# Investigation of Adaptive Radiation Therapy Including Deformable Image Registration, Treatment Planning Modification Strategies, Machine Learning & Deep Learning

by

Paweł Siciarz

A thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Physics and Astronomy

University of Manitoba

Winnipeg, Canada

# Acknowledgments

First and foremost I would like to thank my supervisor, Dr Boyd McCurdy from CancerCare Manitoba. An opportunity to work with and learn from an individual with such extensive knowledge, experience and unmatched leadership qualities was a true honour. Dr McCurdy's insightful and valuable feedback during our countless discussions helped me immensely to grow as a young researcher and to tackle a sea of challenges that the pursuit of the Ph.D. degree had to offer.

Boyd, for your great support, encouragement, and overall awesomeness I will be always grateful!

I would also like to extend my gratitude to the Ph.D. committee members: Dr Boyd McCurdy, Dr Erik Van Uytven, Dr Ryan Rivest, Dr Jason Fiege and Dr Gabriel Thomas for their time, comments and guidance regarding research projects included in this thesis. Special thanks to Erik and Ryan for sharing their hands-on expertise in clinical radiation therapy.

The CancerCare Manitoba Foundation and the University of Manitoba provided financial resources that enabled me to complete this research work. I am very thankful for it.

This thesis was completed with the inclusion of collaborative effort of the following individuals: Timothy Van Beek (software engineering support), Dr Faiez Alshafa (contouring CBCT scans and contour evaluation); Dr Peter B. Greer, Dr Joan A. Hatton,

Dr Philip Wright (CBCT data acquisition and contouring); Dr Nikesh Manumanthappa (CBCT contouring); Dr Salem Alfaifi, Dr Shrinivas Rathod (clinical consultation and study design), Dr Rashmi Koul (data acquisition). Please know, that I greatly appreciate all of your time and contributions.

It's undoubtedly worth mentioning that, apart from research, Medical Physicists working at CancerCare Manitoba created a great educational environment by teaching courses, providing constructive feedback to the Journal Club student presentations and being open to answer any question from inquisitive students. Thank you all for your kindness and knowledge sharing. It was a pleasure to be a part of this community over the last 7 years. At the same time, I would like to thank Dr Stephen Pistorius – the Director of the Medical Physics program who makes it all possible and is very approachable for students' inquiry regarding the program.

I also greatly appreciate the supportive staff at the Medical Physics Department of CancerCare Manitoba, especially, Alana Dahlin, Jovanka Halilovic, Tracy Tyefisher and Luanne Scott who helped me multiple times with any administrative task.

There is also one person from the Department of Physics and Astronomy, the University of Manitoba that I would like to express special gratitude to – Ms Susan Beshta. Susan, I would like to thank you very much for helping me navigate and manage important documents as well as going above and beyond in time-sensitive matters. I cannot thank you enough for that. Working at the same department Dr Ruth Cameron and Andriy Yamchuk are two individuals who I would also like to thank for introducing me to the role of Teaching Assistant and making it a great 3-year experience during my Ph.D. program.

*ACKNOWLEDGMENTS*

I would also like to thank all past and present Medical Physics students for their friendship, companionship, conversations and common experiences including Parandoush Abbasian, Sajjad Aftabi, Princess Anusionwu, Suliman Barhoum, Robert Bergen, Tamar Chighvinadze, Kaiming Guo, Ivan Kutuzov, Peter McCowan, Bryan McIntosh, Azeez Omotayo, Graham Schellenberg, Hongyan Sun, Hongwei Sun, Mohammadreza Teimoorisichani, Troy Teo and Geng Zhang.

I would like to conclude these acknowledgments by expressing my deepest gratitude to people I love the most – my grandmother Sabina Majewska, my mother Renata Majewska and my brother Bartłomiej Siciarz. You gave me all the strength, inspiration and motivation needed to complete this work. Without you, it would have never been created.

Last but not least I would like to thank my dear girlfriend Minh Tam Vu for her love, patients, understanding, and unstoppable positive spirit over the years.

# Abstract

One of the main challenges in radiation therapy treatment is to ensure accurate delivery of the radiation dose that was prescribed by a radiation oncologist. A major contribution to this issue is a change in the patient's anatomy, which may occur during the treatment. To account for these changes, the radiotherapy treatment plan may be adapted. The radiation therapy approach that incorporates such treatment plan modifications is called Adaptive Radiation Therapy (ART). In practice, the complexity of radiation therapy planning, and delivery requires an application of specific tools and procedures in order to make the ART process time-efficient, and the adapted treatment plan sufficiently accurate. The goal of this thesis was to propose and evaluate solutions to four important aspects of adaptive radiation therapy in order to make it more reliable, accurate, and efficient.

