a library of techniques containing manually-entered parameters. This library can be edited at any time [9].

2. *Run AP:* The AP system optimizes target coverage and OAR sparing. An iterative algorithm mimics the planning workflow of the RTT by automatically adding and adjusting objectives, constraints and dose shaping contours to reach the clinical goals. E.g., contours for hot spots and cold spots are automatically drawn and objectives added to diminish them. [9, 23, 24, 25].

3. *Evaluate with scorecard:* For the evaluation of the treatment plan a scorecard with the clinical goals on them can be used. Scorecards promote standardization in the approval of plans [9].

Finally, AP has a *no-compromise* setting that allow the user to prioritise between target and OAR, e.g. sparing the spinal cord over target coverage [23, 25]. AP improves the efficiency since AP can run (step 2) on the background, thus the RTT can continue with other tasks.

3.6.2 RapidPlan

RapidPlan is a module in the Eclipse TPS (Varian Medical Systems). RapidPlan is a knowledge-based automated planning system. This means that the system contains a library of previous treatment plans from institutes that trains the machine learning model [10, 24]. The input is a large data set of treatment plans, including information about the patient's anatomy, the tumor and treatment settings. The machine learning is trained to recognize patterns between the different parameters so the algorithm can predict the optimal treatment plan for a specific patient.
In RapidPlan the machine learning model estimated the DVHs and generate an initial treatment plan, which are both input for the optimization algorithm. The ability for manual adjustments is still provided in RapidPlan, so the RTTs or radiation oncologists can refine the plan if needed.

The key aspects of RapidPlan are the same as in Pinnacle's AP and autoplanning in general: better quality, consistency and efficiency. The brochure of RapidPlan states that in their research 95% of patients had clinically acceptable plans and 53% of patients had an significantly improved treatment plan. Finally, the overall planning time was reduced by 95% [10].

3.6.3 Raystation

RayStation is a commercially available TPS for radiotherapy that is developed by RaySearch Laboratories [3]. RayStation offers a deep learning autoplanning model (DLA) for photon radiotherapy supporting IMRT and VMAT. They have models specialized for different tumor sites, including oropharynx. In figure 1 the schematic workflow of RayStation DLA is displayed. To generate a clinical acceptable treatment plan, the DLA workflow in RayStation follows four general steps:

1. *Imaging & contouring:* The DLA system needs the CT image of the patient and the segmentation as input. The segmentation of the tumor and the OARs can be done manually or automatically since RayStation also has a Deep Learning Segmentation module. The DLA has several mandatory structures that need to exist to start the prediction of the dose distribution.

2. *Dose distribution prediction:* The DLA system predicts the dose distribution based on the image data and the segmentation. For the HNC model the DLA system is trained in collaboration with the Princess Margaret Cancer Center in Toronto (Canada) who provided the patient data. The underlying architecture of the oropharynx DLA model is a U-Net and is validated for VMAT [26, 27]. For this specific cancer site the primary prescription is 7000 cGy and a second dose level (in the electives nodes) of 5425 cGy in 35 fractions. The protocol of the model consists of a number of Regions Of Interests (ROI) with their corresponding clinical goals [27].

3. *Beams:* The beam set-up is set in the model setting and can be selected by a beam template.

4. *Dose mimicking:* In the final part the predicted dose will be optimized to a deliverable dose. The predicted dose itself is not

deliverable as it is not the result of a treatment plan. A deliverable treatment plan will be generated in an optimization of objectives for targets and OARs. The resulting deliverable dose will be compared with the predicted dose. Minimizing the difference between the predicted and deliverable dose will automatically be added to the objective function of the optimization problem. This encourages the deliverable dose to be similar to the predicted dose.

3.7 Plan Evaluation

There are many studies conducted on autoplanning for HNC patients. To assess the quality of the autoplans, they are compared with the clinical plan. In most studies the patient's data is anonymous and the clinical plan is preserved. Before the outcome of the autoplans can be compared to the clinical plan, it is important to understand how the autoplan is evaluated. Multiple studies used dosimetric parameters to evaluate the performance: PTV coverage, maximum dose in serial OARs, mean dose in parallel OARs, and the required time to generate a plan [5, 24].
Another way of evaluating a treatment plan is by a blind comparison by radiation oncologists. In other words, which treatment plan will the doctor choose: the manual plan or the autoplan? The qualitative results from articles that did the blind comparison are presented separately later on.

3.7.1 Plan Objectives

Plan objectives and clinical goals are necessary for plan optimization. These parameters will also help to evaluate the autoplans compared to the clinical plan. In photon beam radiotherapy of HNC it is common to have two dose levels of approximately 70 Gy and 54 Gy for the primary tumor and the elective glands, respectively, in approximately 35 fractions. In the literature the PTV dose coverage is commonly defined as V95% $\geq$ 95%, or the volume that receives 95% of the prescribed dose should be equal to or higher than 95%. In the studies that are consulted for this literature study the exact objectives and clinical goals for the OARs differ per study, institute and research group. Therefore, the results of the studies will not be compared against each other.

3.7.2 Quantitative Results

In 2017 Kusters et al. from the Radboud University Medical Center conducted a study where they evaluated twenty automated IMRT plans with the clinical plan [23]. Pinnacle's AP was used to generate the autoplans. This study shows that the PTV coverage was similar to the clinical plan, while the sparing of OAR was better for the contralateral parotid gland, contralateral submandibular gland, larynx, mandible and brain-stem. The treatment planning time in this particular study was reduced from 1,5 - 3 hours to less than 1 hour.

Similar results are published by Cilla et al. (2021) with AP from Pinnacle [28]. In this study radiotherapy was delivered with VMAT with simultaneously integrated boost (SIB) to fifteen HNC patients. This resulted in a similar PTV coverage compared to the clinical plan and a better sparring of the OARs, especially the spinal cord, brain stem and parotids. The dose reduction in these structures were 13-15%, 9% and 16%, respectively. The overall treatment planning time was less than 30 minutes.

Other autoplanning systems were also used in HNC treatment planning. The study conducted by Tol et al. (2015) compared the clinical plan of fifteen HNC patients to RapidPlan plans of three different models [19]. Two models (30A and 30B) have different compositions of plan libraries and a third model (60) consists of plans from of both

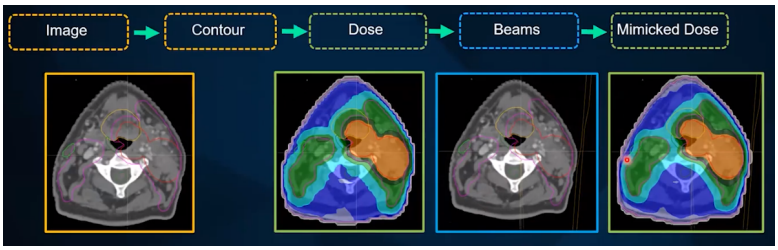


Figure 1: Schematic overview of the deep learning autoplanning process. Source: RaySearch Labatories

libraries. First of all, all the knowledge-based plans were reviewed by a senior HNC specialized radiation onclogisist and the plans were considered satisfactory. In the autoplans PTV coverage and homogeneity index (HI) was improved, but this was not a significant difference. The OAR sparing was slightly better than the clinical plans, e.g. the mean dose in the swallowing muscles was lower in the autoplans. This result is shown in the histogram in figure 2, representing the mean dose in the composite swallowing muscles for the three types of autoplans and the clinical plan.

A study conducted by Kaderka et al. (2019) achieved a similar result for their 52 HNC patients, namely a dose reduction in the OARs [29]. They specifically mention that in their study the dosimetric cost of a lower dose in the OAR meant a small increase in the high dose level PTV coverage and a decrease in the low dose level PTV coverage. Nevertheless, these differences were still clinically acceptable.

In a multi-institute planning study from Krayenbuehl et al. (2018) the results of different autoplanning systems are compared, including AP, RapidPlan and RayStation [24]. In this study they compared sixteen randomly chosen HNC patients from two institutes on PTV coverage, mean and maximum dose to OARs and the planning time. All the autoplanning systems were able to achieve the PTV dose constraint and the constraint for the serial OARs. For the parallel OARs on the other hand, AP was ranked best, followed by RayStation and RapidPlan. When considering the individual parallel OARs, the mean dose in the oral cavity was the lowest in RayStation. Furthermore, the overall planning time was ranked fastest in RapidPlan, followed by AP and RayStation.