The first study focused on the evaluation of several deformable image registration algorithms by registering computed tomography scans to daily CBCT images for five prostate cancer patients treated with intensity-modulated radiation therapy. Results demonstrated that the Dense Anatomical Block Matching (DABM) registration outperformed the other common methods in terms of quantitative and physician evaluation. In the case of Rigid, Affine and B-Spline registration algorithms, the Dice similarity coefficient yielded similar values of 0.85, 0.8, 0.75 and 0.7 for PTV, bladder,

GTV and rectum respectively. Application of the DABM algorithm improved the DSC to around 0.9 for bladder and rectum, 0.85 for GTV and 0.8 for rectum.
Individual physician correction time (the shorter the time the better the performance of image registration) was calculated as the sum of the average correction times for individual structures (rectum, bladder, GTV) were: 5.3 min for DABM, 8.6 min for B-Spline, 9.0 min for Affine and 13.5 min for Rigid registration. The physician needed a total of 7.0 min to segment the rectum, bladder and GTV 'from scratch'. Therefore, the accuracy of the Dense Anatomical Block Matching algorithm was generally at a clinically acceptable level (i.e. within published inter-observer variability range), making DABM a very promising alternative to the existing registration methods when it comes to challenging CT-CBCT deformable image registration problem and its applications such as dose calculation, dose mapping, and contour propagation in adaptive radiation therapy of the pelvic region.

The second study focused on the quantitative evaluation of eight adaptive radiation therapy approaches for 20 prostate cancer patients treated with hypofractionated VMAT. The ART strategies included online and offline methods as well as the application of dose feedback as theoretically available through *in vivo* dosimetry methods. The passing rate (percentage of patients who met specific treatment planning objective) for $D_{99\%}$>3800cGy was 0% for non-ART plans, while for ART plans it was between 0% and 25% and for the original plan it was 100%. In the case of CTV, the passing rate for D99%>4000 cGy objective was >=90% for all treatment plans except non-ART plans for which it was 0%. Treatment plan objectives for bladder were meet by non-ART and ART plans. The passing rate for rectum was 80% and 75% regarding non-ART plan as well as 90% to 100% and 65% to 95% for ART plans with respect to $D_{15\%} < 3200$ cGy and $D_{20\%} < 2800$

cGy objectives, respectively. The performed comprehensive analysis showed that daily online adaptation approaches (ie. every fraction) were the most advantageous, although strategies adapting every other fraction also were impactful while reducing relative workload as well. Offline treatment adaptations were shown to be relatively less beneficial due to increased dose delivered to the bladder and rectum compared to other ART strategies. The findings of this study provided useful insights into the selection of the optimal ART strategy, improving the quality of the decision-making process based on the quantitatively evaluated dosimetric benefits.

The third study aimed to utilize a deep learning network to automatically contour critical organs on the computed tomography (CT) scans of head and neck cancer patients who underwent radiation therapy treatment. Trained U-net models were able to contour 25 critical organs on unseen CT image data sets at a clinically acceptable accuracy (i.e. within published interobserver variability range) in 6.8 s per patient (compared to several hours for manual contouring performed by a radiation oncologist). Relatively high accuracy and short contouring time could allow for the implementation of the model within a clinical ART workflow, which would lead to a significant decrease in the time required to create a new adapted treatment plan.

The objective of the fourth study was to use machine learning methods to build a decision support system that would classify previously delivered VMAT plans of brain tumor patients into those that met PTV treatment planning objectives and those for which PTV objectives were not met due to the priority given to one or more OARs (i.e. a trade-off was required); those plans, however, were still clinically acceptable and delivered. Among those evaluated, the Logistic Regression model achieved the highest accuracy of

98.3%±4.1% on the testing data and can be used by radiation oncologists to support their decision-making process in terms of potential treatment plan adaptations and plan approvals as well as by medical physicists in a data-driven quality assurance program. The direct benefit to ART comes from eliminating the need for time consuming search and analysis of previously delivered plans that could help classify a new plan as the one that may potentially require modifications, as well as for treatment plan quality assurance.

The research studies conducted and described in this dissertation advance several key aspects of adaptive radiation therapy by innovative application of existing but understudied methods in terms of deformable image registration and ART strategies as well as through the development of new custom solutions based on artificial intelligence for medical image segmentation and radiation therapy plan classification. Presented results provide a solid foundation for the potential future development of software applications that could be integrated with the existing radiation therapy planning software and improve its functionality in order to deliver more efficient, personalized, and higher quality treatment to cancer patients.

# Contents

# List of Figures

## Chapter 1



## Chapter 2

## Chapter 3

## Chapter 4

## Chapter 5

## Chapter 6

## Chapter 8

# List of Tables

## Chapter 3



## Chapter 4

## Chapter 5



## Chapter 8

# List of Abbreviations

| Abbreviation | Description |
|---|---|
| **ANN** | Artificial Neural Network |
| **ART** | Adaptive Radiation Therapy |
| **AUC** | Area Under the Curve |
| **CBCT** | Cone Beam Computed Tomography |
| **CCS** | Cancer Care Society |
| **CI** | Conformity Index |
| **CLAHE** | Contrast Limited Adaptive Histogram Equalization |
| **CNN** | Convolutional Neural Network |
| **COM** | Center of the Mass |
| **CT** | Computed Tomography |
| **CTV** | Clinical Target Volume |
| **DAMB** | Dense Anatomical Block Matching |
| **DBSCAN** | Density-Based Spatial Clustering of Applications with Noise |
| **dCT** | deformed Computed Tomography |