3.7.3 Qualitative Results

Cilla et al. (2021) did not only compare the clinical plans and autoplans on dose parameters, but also did a blind test with two radiation oncologists [28]. In 80% of the cases they chose the autoplan over the clinical plan.

In the article of Hansen et al. (2016) qualitative research was conducted in Danish University Hospital [30]. They report promising result that from 29 out of the 30 patients the autoplan was chosen over the manual treatment by three senior radiation oncologists. Quantitatively, the OARs where spared better, varying from 0.5 to 6.5 Gy reduction. The treatment planning time was halved when using AP in Pinnacle.

An even broader study was done by Olanrewaju et al. (2021) on fifty HNC patients [31]. Their treatment plans were re-created by the Radiation Planning Assistant (RPA), which is a web-based service developed by the MD Anderson Cancer Center, University of Texas. The plan optimization and dose calculation is performed by Eclipse TPS (Varian Medical Systems). Finally, the plans were reviewed by fourteen different radiation oncologists from different institutes from the USA. In terms of DVH metrics for coverage and OAR constraints the autoplans were quite similar to the clinical plan. The physicians reviewed both the clinical plan and the autoplan and they were found useful in 78% and 88% of the cases, respectively. After asking for a preference 27% of the reviewers chose the clinical plan, 47% chose the autoplan and 25% stated that the plans were equivalent.

4 Discussion

In most of the cases the evaluation of the autoplan showed a similar or slightly better treatment plan

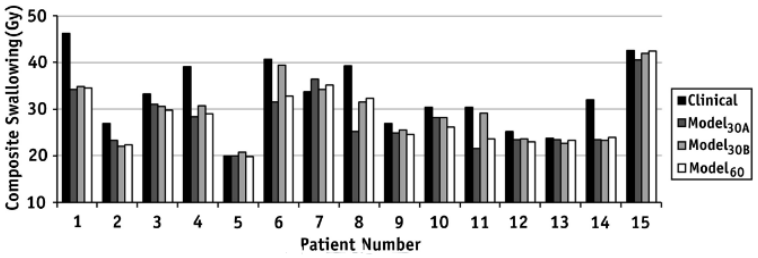


Figure 2: Histogram of mean dose to the composite swallowing muscles for 3 RapidPlan knowledge-based automated plans and the clinical plan [19].

in comparison with the clinical plan. In general the sparing of OAR is better in the autoplans. The results from the studies that are used in this paper all resulted from institute specific data. This means that the autoplans are compared with clinical plans from the same institute. This makes it impossible to compare the results of the different studies against each other. Separately, the studies show good results, but comparing the data will provide a distorted view on the results.

Furthermore, the results are interpreted as if the clinical plan is the ground truth. Meaning that the autoplans are evaluated relative to the clinical plan, but it might be the case that not all the clinical plans were the best possible solutions. If the clinical plan was manually planned by another RTT or approved by another radiation oncologist, the results might have been different. After all, it has been said that one of the reasons to use autoplanning comes from the desire for improved consistency in treatment plans.

The final discussion point is the chance of publication bias. In the articles the researchers probably only publish the best and or mention-worthy results. Thus it seems like tumor coverage in autoplanning is improving or similar, the dose in the OAR is reduced and the overall treatment time is decreased in all of the cases, but this is only true for the *published* results.

5 Conclusion

Autoplanning has the potential to greatly improve the quality, consistency and efficiency of treatment planning for patients with HNC, e.g. by reducing the time required to create a treatment plan and limiting variations between treatment plans. The use of i-Cycle, atlas-based planning, knowledge-based planning and machine learning algorithms are some of the techniques used in autoplanning systems. Commercially available TPS like Pinnacle, Ecplise and RayStation offer autoplanning modules as well. Multiple studies have reported the benefits of these systems, including a similar or slightly improved PTV coverage while sparing the OARs better in comparison to the clinical plan.

In conclusion, it seems that autoplanning systems positively contribute to the quality, consistency and efficiency of treatment planning in photon beam radiotherapy for patients with HNC, when applied to previously treated patients. Needless to say, the final approval of the patient's treatment plan before actual radiation treatment should stay with the radiation oncologist and medical physicist.

References

[1] A. F. Osman and N. M. Tamam, "Attention-aware 3d u-net convolutional neural network for knowledge-based planning 3d dose distribution prediction of head-and-neck cancer," *Journal of Applied Clinical Medical Physics*, vol. 23, no. 7, p. e13630, 2022.

[2] M. Wang, Q. Zhang, S. Lam, J. Cai, and R. Yang, "A review on application of deep learning algorithms in external beam radiotherapy automated treatment planning," *Frontiers in oncology*, p. 2177, 2020.

[3] C. Wang, X. Zhu, J. C. Hong, and D. Zheng, "Artificial intelligence in radiotherapy treatment planning: present and future," *Technology in cancer research & treatment*, vol. 18, 2019.

[4] L. Q. Chow, "Head and neck cancer," *New England Journal of Medicine*, vol. 382, no. 1, pp. 60–72, 2020.

[5] I. Hazell, K. Bzdusek, P. Kumar, C. R. Hansen, A. Bertelsen, J. G. Eriksen, J. Johansen, and C. Brink, "Automatic planning of head and neck treatment plans," *Journal of Applied Clinical Medical Physics*, vol. 17, pp. 272–282, 2016.

[6] Z. Ouyang, Z. Liu Shen, E. Murray, M. Kolar, D. LaHurd, N. Yu, N. Joshi, S. Koyfman, K. Bzdusek, and P. Xia, "Evaluation of auto-planning in imrt and vmat for head and neck cancer," *Journal of applied clinical medical physics*, vol. 20, no. 7, pp. 39–47, 2019.

[7] M.-C. Biston, M. Costea, F. Gassa, A.-A. Serre, P. Voet, R. Larson, and V. Grégoire, "Evaluation of fully automated a priori mco treatment planning in vmat for head-and-neck cancer," *Physica Medica*, vol. 87, pp. 31–38, 2021.

[8] RaySearch Laboratories AB (publ), "Machine learning - automated treatment planning," *P.O. Box 3297, SE-103 65 Stockholm, Sweden*, 2019. https://www.raysearchlabs.com/siteassets/about-overview/media-center/wp-re-ev-n-pdfs/white-papers/whitepaper_ml_automatedplanning_raystation.pdf.

[9] Philips Radiation Oncology, "Personalized therapy planning - philips pinnacle evolution," *Koninklijke Philips N.V.*, 2020.

[10] Varian - Siemens Healthineers Company, "Rapidplan - unleash the power of your data with knowledge-based planning," *Varian Medical Systems, Inc.*, 2022.

[11] R. Van Haveren, W. Ogryczak, G. M. Verduijn, M. Keijzer, B. J. Heijmen, and S. Breedveld, "Fast and fuzzy multi-objective radiotherapy treatment plan generation for head and neck cancer patients with the lexicographic reference point method (lrpm)," *Physics in Medicine & Biology*, vol. 62, no. 11, p. 4318, 2017.

[12] R. Van Haveren, B. J. Heijmen, and S. Breedveld, "Automatic configuration of the reference point method for fully automated multi-objective treatment planning applied to oropharyngeal cancer," *Medical Physics*, vol. 47, no. 4, pp. 1499–1508, 2020.

[13] Y. Wang, B. J. Heijmen, and S. F. Petit, "Knowledge-based dose prediction models for head and neck cancer are strongly affected by interorgan dependency and dataset inconsistency," *Medical physics*, vol. 46, no. 2, pp. 934–943, 2019.

[14] C. McIntosh, M. Welch, A. McNiven, D. A. Jaffray, and T. G. Purdie, "Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method," *Physics in Medicine & Biology*, vol. 62, no. 15, p. 5926, 2017.

[15] C. McIntosh and T. G. Purdie, "Voxel-based dose prediction with multi-patient atlas selection for automated radiotherapy treatment planning," *Physics in Medicine & Biology*, vol. 62, no. 2, p. 415, 2016.

[16] V. Kearney, J. W. Chan, G. Valdes, T. D. Solberg, and S. S. Yom, "The application of artificial intelligence in the imrt planning process for head and neck cancer," *Oral Oncology*, vol. 87, pp. 111–116, 2018.

[17] P. Giraud, P. Giraud, A. Gasnier, R. El Ayachy, S. Kreps, J.-P. Foy, C. Durdux, F. Huguet, A. Burgun, and J.-E. Bibault, "Radiomics and machine learning for radiotherapy in head and neck cancers," *Frontiers in oncology*, vol. 9, p. 174, 2019.

[18] P. Meyer, V. Noblet, C. Mazzara, and A. Lallement, "Survey on deep learning for radiotherapy," *Computers in biology and medicine*, vol. 98, pp. 126–146, 2018.

[19] J. P. Tol, A. R. Delaney, M. Dahele, B. J. Slotman, and W. F. Verbakel, "Evaluation of a knowledge-based planning solution for head and neck cancer," *International Journal of Radiation Oncology* Biology* Physics*, vol. 91, no. 3, pp. 612–620, 2015.

[20] Y. Ge and Q. J. Wu, "Knowledge-based planning for intensity-modulated radiation therapy: a review of data-driven approaches," *Medical physics*, vol. 46, no. 6, pp. 2760–2775, 2019.

[21] Z. Liu, X. Chen, K. Men, J. Yi, and J. Dai, "A deep learning model to predict dose–volume histograms of organs at risk in radiotherapy treatment plans," *Medical Physics*, vol. 47, no. 11, pp. 5467–5481, 2020.

[22] A. T. Chang, A. W. Hung, F. W. Cheung, M. C. Lee, O. S. Chan, H. Philips, Y.-T. Cheng, and W.-T. Ng, "Comparison of planning quality and efficiency between conventional and knowledge-based algorithms in nasopharyngeal cancer patients using intensity modulated radiation therapy," *International Journal of Radiation Oncology* Biology* Physics*, vol. 95, no. 3, pp. 981–990, 2016.

[23] J. Kusters, K. Bzdusek, P.Kumar, P. van Kollenburg, M.C.Kunze-Busch, M.Wendling, T. Dijkema, and J. Kaanders, "Automated imrt planning in pinnacle," *Strahlentherapie und Onkologie*, vol. 193, pp. 1031–1038, 2017.

[24] J. Krayenbuehl, M. Zamburlini, S. Ghandour, M. Pachoud, S. Tanadini-Lang, J. Tol, M. Guckenberger, and W. Verbakel, "Planning comparison of five automated treatment planning solutions for locally advanced head and neck cancer," *Radiation Oncology*, vol. 13, no. 1, pp. 1–8, 2018.

[25] J. Krayenbuehl, I. Norton, G. Studer, and M. Guckenberger, "Evaluation of an automated knowledge based treatment planning system for head and neck," *Radiation Oncology*, vol. 10, no. 1, pp. 1–8, 2015.

[26] D. Nguyen, T. Long, X. Jia, W. Lu, X. Gu, Z. Iqbal, and S. Jiang, "A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from

patient anatomy using deep learning," *Scientific reports*, vol. 9, no. 1, p. 1076, 2019.

[27] RaySearch Laboratories AB (publ), "Deep learning planning models," 2022.

[28] S. Cilla, F. Deodato, C. Romano, A. Ianiro, G. Macchia, A. Re, M. Buwenge, L. Boldrini, L. Indovina, V. Valentini, *et al.*, "Personalized automation of treatment planning in head-neck cancer: A step forward for quality in radiation therapy?," *Physica Medica*, vol. 82, pp. 7–16, 2021.

[29] R. Kaderka, R. C. Mundt, N. Li, B. Ziemer, V. N. Bry, M. Cornell, and K. L. Moore, "Automated closed-and open-loop validation of knowledge-based planning routines across multiple disease sites," *Practical Radiation Oncology*, vol. 9, no. 4, pp. 257–265, 2019.

[30] C. R. Hansen, A. Bertelsen, I. Hazell, R. Zukauskaite, N. Gyldenkerne, J. Johansen, J. G. Eriksen, and C. Brink, "Automatic treatment planning improves the clinical quality of head and neck cancer treatment plans," *Clinical and translational radiation oncology*, vol. 1, pp. 2–8, 2016.

[31] A. Olanrewaju, L. E. Court, L. Zhang, K. Naidoo, H. Burger, S. Dalvie, J. Wetter, J. Parkes, C. J. Trauernicht, R. E. McCarroll, *et al.*, "Clinical acceptability of automated radiation treatment planning for head and neck cancer using the radiation planning assistant," *Practical radiation oncology*, vol. 11, no. 3, pp. 177–184, 2021.

7.2 Model protocol

MODEL PROTOCOL

ROI	Clinical goals
PTV_High	At least 6650 cGy dose at 98.0 % volume
PTV_High	At most 7490 cGy dose at 1.8 cm^3 volume
PTV_Low	At least 5320 cGy dose at 98.0 % volume
SpinalCord	At most 4500 cGy dose at 0.0 % volume
Brain	At most 5800 cGy dose at 1.0 cm^3 volume
Brain	At most 6800 cGy dose at 0.0 % volume
Cochlea_L	At most 4500 cGy dose at 50.0 % volume
Cochlea_L	At most 5500 cGy dose at 5.0 % volume
Cochlea_R	At most 4500 cGy dose at 50.0 % volume
Cochlea_R	At most 5500 cGy dose at 5.0 % volume
Parotid_L	At most 2600 cGy dose at 50.0 % volume
Parotid_R	At most 2600 cGy dose at 50.0 % volume
Musc_Constrict_I	At most 5500 cGy dose at 50.0 % volume
Musc_Constrict_M	At most 5500 cGy dose at 50.0 % volume
Musc_Constrict_S	At most 5500 cGy dose at 50.0 % volume
Brainstem	At most 5400 cGy dose at 0.0 % volume
Glnd_Submand_L	At most 3500 cGy dose at 50.0 % volume
Glnd_Submand_R	At most 3500 cGy dose at 50.0 % volume
Cavity_Oral	At most 3000 cGy dose at 50.0 % volume
Esophagus_S	At most 3000 cGy dose at 50.0 % volume
OpticNrv_L	At most 5400 cGy dose at 0.0 % volume
OpticNrv_R	At most 5400 cGy dose at 0.0 % volume
Eye_L	At most 3000 cGy dose at 0.0 % volume
Eye_R	At most 3000 cGy dose at 0.0 % volume
Glnd_Lacrimal_L	At most 2500 cGy dose at 50.0 % volume
Glnd_Lacrimal_R	At most 2500 cGy dose at 50.0 % volume
Glottis	At most 4000 cGy dose at 50.0 % volume
Larynx_SG	At most 4000 cGy dose at 50.0 % volume
Bone_Mandible	At most 7200 cGy dose at 0.0 % volume
Glnd_Lacrimal_R	At most 2500 cGy dose at 50.0 % volume
Glottis	At most 4000 cGy dose at 50.0 % volume
Larynx_SG	At most 4000 cGy dose at 50.0 % volume
Bone_Mandible	At most 7200 cGy dose at 0.0 % volume

Figure 37: Model protocol for RayStation DLAP RSL-OROPHARYNX-700-SIB [28]

7.3 Model overview

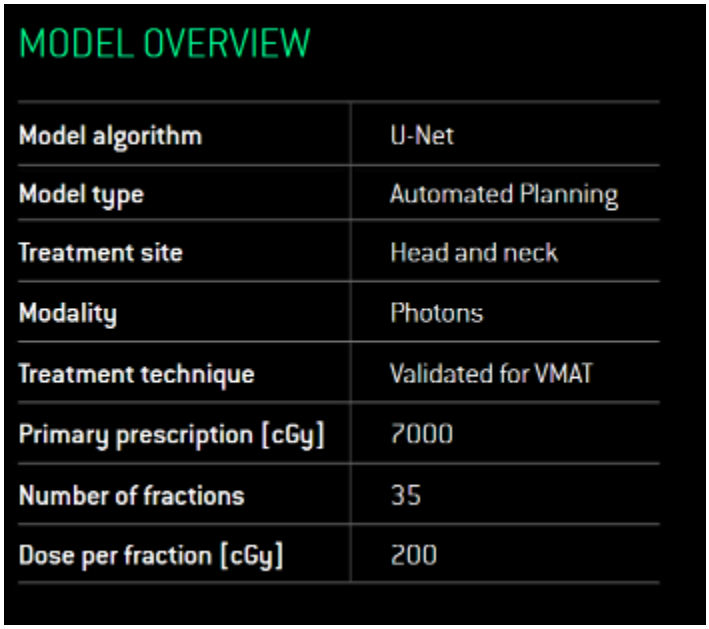


Figure 38: Model overview for RayStation DLAP RSL-OROPHARYNX-700-SIB [28]