| | |
|---|---|
| **DF** | Dose Feedback |
| **DIR** | Deformable Image Registration |
| **DL** | Deep Learning |
| **DNA** | Deoxyribonucleic acid |
| **DNN** | Deep Neural Network |
| **DSC** | Dice Similarity Coefficient |
| **DVF** | Deformation Vector Field |
| **DVH** | Dose Volume Histogram |
| **EBRT** | External Beam Radiation Therapy |
| **FFD** | Free Form Deformation |
| **FN** | False Negatives |
| **FOV** | Field of View |
| **FP** | False Positives |
| **GFFD** | Gradient Free Form Deformation |
| **GPU** | Graphical Processor Unit |
| **Grad-CAM** | Gradient-weighted Class Activation Mapping |
| **GTV** | Gross Tumor Volume |
| **HD** | Hausdorff Distance |
| **HI** | Homogeneity Index |
| **ICE** | Inverse Consistency Error |
| **ILSVRC** | The ImageNet Large Scale Visual Recognition Challenge |

| | |
|---|---|
| **IMRT** | Intensity Modulated Radiation Therapy |
| **ITK** | Insight Toolkit |
| **KBP** | Knowledge-Based Planning |
| **LET** | Linear Energy Transfer |
| **MAD** | Median Absolute Deviation |
| **ML** | Machine Learning |
| **MLC** | Multi-Leaf Collimator |
| **MNM** | Modified Newton Method |
| **MRI** | Magnetic Resonance Imaging |
| **NTCP** | Normal Tissue Complication Probability |
| **OAR** | Organs at Risk |
| **OBI** | On-Board Imager |
| **PCA** | Principal Component Analysis |
| **pCT** | Planning (Pre-treatment) Compute Tomography |
| **PET** | Positron Emission Tomography |
| **PTV** | Planning Target Volume |
| **QA** | Quality Assurance |
| **RF** | Random Forest |
| **ROC** | Receiver Operating Characteristic |
| **SD** | Standard Deviation |
| **SE** | Standard Error |

| | |
|---|---|
| **SHAP** | SHapley Additive exPlanations |
| **SVM** | Support Vector Machines |
| **TCP** | Tumor Control Probability |
| **TN** | True Negatives |
| **TP** | True Positives |
| **TPS** | Treatment Planning System |
| **VMAT** | Volumetric Modulated Arc Therapy |
| **WHO** | World Health Organization |

# Chapter 1

# Rationale

This chapter introduces basic concepts of radiation therapy treatment, including adaptive radiation therapy (ART). It explains the rationale behind certain components of the ART process that are challenging, and which would benefit from improvements in order to make plan adaptation more accurate and efficient, and therefore more suitable for clinical implementation. The chapter provides the formulation of the purpose of this dissertation and also summarizes the scientific contribution of the thesis research.

## 1.1 Introduction to Radiation Therapy

Cancer is one of the world's largest health issues. It is the leading cause of death in Canada and the second most common cause of death (exceeded only by heart disease) in the United States. According to the Canadian Cancer Society (CCS), approximately 49% of males and 45% of females will develop cancer in their lifetime, and 25% of Canadians will die from cancer. The CCS also reports that the number of cancer patients is expected to increase every year in Canada[1] for the foreseeable future.

According to the World Health Organization (WHO), radiotherapy is one of the major treatment options in cancer management[2,3]. Depending on the stage and type of cancer, approximately 50–65% of all cancer patients require radiotherapy during their disease[2,3]. Together with other modalities such as surgery and chemotherapy, radiation therapy plays an important role in the treatment of approximately 40% of patients who are treated for cancer[4]**.** Radiotherapy is a multi-stage, complex process that utilizes the energy of ionizing radiation to kill cancerous cells, and advanced technologies to deliver these treatments. The rapid development of technology combined with the introduction of new irradiation techniques have always aimed to deliver the highest possible radiation dose to the volume treated, with maximal sparing of normal, healthy tissues while maintaining the economic efficiency of the treatment[5-7].

Although there are numerous methods of delivering a high dose of radiation to the diseased tissue target(s), this thesis focuses solely on external beam radiation therapy (EBRT) using photons due to the scope of pursued research. EBRT is a widely available technique often utilizing a linear accelerator (or 'linac', Figure 1.1a) that generates a high-energy beam of photons (i.e. x-rays) directed into a patient in order to deliver a radiation dose to the tumor cells (i.e. target) inside the patient's body. Typically the dose is delivered to the target structure from multiple beam angles as shown in Figure 1.1b, such that the overlapping region of the beams receives a much higher dose compared to the surrounding healthy tissues. To further protect healthy organs, the shape of radiation beams is conformed to the shape of the tumor at the specific beam angle. Sections 2.1.6 and 2.1.7 introduce more advanced irradiation techniques, where the intensity of the beams is modulated to achieve a highly conformal dose distribution in the treated volume[8,9].

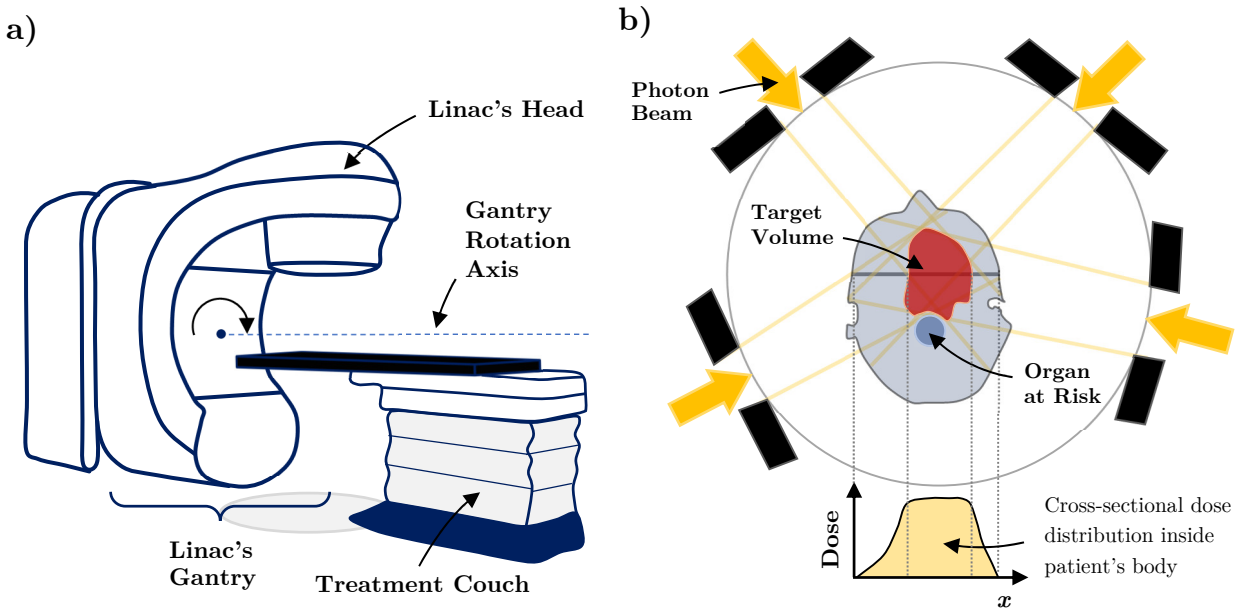



**Figure 1. 1. a)** A modern linear accelerator (linac) used to deliver a radiation dose to the treated volume; **b)** Diagram illustrating the basic concept of dose delivery using multiple external photon beams during a radiotherapy session[9].

## 1.2 Radiation Therapy Workflow

The radiotherapy process consists of several steps. First, a radiation oncology patient is diagnosed, referred for radiation treatment, and then prescribed a particular amount of therapeutic dose. Then an immobilization device (generic or custom-made) is frequently used to help ensure reproducible patient positioning during imaging data acquisition and treatment delivery. In the next step, computed tomography (CT) scans of the patient's anatomy are obtained using a specialized CT scanner, called a CT 'simulator'. The patient's geometry during this scan replicates (i.e. simulates) the intended geometry of the treatment set-up at the linac. Depending on the treated lesion, additional imaging modalities, such as MRI (Magnetic Resonance Imaging) and PET (Positron Emission Tomography) can also be used in combination with CT to improve the accuracy of tumor

definition by providing additional anatomical information (via different tissue contrast) or physiological information (via functional imaging techniques). Based on the imaging data, the radiation oncologist delineates target structures and organs-at-risk (OARs) and prescribes the radiation treatment which details the radiation dose to be delivered to the target(s), the dose limits of the OARs, and the dose fractionation schedule. After that, radiation therapy treatment planners (also called 'dosimetrists') create a custom treatment plan in the computerized treatment planning system (TPS) using the irradiation technique and prescription specified by the radiation oncologist. The radiation dose that will be delivered to the patient is accurately calculated in the TPS using dose calculation algorithms and based on the tissue attenuation information derived from the CT simulation scan. Once completed, the treatment plan quality is verified by a medical physicist, given final approval by a radiation oncologist, and then delivered to the patient by a radiation therapist according to the specified fractionation scheme. To ensure the successful delivery of each dose fraction, the patient position is carefully verified using imaging techniques just before the irradiation session, by a radiation therapist. Every radiation therapy department also implements numerous additional quality assurance procedures to maintain and monitor the safety and accuracy of the entire radiation treatment procedure. After the patient finishes his/her radiotherapy course, the outcomes of the treatment are evaluated during follow-up appointments[10-12].