7.4 Validation report DLAP RSL-OROPHARYNX-700-SIB

Validation report for model RSL-Oropharynx-7000-SIB (3.0)

Model Overview

Model name	RSL-Oropharynx-7000-SIB (3.0)
Model algorithm	U-Net
Scripting environment name	ML Planning 10B (v3.1.0)
Model type	Automated Planning
Model originator	RaySearch Laboratories
Scripting environment originator	RaySearch Laboratories

Model Validation Data

Body site	HN
Image modality	CT
Image patient position	HFS
Treatment position	HeadFirstSupine
Treatment modality	Photons
Treatment techniques	VMAT (2 arcs, [2.0, 2.0] degree gantry spacing)
Energy	6
Prescribed dose	7000.0
Number of fractions	35
Dose per fraction	200.0

RaySearch Laboratories AB (publ)
Reg. no. 556322-6157
Box 3297
SE-103 65 Stockholm
Sweden

Visitors: Eugeniavägen 18
info@raysearchlabs.com
www.raysearchlabs.com
Phone: +46 (0)8 510 530 00

Required ROIs	CTV_High, CTV_Low, PTV_High, PTV_Low, Brainstem, SpinalCord, Parotid_L, Parotid_R, Esophagus, External, Musc_Constrict_I, Musc_Constrict_M, Musc_Constrict_S, PTV_Low-PTV_High+5mm
Optional ROIs	Glnd_Submand_L, Glnd_Submand_R, Cavity_Oral, Bone_Mandible
Patient Age	Adults
Gender	Male and Female
Treatment intent	Curative
Validation data treatment technique	VMAT, 2 arcs, [2.0, 2.0] degree gantry spacing, 6MV
Other considerations	

Training Data

Number of plans	100
Origin	Princess Margaret Cancer Centre
Training completion date	2021-09-18
Treatment position	HFS
Modality	Photons
Treatment techniques	VMAT
Energy	6
Prescribed dose	7000/5600 cGy (simultaneous integrated boost)
Dose per fraction	200/160 cGy
Type of plans	All plans are clinically approved, peer-reviewed, and used for delivery
Other considerations	

Validation

Number of plans	10
Origin	RaySearch Laboratories contracted data and data part of the OPC-Radiomics dataset[1]
TPS	RayStation
Validation system	RayStation 10B
Validation date	February 11, 2022
Validated by	RaySearch Laboratories

Validation Details

This section describes the validation details.

Validation data set selection

For the validation of the ML planning results ten patients were randomly selected. All relevant OARs were delineated for the patients. The model was validated for the DAHANCA 2020 protocol (arm 2, 35x 2 Gy).

Model was validated on a Elekta Versa machine using 6MV, setup with 2 full arcs with 2-degree gantry spacing. A list of the clinical goals used for validation can be found in the table below.

Clinical Goals

ROI	Clinical goal	Tier
PTV_54.25	At least 5154.0 cGy dose at 98.0 % volume	
PTV_70	At least 6650.0 cGy dose at 98.0 % volume	
PTV_70	At most 7490.0 cGy dose at 1.8 cm^3 volume	
SpinalCord	At most 4500.0 cGy dose at 0.0 % volume	Absolute
Brain	At most 5800.0 cGy dose at 1.0 cm^3 volume	Should
Brain	At most 6800.0 cGy dose at 0.0 % volume	Should
Cochlea_L	At most 4500.0 cGy dose at 50.0 % volume	Should

Cochlea_L	At most 5500.0 cGy dose at 5.0 % volume	Should
Cochlea_R	At most 4500.0 cGy dose at 50.0 % volume	Should
Cochlea_R	At most 5500.0 cGy dose at 5.0 % volume	Should
Parotid_L	At most 2600.0 cGy dose at 50.0 % volume	Should
Parotid_R	At most 2600.0 cGy dose at 50.0 % volume	Should
Musc_Constrict_I	At most 5500.0 cGy dose at 50.0 % volume	Should
Musc_Constrict_M	At most 5500.0 cGy dose at 50.0 % volume	Should
Musc_Constrict_S	At most 5500.0 cGy dose at 50.0 % volume	Should
Brainstem	At most 5400.0 cGy dose at 0.0 % volume	Absolute
Glnd_Submand_L	At most 3500.0 cGy dose at 50.0 % volume	Should
Glnd_Submand_R	At most 3500.0 cGy dose at 50.0 % volume	Should
Cavity_Oral	At most 3000.0 cGy dose at 50.0 % volume	Should
Esophagus_S	At most 3000.0 cGy dose at 50.0 % volume	Should
OpticNrv_L	At most 5400.0 cGy dose at 0.0 % volume	Must
OpticNrv_R	At most 5400.0 cGy dose at 0.0 % volume	Must
Eye_L	At most 3000.0 cGy dose at 0.0 % volume	Must
Eye_R	At most 3000.0 cGy dose at 0.0 % volume	Must
Glnd_Lacrimal_L	At most 2500.0 cGy dose at 50.0 % volume	Must
Glnd_Lacrimal_R	At most 2500.0 cGy dose at 50.0 % volume	Must
Glottis	At most 4000.0 cGy dose at 50.0 % volume	Should
Larynx_SG	At most 4000.0 cGy dose at 50.0 % volume	Should

Bone_Mandible	At most 7200.0 cGy dose at 0.0 % volume	Should
Pituitary	At most 2000.0 cGy dose at 50.0 % volume	Should

Results

Summary

For 10/10 plans acceptable target coverage was achieved. In 10/10 plans all the OAR clinical goals were met for the absolute and must tiers. No plans however passed all of the should tier clinical goals.

Dosimetry

Patient Id	Roi	D99 (cGy)	D98 (cGy)	D95 (cGy)	D50 (cGy)	D2 (cGy)	D1 (cGy)
A	PTV_70	6597	6650	6738	7054	7242	7262
A	PTV_54.25	5166	5206	5266	5565	7193	7220
B	PTV_70	6594	6663	6755	7044	7184	7195
B	PTV_54.25	5140	5190	5245	5470	7143	7164
C	PTV_70	6571	6650	6719	7047	7202	7212
C	PTV_54.25	5137	5180	5237	5496	7154	7180
D	PTV_70	6648	6708	6812	7028	7169	7184
D	PTV_54.25	5176	5209	5262	6101	7143	7163
E	PTV_70	6611	6669	6753	7072	7242	7261
E	PTV_54.25	5153	5196	5255	5506	7188	7214
F	PTV_70	6592	6660	6749	7036	7187	7201
F	PTV_54.25	5151	5192	5245	5472	7139	7163
G	PTV_70	6628	6679	6778	7033	7163	7179

G	PTV_54.25	5129	5168	5223	5455	7114	7136
H	PTV_70	6585	6651	6753	7028	7183	7201
H	PTV_54.25	5121	5168	5234	5468	7123	7149
I	PTV_70	6594	6650	6715	7055	7219	7235
I	PTV_54.25	5117	5174	5243	5487	7134	7166
J	PTV_70	6590	6650	6728	7042	7203	7220
J	PTV_54.25	5115	5177	5243	5486	7129	7161

Unfulfilled clinical goals

No plans passed all the “Should” tier clinical goals due to varying degrees of PTV overlap. For all cases most of the failing OAR clinical goals such as glottic larynx, supraglottic larynx, submandibulars and PCMs had a large overlap with the PTVs which made them impossible to spare without sacrificing target coverage. OARs with less overlap with the PTVs such as the esophagus and parotids still have some overlap which complicates sparing. However, sparing should be possible with some manual post processing of the plans or using a specialized strategy.