## 1.3 Overview of Adaptive Radiation Therapy

In radiotherapy, patient-specific parameters can have a large impact on the outcome of the treatment[13,14] yet in current practice, some of the information used while planning the patient's therapy is still based on general population statistics, which in some cases makes

an individual's therapy suboptimal[15]. A meaningful example is that the delivery of a full therapeutic dose of radiation treatment is usually planned only once, about 1-2 weeks ahead of the start of radiation treatment. Then the total prescribed radiation dose delivery is spread out into smaller daily doses often given over several weeks ignoring potential changes in soft tissue over that time frame. However, it is known that the patient anatomy including the local tumor anatomy can change significantly over this time period for some patients. Even in hypofractionation schemes which are recently gaining clinical interest, where the prescribed radiation dose is delivered in fewer fractions over a shorter time period (i.e. 1-8 fractions in less than 2 weeks), changes in the patient's anatomy can be substantial[16]. The use of hypofractionation is growing as a standard option in modern radiotherapy, however, it dramatically increases the effect of a geometric delivery error as compared to conventional fractionation approaches, where the impact of such an error would be reduced by delivering the treatment over many more fractions. One solution for this problem is to more closely monitor the changes in the patient's anatomy over the entire course of the radiotherapy treatment and, if necessary, adapt the treatment to ensure that the radiation dose is delivered to the patient as planned.

This strategy of adapting the radiation therapy treatment plan to the patient's changing anatomy, known generally as Adaptive Radiation Therapy (ART), has been an active area of research in recent years[17-23]. Typically, an ART strategy relies on frequent imaging throughout the course of radiation treatment. While daily imaging is routinely used to make patient position adjustments ahead of the radiation treatment delivery, it is usually based primarily on the alignment of bony anatomy in 2D projection imaging or 3D volumetric imaging. In contrast, ART strategies go beyond simple patient repositioning

and involve a modification of the treatment plan (when needed). This modification is based on detailed 3D anatomical information including the tumor, organs at risk, and patient shape. The premise for ART is that if the patient anatomy during the treatment course can differ significantly from the anatomy observed in the originally developed treatment plan, then the treatment plan can be modified for the remaining fractions to compensate for the anatomical changes that have occurred, thus ensuring the tumor still receives the intended therapeutic dose and also keeping the OARs within their dose tolerances.

## 1.4 Practical Challenges of Adaptive Radiation Therapy

### 1.4.1 CT to CBCT Deformable Image Registration

Deformable image registration (DIR) in radiation therapy is a technique that establishes a mathematical correspondence between two (most often) or more sets of 2D, 3D (most often) or 4D volumetric anatomical images, acquired at different points in time, based on a predetermined similarity metric (note, DIR can also be applied to simple two-dimensional images). The application of a DIR algorithm to a pair of volumetric image sets generates two results. The first is a three-dimensional vector field (DVF – deformable vector field) which is used to map one image set (termed the 'floating' or 'moving' image) to the other image set (the 'target' or 'fixed' image). The second result is the deformed image set, which is the moving image set transformed with the DVF. In addition to transforming the moving image set, a deformable vector field can also map any volumetric or single-point(s) attributes associated with a transformed image.

In the context of adaptive radiation therapy, DIR is used mainly for anatomical contour propagation in order to reduce manual segmentation time, for daily dose estimation to monitor the quality of dose delivery, and for the creation of a new treatment plan to account for changes in the patient's anatomy in order to ensure that treatment objectives are still met[24]. All these applications require the DIR algorithm to be sufficiently accurate, which is challenging considering that the target image is usually represented by a Cone Beam Computed Tomography (CBCT) scan. CBCT is an x-ray imaging system attached to the linear accelerator and employs a broad, conical x-ray beam. It is used to visualize and verify the setup accuracy of the daily anatomy of the patient when placed in the treatment position, immediately before the radiation dose is delivered, but has much poorer image quality compared to conventional CT due to contamination of the projection image by patient-generated scattered x-rays[25]. The significant difference in image quality between CT and CBCT makes this image registration application very challenging[26,27].

### 1.4.2 Adaptive Radiation Therapy Strategies

The selection of a specific ART strategy is one of the most important aspects of ART in terms of practical implementation in the clinical environment because it needs to balance the time, economic and human resources that are required to perform ART effectively, against the quantitative dosimetric improvements in the patient treatment. ART strategies must consider and justify the frequency and effort of treatment plan modifications that additionally depend on the type of radiation therapy the patient is undergoing and must be evaluated for the specific clinical case being treated.

For these reasons, it is clear that the process of selecting a suitable ART strategy is very challenging. At the moment, the main solution to this problem is a comprehensive evaluation of various ART strategies retrospectively and then selecting those approaches that maximize the treatment benefits within the allocated clinical resources. It is important to note that the evaluation of the trade-off between the dosimetric improvement in treatment with ART versus the additional resources used to achieve that improvement, is not a purely scientific one but should be based on a cost-benefit analysis that could vary between institutions and even between specific patient situations.

### 1.4.3 Anatomy Segmentation

Anatomy segmentation has always been a challenging step in adaptive radiation therapy (and radiation therapy in general) as it is a complex task that can involve different image modalities and image qualities and is therefore very time-consuming. At the same time, the accurate delineation of the target(s) and OARs is crucial for the development of the treatment plan and the evaluation of the radiation dose delivered to the patient.

There have been numerous approaches to automate the segmentation process including edge detection[28], region growing[29], intensity thresholding[30], machine learning[31] as well as single and multi-atlases of imaging data[32]. However, the main disadvantages of these methods include the limited use of prior anatomical knowledge, constrained flexibility, and/or suboptimal accuracy. A possible improvement upon those limitations has become possible due to the recent advances in deep learning techniques, which currently are considered the state-of-the-art method for medical image segmentation and have opened many new unexplored research opportunities[33,34,35].

### 1.4.4 Patient Selection

Not every patient will benefit from the application of ART[20,38]. The selection of those patients who would likely benefit from plan adaptation is an important aspect of adaptive radiation therapy because it helps maximize efficient resource allocation in a clinic. To develop a system that would facilitate the patient selection process for possible ART, it is necessary to first analyze historical treatment plans and, based on that analysis, establish selection criteria. Then once a new treatment plan is created it can be compared to similar past treatment plans and be evaluated against the selection criteria. However, the manual analysis of historical plans and then the proper classification of a new plan are complex and time-consuming tasks that generally cannot be performed in real-time for all patients. Therefore, the efficient automation of this analysis would be of great importance in the clinical environment. Recently developed technology allows building a machine learning-based system that is trained through input of historical plans that have been classified into those that would likely benefit from ART and those that would not. Then, the trained machine learning (ML) model can be used to classify new plans and direct radiation oncology treatment towards using ART strategies for those patients who are most likely to benefit[38,39].

## 1.5 Research Purpose

The purpose of the research work presented in this dissertation was to propose and evaluate solutions to some of the challenges facing the clinical implementation of ART in the radiotherapy clinic. Specifically, the four problems described in the preceding sections 1.4.1. - 1.4.4. are investigated, in order to make adaptive radiation therapy more reliable,

accurate, and efficient for selected clinical disease sites. The following section briefly outlines the scientific contributions associated with the proposed solutions.

## 1.6 Scientific Contribution

This section summarizes the scientific contributions of each of the four research projects comprising this thesis, detailed in chapters 3 to 6, that have contributed to advance several aspects of adaptive radiation therapy.

### 1.6.1 CT to CBCT Deformable Image Registration

The purpose of this work, described in Chapter 3, was to evaluate the performance of the non-linear Dense Anatomical Block Matching (DABM) algorithm for CT-CBCT image registration. The DABM method had not been previously investigated for this challenging multi-modal image registration problem. Five prostate patients that underwent intensity-modulated radiation therapy (IMRT) were used in this work. Pre-treatment CT (pCT) data sets and mid-treatment CBCT data sets acquired during the radiotherapy course were used to validate the algorithm performance and benchmark against other commonly used DIR algorithms such as Rigid, Affine, and B-spline. After registration, anatomical structures delineated on the pCT were deformed using the obtained deformation vector fields, then propagated to the CBCT images and compared to the analogous contours delineated on the CBCT by an experienced radiation oncologist. The results indicated that for the patients and anatomical structures considered in this study, both the accuracy and the consistency of the DABM algorithm were considerably better than the other evaluated commonly-implemented registration methods. Additionally, generated DVFs had a well-preserved topology (i.e. did not contain unrealistic deformations of the patient's

anatomy). Therefore, DABM was shown to be a promising option for applications in ART-related tasks such as image deformation, contour propagation, and dose mapping.

### 1.6.2 Adaptive Radiation Therapy Strategies

Chapter 4 describes a study that performed a comprehensive evaluation of eight adaptive radiation therapy strategies applied to 20 prostate cancer patients. Although ART is an active area of research, these strategies have not yet been explored for hypofractionated VMAT treatment for prostate cancer. The performance of various ART strategies (eight in total consisting of both offline and online approaches) was quantified by the comparison of dose and dose-volume metrics for the original treatment plan and adapted plans with the consideration of target and critical structures. Note that the online and offline ART approaches are described in more detail in section 2.3 and also Chapter 4. It was found that non-adapted plans (original plans) deviated unfavourably from the intended delivery, while the best performing online ART strategies resulted in a significant reduction of those deviations. Additionally, most ART strategies improved the CTV (Clinical Target Volume) coverage as measured with the homogeneity index by up to 30% relative to the intended plan. Using the same metric, PTV (Planning Target Volume) coverage improved by around 20% with respect to the non-adapted plan. In conclusion, daily on-line adaptation approaches were the most impactful while offline treatment adaptations were shown to be less beneficial due to increased dose delivered to bladder and rectum compared to other ART strategies. Among online plan adaptations considered in this study, the strategy of performing online ART during the first, third, and fifth fractions, delivered almost the same dosimetric benefit as the daily on-line adaptation approach, but at a significantly reduced workload. The detailed analysis of dosimetric metrics for all

the structures and for all considered ART approaches is included in Chapter 4. The time for each ART strategy was estimated with the consideration of plan optimization, dose calculation, and deformable image registration. The calculation time did not include CBCT image acquisition, data transfer, and new adapted treatment plan verification (i.e. assumes conventional QA (Quality Assurance) would not be required or could be performed in real-time).