RSL-Oropharynx-7000-SIB-Validation-A

PTV_54.25	At least a conformity index of 0.9 at 5154.0 cGy isodose	**0.79 CI**
Parotid_L	At most 2600.0 cGy dose at 50.0 % volume	4592 cGy
Parotid_R	At most 2600.0 cGy dose at 50.0 % volume	3960 cGy
Musc_Constrict_I	At most 5500.0 cGy dose at 50.0 % volume	7092 cGy
Musc_Constrict_M	At most 5500.0 cGy dose at 50.0 % volume	7051 cGy
Musc_Constrict_S	At most 5500.0 cGy dose at 50.0 % volume	7111 cGy
Glnd_Submand_L	At most 3500.0 cGy dose at 50.0 % volume	7052 cGy
Glnd_Submand_R	At most 3500.0 cGy dose at 50.0 % volume	6759 cGy
Cavity_Oral	At most 3000.0 cGy dose at 50.0 % volume	3353 cGy
Esophagus_S	At most 3000.0 cGy dose at 50.0 % volume	4969 cGy

Glottis	At most 4000.0 cGy dose at 50.0 % volume	6975 cGy
Larynx_SG	At most 4000.0 cGy dose at 50.0 % volume	7052 cGy

RSL-Oropharynx-7000-SIB-Validation-B

PTV_54.25	At least a conformity index of 0.9 at 5154.0 cGy isodose	**0.86 CI**
Musc_Constrict_I	At most 5500.0 cGy dose at 50.0 % volume	6834 cGy
Musc_Constrict_M	At most 5500.0 cGy dose at 50.0 % volume	6663 cGy
Musc_Constrict_S	At most 5500.0 cGy dose at 50.0 % volume	6028 cGy
Glnd_Submand_L	At most 3500.0 cGy dose at 50.0 % volume	5543 cGy
Glnd_Submand_R	At most 3500.0 cGy dose at 50.0 % volume	7044 cGy
Cavity_Oral	At most 3000.0 cGy dose at 50.0 % volume	4973 cGy
Esophagus_S	At most 3000.0 cGy dose at 50.0 % volume	4503 cGy
Glottis	At most 4000.0 cGy dose at 50.0 % volume	6910 cGy
Larynx_SG	At most 4000.0 cGy dose at 50.0 % volume	7038 cGy

RSL-Oropharynx-7000-SIB-Validation-C

PTV_54.25	At least a conformity index of 0.9 at 5154.0 cGy isodose	**0.82 CI**
Musc_Constrict_M	At most 5500.0 cGy dose at 50.0 % volume	5553 cGy
Musc_Constrict_S	At most 5500.0 cGy dose at 50.0 % volume	6894 cGy
Glnd_Submand_L	At most 3500.0 cGy dose at 50.0 % volume	6150 cGy
Glnd_Submand_R	At most 3500.0 cGy dose at 50.0 % volume	6862 cGy
Cavity_Oral	At most 3000.0 cGy dose at 50.0 % volume	4700 cGy
Esophagus_S	At most 3000.0 cGy dose at 50.0 % volume	3772 cGy
Glottis	At most 4000.0 cGy dose at 50.0 % volume	4843 cGy

Larynx_SG	At most 4000.0 cGy dose at 50.0 % volume	5375 cGy

RSL-Oropharynx-7000-SIB-Validation-D

PTV_54.25	At least a conformity index of 0.9 at 5154.0 cGy isodose	**0.8 CI**
Parotid_L	At most 2600.0 cGy dose at 50.0 % volume	4667 cGy
Parotid_R	At most 2600.0 cGy dose at 50.0 % volume	3810 cGy
Musc_Constrict_M	At most 5500.0 cGy dose at 50.0 % volume	5931 cGy
Musc_Constrict_S	At most 5500.0 cGy dose at 50.0 % volume	6589 cGy
Glnd_Submand_L	At most 3500.0 cGy dose at 50.0 % volume	7048 cGy
Glnd_Submand_R	At most 3500.0 cGy dose at 50.0 % volume	5816 cGy
Cavity_Oral	At most 3000.0 cGy dose at 50.0 % volume	7031 cGy
Esophagus_S	At most 3000.0 cGy dose at 50.0 % volume	4977 cGy
Glottis	At most 4000.0 cGy dose at 50.0 % volume	5977 cGy
Larynx_SG	At most 4000.0 cGy dose at 50.0 % volume	7048 cGy

RSL-Oropharynx-7000-SIB-Validation-E

PTV_54.25	At least a conformity index of 0.9 at 5154.0 cGy isodose	**0.8 CI**
Parotid_L	At most 2600.0 cGy dose at 50.0 % volume	2774 cGy
Musc_Constrict_M	At most 5500.0 cGy dose at 50.0 % volume	6982 cGy
Musc_Constrict_S	At most 5500.0 cGy dose at 50.0 % volume	5583 cGy
Glnd_Submand_L	At most 3500.0 cGy dose at 50.0 % volume	6967 cGy
Glnd_Submand_R	At most 3500.0 cGy dose at 50.0 % volume	6339 cGy
Cavity_Oral	At most 3000.0 cGy dose at 50.0 % volume	5199 cGy
Glottis	At most 4000.0 cGy dose at 50.0 % volume	4354 cGy

Larynx_SG	At most 4000.0 cGy dose at 50.0 % volume	7064 cGy

RSL-Oropharynx-7000-SIB-Validation-F

PTV_54.25	At least a conformity index of 0.9 at 5154.0 cGy isodose	**0.85 CI**
Musc_Constrict_M	At most 5500.0 cGy dose at 50.0 % volume	6288 cGy
Musc_Constrict_S	At most 5500.0 cGy dose at 50.0 % volume	6505 cGy
Glnd_Submand_L	At most 3500.0 cGy dose at 50.0 % volume	6699 cGy
Glnd_Submand_R	At most 3500.0 cGy dose at 50.0 % volume	6917 cGy
Cavity_Oral	At most 3000.0 cGy dose at 50.0 % volume	5597 cGy
Esophagus_S	At most 3000.0 cGy dose at 50.0 % volume	3832 cGy
Glottis	At most 4000.0 cGy dose at 50.0 % volume	4871 cGy
Larynx_SG	At most 4000.0 cGy dose at 50.0 % volume	7067 cGy

RSL-Oropharynx-7000-SIB-Validation-G

PTV_54.25	At least a conformity index of 0.9 at 5154.0 cGy isodose	**0.82 CI**
Parotid_L	At most 2600.0 cGy dose at 50.0 % volume	4544 cGy
Parotid_R	At most 2600.0 cGy dose at 50.0 % volume	5393 cGy
Glnd_Submand_L	At most 3500.0 cGy dose at 50.0 % volume	5469 cGy
Glnd_Submand_R	At most 3500.0 cGy dose at 50.0 % volume	6833 cGy
Cavity_Oral	At most 3000.0 cGy dose at 50.0 % volume	3733 cGy
Esophagus_S	At most 3000.0 cGy dose at 50.0 % volume	4031 cGy
Glottis	At most 4000.0 cGy dose at 50.0 % volume	4497 cGy
Larynx_SG	At most 4000.0 cGy dose at 50.0 % volume	5141 cGy

RSL-Oropharynx-7000-SIB-Validation-H

PTV_54.25	At least a conformity index of 0.9 at 5154.0 cGy isodose	**0.78 CI**
Musc_Constrict_M	At most 5500.0 cGy dose at 50.0 % volume	5620 cGy
Glnd_Submand_L	At most 3500.0 cGy dose at 50.0 % volume	6304 cGy
Glnd_Submand_R	At most 3500.0 cGy dose at 50.0 % volume	3959 cGy
Larynx_SG	At most 4000.0 cGy dose at 50.0 % volume	6978 cGy

RSL-Oropharynx-7000-SIB-Validation-I

PTV_54.25	At least a conformity index of 0.9 at 5154.0 cGy isodose	**0.79 CI**
Glnd_Submand_R	At most 3500.0 cGy dose at 50.0 % volume	5592 cGy
Larynx_SG	At most 4000.0 cGy dose at 50.0 % volume	4458 cGy