### 1.6.3 Anatomy Segmentation

This research project, described in Chapter 5, utilized a deep learning model to perform automatic segmentation of organs-at-risk on computed tomography scans of head and neck cancer patients. The model was based on the U-Net architecture with embedded Inception-ResNet-v2 blocks and had not yet been investigated or validated for contouring of the head and neck region. Additionally, this was the first study that provided a comprehensive interpretation of contour predictions for the head and neck anatomy in the form of Gradient-weighted Class Activation Maps. The study included a total of 964 patients (from a public access database) and 25 critical organs contoured. The model was trained for each anatomical structure with 65% of training and 25% of validation data and was evaluated on the remaining 10% of CT scans delineated by radiation oncologists. The performance metrics included Dice Score, Jaccard Index, and Hausdorff Distances and were calculated comparing the model-generated segmentation to the reference (i.e. physician-generated) segmentations. This work showed that the U-Net architecture with Inception-ResNet-v2 blocks achieved satisfactory accuracy (almost all structure results were better than the literature, and better than inter-physician variability), very short contouring times, and realistic prediction reasoning, making it a feasible clinical solution

for segmentation of the head and neck regional anatomy to both prepare and then adapt radiation therapy treatment plans.

### 1.6.4 Patient Selection

The objective of the work described in Chapter 6 was to utilize a machine learning model to create a novel decision support system for brain tumor patients. The binary classification model was developed based on 79 treated patients for whom two-arc VMAT treatment plans had been created. Data for model training consisted of five dosimetric and five geometric features extracted from the previously delivered radiation therapy treatment plans. The target variable included class-0, corresponding to plans for which the PTV treatment planning objective was met, and class-1, which was associated with plans for which the PTV objective was not met due to the priority trade-off to meet one or more competing OAR constraints. Four common training models were initially evaluated: Support Vector Machine, Elastic Net, Logistic Regression, and Random Forest using double-nested cross-validation and the AUC (area under receiver operating characteristic curve) metric. The model predictions were explained both globally and locally with Shapely additive explanation (SHAP) interaction values. The study results showed that the highest-performing model was Logistic Regression, achieving an accuracy of 93.8±4.1% and AUC of 0.98±0.02 on the test data. In addition to analysing the model performance, this research work also provided model explainability which is rarely performed in applications of ML in radiation therapy but has been identified as a strong need in the research community in order to properly interpret model results. Therefore, the included explainability analysis may be particularly valuable to both medical physics and radiation oncology professionals working in a clinical environment. From an adaptive

radiation therapy standpoint, new treatment plans classified as those for which the PTV objective was not met due to the OARs priority trade-offs (class-1) would be considered as the most suitable for plan adaptation throughout the treatment to meet the PTV objective while minimizing or eliminating the mentioned trade-offs.

## 1.7 Summary

Adaptive radiation therapy is an advanced form of radiation therapy that involves additional procedures performed to modify the initially created treatment plan. ART facilitates the delivery of the prescribed highly conformal radiation doses to the target structure(s) while minimizing irradiation of critical organs and other healthy tissues, by accounting for changes in the patient's anatomy throughout the course of treatment.

Due to the strong interest in ART, many different approaches for radiation treatment adaptation have been investigated[39-41]. However, despite many promising results from those research efforts, there are still several challenges to overcome to make plan adaptation more reliable and clinically feasible for a large number of patients. Several of these challenges are associated with the suboptimal performance of tools that are used in a few crucial steps in the ART framework, including:

a) the application of CT-CBCT deformable image registration
b) selection of specific ART strategies
c) automated and accurate segmentation of patient anatomy
d) efficient categorization of patients who would be most likely to benefit from ART

This dissertation aims to propose and evaluate solutions in these four challenging areas in order to make adaptive radiation therapy more reliable, accurate, and efficient for selected clinical disease sites.

These specific research goals were achieved through the following scientific contributions:

a) identification and evaluation of dense anatomical block matching algorithm for CT-CBCT deformable image registration in the case of prostate cancer patients
b) design and evaluation of eight adaptive radiation therapy strategies as well as a selection of the most promising ART approach for hypofractionated prostate cancer patients
c) development of data processing and deep learning model training pipelines for automatic segmentation of computed tomography scans for radiation therapy of head and neck cancer patients
d) development of a machine learning based system for identification of those treatment plans that would most likely require plan adaptation throughout the treatment.

These four investigations are described in detail in Chapters 3-6 of this thesis, while Chapter 2 provides an introduction to the main concepts required to understand the research chapters. The topics covered in Chapter 2 include fundamentals of radiation therapy, deformable image registration, adaptive radiation therapy, machine learning, and deep learning. The thesis concludes in Chapter 7 which gives a summary of the scientific work performed here, as well as opportunities for future investigations. An appendix is also provided to supplement information in the main scientific chapters.