RSL-Oropharynx-7000-SIB-Validation-J

PTV_54.25	At least a conformity index of 0.9 at 5154.0 cGy isodose	**0.78 CI**
Glnd_Submand_L	At most 3500.0 cGy dose at 50.0 % volume	5796 cGy
Glnd_Submand_R	At most 3500.0 cGy dose at 50.0 % volume	5616 cGy
Cavity_Oral	At most 3000.0 cGy dose at 50.0 % volume	3587 cGy
Larynx_SG	At most 4000.0 cGy dose at 50.0 % volume	4394 cGy
Bone_Mandible	At most 7200.0 cGy dose at 0.0 % volume	7211 cGy

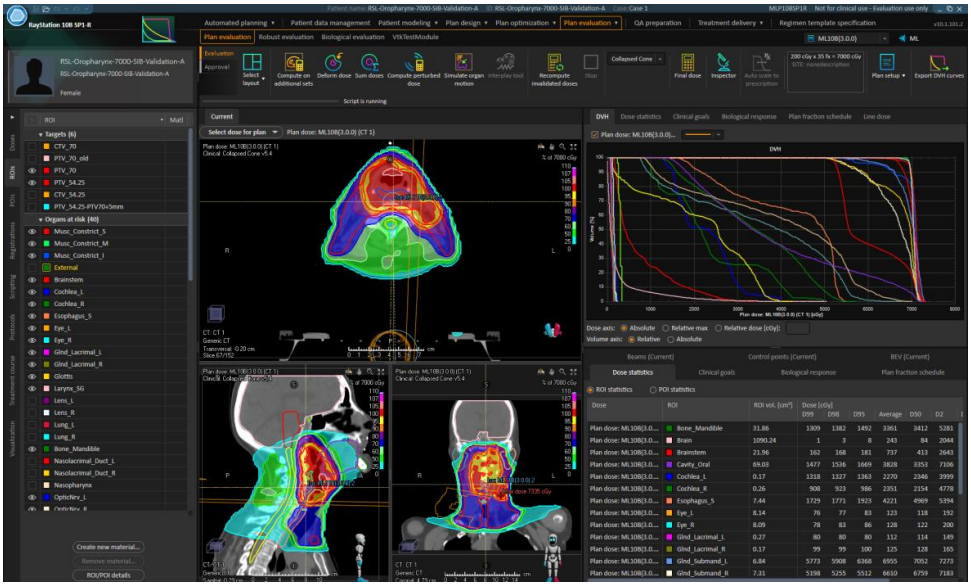
Conclusion

In 10/10 plans proper target coverage and sparing of “Absolute” and “Must” tier clinical goals was achieved. The dose distributions for all plans were visually inspected and found to be clinically acceptable.

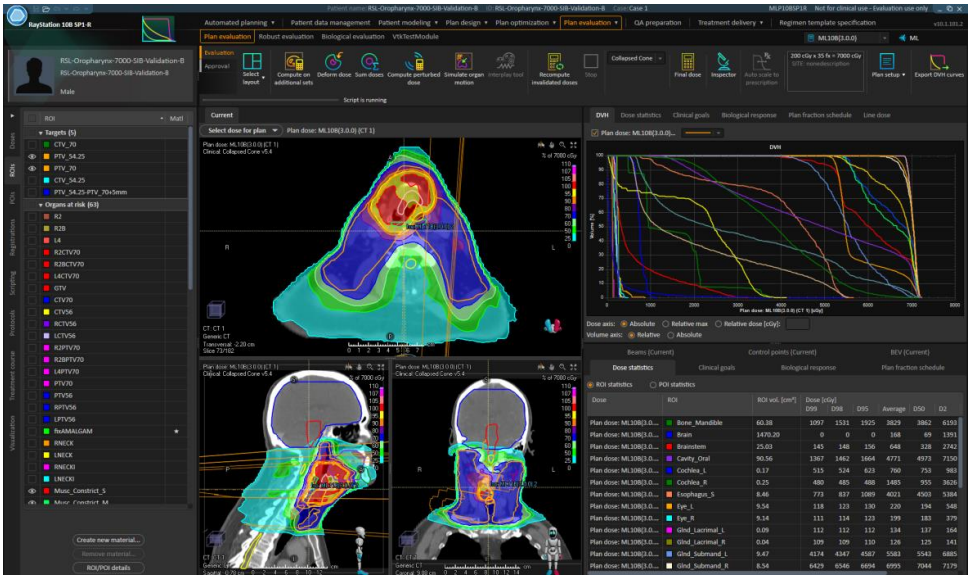
Appendix

Dose distribution images

RSL-Oropharynx-7000-SIB-Validation-A



RSL-Oropharynx-7000-SIB-Validation-B

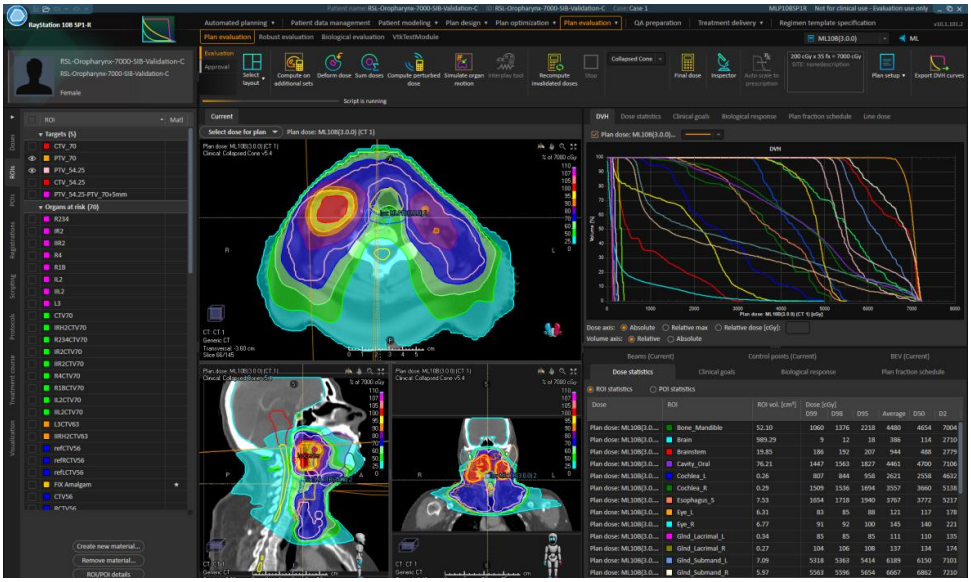


RaySearch Laboratories AB (publ)

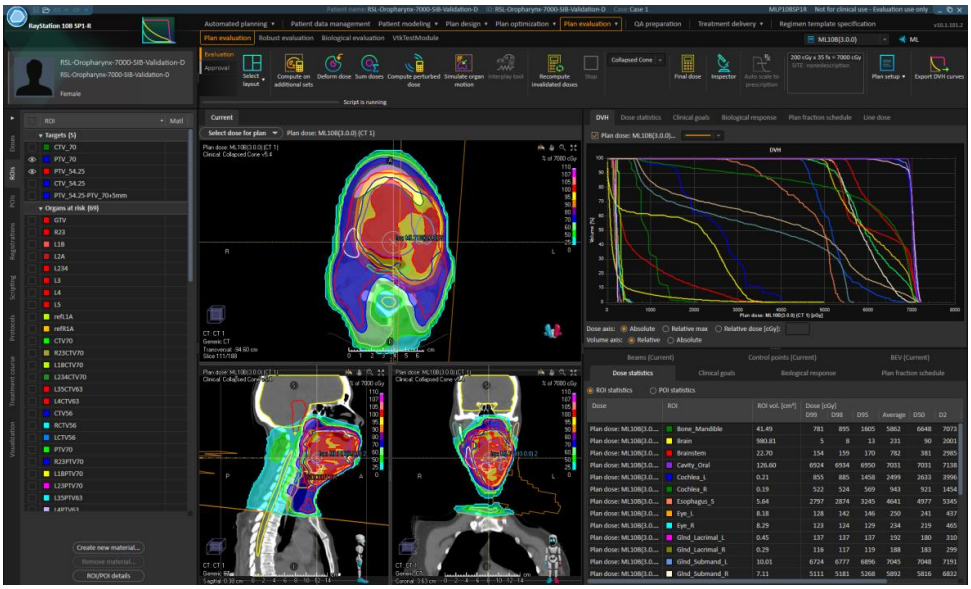
Reg. no. 556322-6157
Box 3297
SE-103 65 Stockholm
Sweden

Visitors: Eugeniavägen 18
info@raysearchlabs.com
www.raysearchlabs.com
Phone: +46 (0)8 510 530 00

RSL-Oropharynx-7000-SIB-Validation-C



RSL-Oropharynx-7000-SIB-Validation-D



RaySearch Laboratories AB (publ)

Reg. no. 556322-6157
Box 3297
SE-103 65 Stockholm
Sweden

Visitors: Eugeniavägen 18
info@raysearchlabs.com
www.raysearchlabs.com
Phone: +46 (0)8 510 530 00

RSL-Oropharynx-7000-SIB-Validation-E

RSL-Oropharynx-7000-SIB-Validation-F

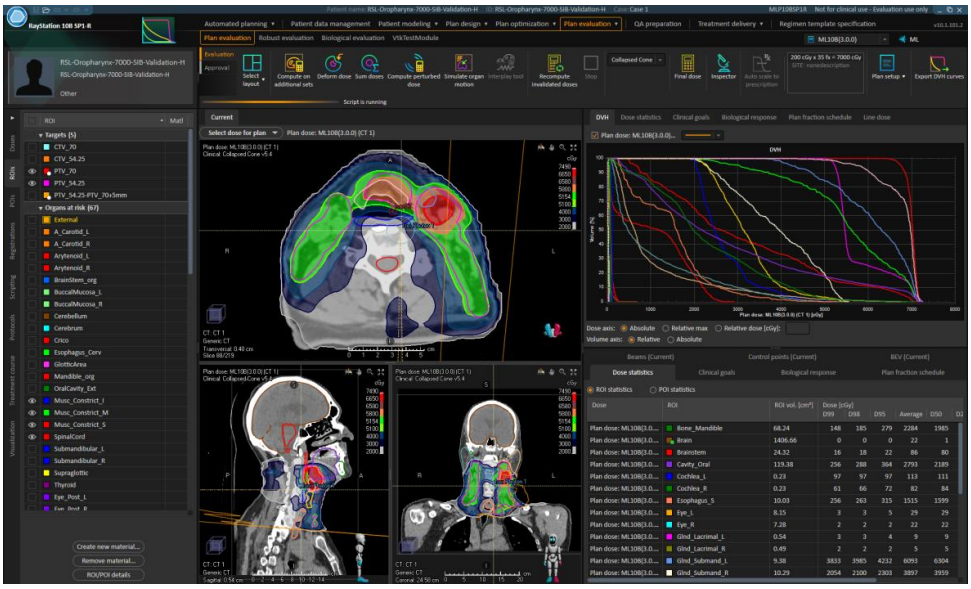
RaySearch Laboratories AB (publ)

Reg. no. 556322-6157
Box 3297
SE-103 65 Stockholm
Sweden

Visitors: Eugeniavägen 18
info@raysearchlabs.com
www.raysearchlabs.com
Phone: +46 (0)8 510 530 00

RSL-Oropharynx-7000-SIB-Validation-G

RSL-Oropharynx-7000-SIB-Validation-H



RaySearch Laboratories AB (publ)

Reg. no. 556322-6157
Box 3297
SE-103 65 Stockholm
Sweden

Visitors: Eugeniavägen 18
info@raysearchlabs.com
www.raysearchlabs.com
Phone: +46 (0)8 510 530 00

RSL-Oropharynx-7000-SIB-Validation-I

RSL-Oropharynx-7000-SIB-Validation-J

RaySearch Laboratories AB (publ)

Reg. no. 556322-6157
Box 3297
SE-103 65 Stockholm
Sweden

Visitors: Eugeniavägen 18
info@raysearchlabs.com
www.raysearchlabs.com
Phone: +46 (0)8 510 530 00

Monitor Units

Patient Id	Total MU
A	670.49
B	709.39
C	696.81
D	633.13
E	769.17
F	697.43
G	784.69
H	893.31
I	755.26
J	762.65

7.5 Associated ROIs in DLAP

ROI	Required or optional
CTV high	Required
CTV low	Required
PTV high	Required
PTV low	Required
Brainstem	Required
Brain	Optional
Spincal corad	Required
Parotid left	Required
Parotid right	Required
Submandible gland left	Optional
Submandible gland right	Optional
PCM inferior	Required
PCM medius	Required
PSM superior	Required
Cricopharyngeus	Required
Oral cavity	Optional
Esophagus	Required
Mandible	Optional
Mandible-PTV	Optional
Trachea	Required
Glottic area	Required
Larynx	Required
PTV low - PTV hogh + 5mm	Required
PTV low - PTV hogh + 10mm	Required
Cochlea left	Required
Cochlea right	Required
Lens left	Optional
Lens right	Optional

7.6 White-Paper RayStation

RayStation

MACHINE LEARNING AUTOMATED TREATMENT PLANNING

Treatment planning is often a time consuming and complex process where sharing of knowledge and expertise between cancer centers is cumbersome and therefore not a reality today. RaySearch already has a strong focus on automation and with machine learning in RayStation, this is taken to a new level.
The machine learning automated treatment planning method learns from historical patient and plan data and infers a 3D spatial dose on a new patient geometry. Together with a powerful mimicking optimization, deliverable treatment plans are generated in minutes. This new approach to planning can improve efficiency, reduce treatment plan variability and facilitate knowledge sharing. University Medical Center Groningen (UMCG) has conducted a clinical study on the method for Head and Neck cancer VMAT cases, showing promising results [1].

MACHINE LEARNING PLANNING IN RAYSTATION

The machine learning treatment planning approach in RayStation utilizes models that have learned the relation between patient geometry, dose shape and tradeoffs from historical treatment plans. A machine learning model is trained by providing treatment plans to the training framework where the model learns to infer dose spatially on a new patient geometry. This can be compared to a dosimetrist learning over time by creating treatment plans for new patient cases. After a trained model has inferred dose to the patient, a dose mimicking optimization is performed to generate an optimized deliverable treatment plan.

The model itself does not contain any personal data and can therefore easily be shared between clinics. In the RayStation 8B* version released in December 2018, the machine learning planning supports VMAT, IMRT, and TomoTherapy treatments. Future releases will support protons and other delivery techniques as well.

The machine learning method for planning [2] has been developed in a collaboration with Princess Margaret Cancer Centre in Toronto, Canada, and is the first machine learning application for treatment planning in a treatment planning system on the radiation oncology market today.

> *"Machine learning is a natural fit for automating the complex treatment-planning process. It will enable us to generate highly personalized radiation treatment plans more efficiently, thereby allowing clinical resources or specialist technical staff to dedicate more time to patient care."*
>
> *Tom Purdie, Medical Physicist,*
> *Princess Margaret Cancer Centre, Toronto, Canada*

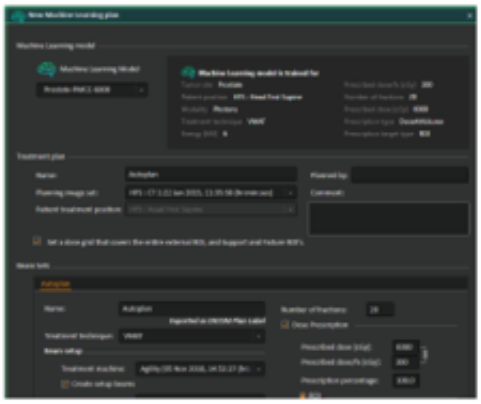


Figure 1. Dialog box in RayStation for creating a machine learning plan.

GENERATING MACHINE LEARNING PLANS

With the machine learning planning module in RayStation, it is possible to generate one or multiple deliverable plans in minutes. This is done by applying one or many trained models to the patient, where each model typically is associated with a treatment site, delivery technique and protocol. Each model can produce one or multiple spatial doses for the patient based on learned tradeoffs from the patient anatomy, tumor size and location. The inferred doses are then mimicked in RayStation to retrieve a deliverable plan.

RayStation will come with pre-trained models from leading cancer clinics. Clinics can also train models using their own data. Both model training and treatment plan generation can be accessed via scripting to fully automate the planning process.

MACHINE LEARNING METHOD

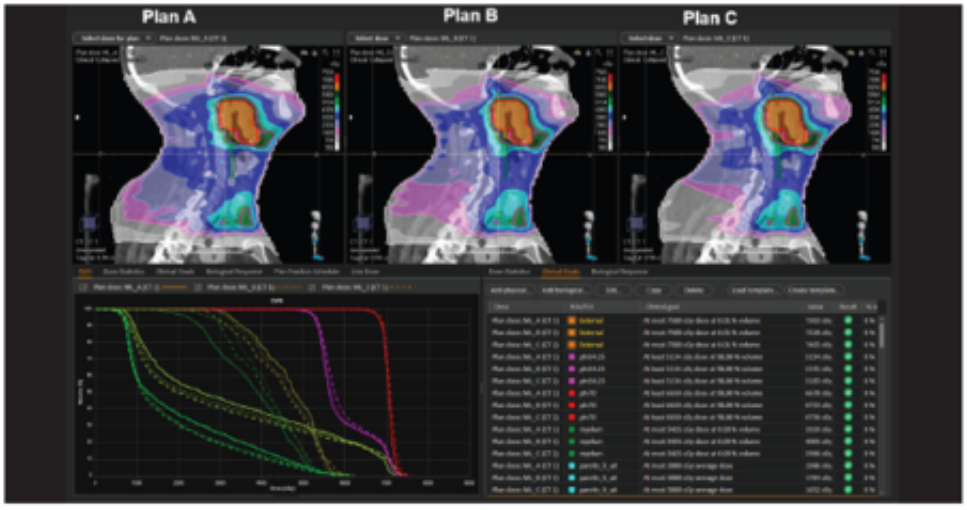
Unlike traditional treatment planning methods, where the dose is generated at the end of the workflow, the dose is inferred at the beginning of the workflow as input to the mimicking optimization, see Figure 2.

The mimicking optimization is performed to create a deliverable dose for the selected treatment machine and beam setup. This optimization will also strive to improve the dose when possible; spatially and optionally through clinical goals.

The machine learning framework can also infer multiple doses based on strategies defining tradeoffs and goals for the plan, see Figure 3. This makes it possible to push the machine learning plans to eventually create better plans than the plans used in the training set. The multiple plan option helps to get a better understanding of the tradeoffs for the patient. Any of the generated plans can be selected for delivery or for post-processing.

Figure 2. Schematic overview of the treatment planning process. Top: traditional workflow. Bottom: machine learning workflow.

Figure 3. Automatically generated deliverable plans for a Head and Neck cancer case based on three different strategies: Plan A: Standard (solid line), Plan B: Greedy (dashed), Plan C: Avoid Xerostomia (dash-dotted).



CLINICAL STUDY AT UMCG

CT scans, structures and doses of 71 primary Head and Neck cancer patients from UMCG, treated with dual arc VMAT and two dose levels; 70 Gy and 54.25 Gy delivered in 35 fractions, were collected. Patient selection was restricted to tumors localized in the oropharynx, larynx, oral cavity, nasopharynx and paranasal sinuses. A repeated random subset validation approach was applied where the patient data was split into four sets; using 8 patients for testing and the remaining 63 patients for training in each validation set.

The machine learning method in [2] was used for training. The trained models were used to predict three spatial doses for each patient in the test set, corresponding to three defined strategies: *Standard, Greedy, and Avoid Xerostomia*. In the final step, each predicted dose was input to a mimicking optimization algorithm to generate a deliverable dual arc VMAT plan. The predicted and mimicked doses of the patients in the test set were compared against the dosimetrist-optimized clinical plans, denoted below by reference plans.

Plans were compared in terms of the following dose metrics:

- Targets: D98 > 95%, D2 < 107%;
- OARs: Dmean, Dmax.

One mimicked plan out of the three automatically generated plans per patient was selected based on D98 to both targets and Dmean to OARs. No further post-processing was performed on the mimicked plans.

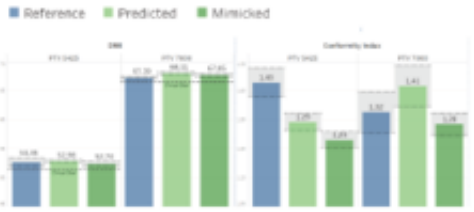


Figure 4. Dose metric comparison for the two target volumes, D98 (left) and conformity index (right), with grey area showing the 95% confidence interval.

RESULTS

The machine learning plans were generated automatically via scripting. The average run time per plan was 29 minutes, where the dose prediction took 4 minutes and the mimicking optimization for the dual arc VMAT took 25 minutes on an Intel® i9-7940X CPU.

The predicted dose was in accordance with the reference dose for all plans. The mimicked plans had adequate target coverage for primary and elective target volumes according to clinical goals in 31/32 (97%) of the cases. For reference plans, 30/32 (94%) had adequate target coverage based on the same criteria. Target conformity was better in the mimicked plans compared to the reference plans, see Figure 4.

The maximum dose to brain, brainstem, and spinal cord was lower for the mimicked plans than for the reference plans. Overall, the average dose to parotids, oral cavity, pharynx and supraglottic were similar, see Figure 5.

CONCLUSIONS

The machine learning planning approach is an efficient method to generate high-quality treatment plans automatically. In a retrospective Head and Neck study with UMCG, the machine learning planning approach is capable of generating deliverable dual arc VMAT plans with adequate target coverage in 97% (31/32) of the cases. The dose to the OARs was similar to the dose of the clinical reference plans.

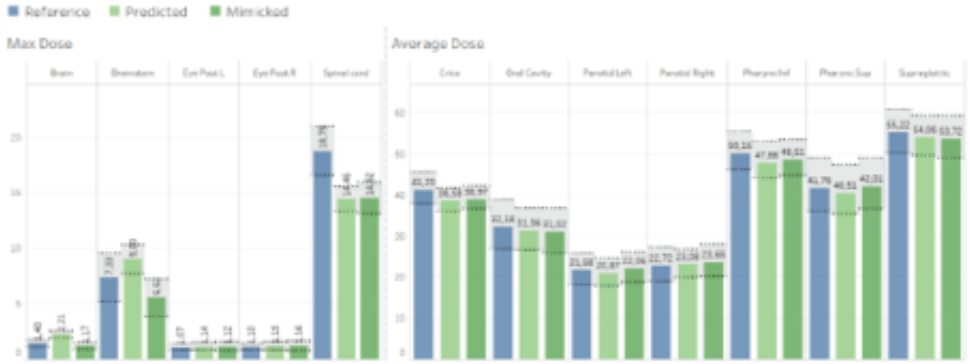
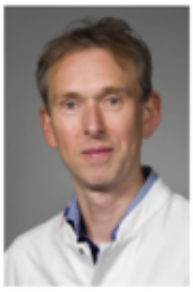


Figure 5. *Dose metric comparison for OARs, max dose (left) and average dose (right), with grey area showing the 95% confidence interval.*



"It was great to work together with RaySearch on the forefront of this very promising new technology. Results were good from the start and quality of dose predictions and dose mimicking could be further improved within a short time frame, thanks to the dedicated RaySearch team. It is now ready for prime time to support our efforts to give each patient the best possible treatment."

***Erik Korevaar**, Medical Physicist,*
University Medical Center Groningen Netherlands

REFERENCES

[1] van Bruggen IG, Kierkels RGJ, Holmström M, Lidberg D, Berggren K, Both S, Langendijk JA, Löfman F, Korevaar EW.
Fully automated treatment planning of deliverable VMAT by machine learning dose prediction and mimicking optimization in HNC, ICCR abstract, 2019.

[2] McIntosh C, Welch M, McNiven A, Jaffray DA, Purdie TG.
Fully Automated Treatment Planning for Head and Neck Radiotherapy using a Voxel-Based Dose Prediction and Dose Mimicking Method, Phys.Med.Biol., 2016.

* Subject to regulatory clearance in some markets.

For more information or to book a demo,
visit www.raysearchlabs.com

RaySearch Laboratories

Machine Learning – Automated treatment planning - 2019-03-28