## 1.8 References

1. Poirier AE, Ruan Y, Walter SD, et al. The future burden of cancer in Canada: Long-term cancer incidence projections 2013–2042. *Cancer Epidemiology.* 2019;59:199-207.

2. Wild CP, Stewart BW, Wild C. *World cancer report 2014.* World Health Organization Geneva, Switzerland; 2014.

3. Baskar R, Itahana K. Radiation therapy and cancer control in developing countries: Can we save more lives? *International journal of medical sciences.* 2017;14(1):13.

4. Baskar R, Lee KA, Yeo R, Yeoh K-W. Cancer and radiation therapy: current advances and future directions. *International journal of medical sciences.* 2012;9(3):193.

5. Washington CM, Leaver DT. *Principles and Practice of Radiation Therapy-E-Book.* Elsevier Health Sciences; 2015.

6. Martin A, Thomas S, Harden S, Burnet N. Evaluating competing and emerging technologies for stereotactic body radiotherapy and other advanced radiotherapy techniques. *Clinical Oncology.* 2015;27(5):251-259.

7. Garcia R, Nyström H, Fiorino C, Thwaites D. Does a too risk-averse approach to the implementation of new radiotherapy technologies delay their clinical use? *The British journal of radiology.* 2015;88(1051):20150124.

8. Gardner SJ, Kim J, Chetty IJ. Modern Radiation Therapy Planning and Delivery. *Hematology/Oncology Clinics.* 2019;33(6):947-962.

9. Brady LW, Heilmann H, Molls M. *New technologies in radiation oncology.* Springer; 2006.

10. Lee S, Cao YJ, Kim CY. Physical and radiobiological evaluation of radiotherapy treatment plan. *Evolution of ionizing radiation research Rijeka, Croatia: InTech.* 2015.109-150.

11. Perez CA, Vijayakumar S, Levitt S, Purdy JA. *Technical Basis of Radiation Therapy Practical Clinical Applications.* Springer; 2012.

12. Khan FM, Gibbons JP. *Khan's The Physics of Radiation Therapy.* Wolters Kluwer Health; 2014.

13. Nabavizadeh, Nima, et al. "Image guided radiation therapy (IGRT) practice patterns and IGRT's impact on workflow and treatment planning: Results from a national survey of American Society for Radiation Oncology members." *International Journal of Radiation Oncology* Biology* Physics* 94.4 (2016): 850-857.

14. Arabloo, Jalal, et al. "Health technology assessment of image-guided radiotherapy (IGRT): A systematic review of current evidence." *Medical journal of the Islamic Republic of Iran* 30 (2016): 318.

15. Li XA. *Adaptive Radiation Therapy.* CRC Press; 2011.

16. Keall P, Nguyen DT, O'Brien R, et al. Stereotactic prostate adaptive radiotherapy utilising kilovoltage intrafraction monitoring: the TROG 15.01 SPARK trial. *BMC cancer.* 2017;17(1):180.

17. Bohoudi O, Bruynzeel A, Senan S, et al. Fast and robust online adaptive planning in stereotactic MR-guided adaptive radiation therapy (SMART) for pancreatic cancer. *Radiotherapy and Oncology.* 2017;125(3):439-444.

18. Keall P, Booth J, Nguyen D, et al. The first clinical implementation of real-time adaptive radiation therapy using a standard linear accelerator. *International Journal of Radiation Oncology • Biology • Physics.* 2017;99(2):S223-S224.

19. Lamb J, Cao M, Kishan A, et al. Online adaptive radiation therapy: implementation of a new process of care. *Cureus.* 2017;9(8).

20. Lim-Reinders S, Keller BM, Al-Ward S, Sahgal A, Kim A. Online Adaptive Radiation Therapy [published online ahead of print 2017/09/17]. *Int J Radiat Oncol Biol Phys.* 2017;99(4):994-1003.

21. Liu C, Kim J, Kumarasiri A, et al. An automated dose tracking system for adaptive radiation therapy. *Computer methods and programs in biomedicine.* 2018;154:1-8.

22. Posiewnik M, Piotrowski T. A review of cone-beam CT applications for adaptive radiotherapy of prostate cancer. *Physica Medica.* 2019;59:13-21.

23. Veresezan O, Troussier I, Lacout A, et al. Adaptive radiation therapy in head and neck cancer for clinical practice: state of the art and practical challenges. *Japanese Journal of Radiology.* 2017;35(2):43-52.

24. Rigaud, Bastien, et al. "Deformable image registration for radiation therapy: principle, methods, applications and evaluation." *Acta Oncologica* 58.9 (2019): 1225-1237.

25. Srinivasan, Kavitha, Mohammad Mohammadi, and Justin Shepherd. "Applications of linac-mounted kilovoltage Cone-beam Computed Tomography in modern radiation therapy: A review." *Polish journal of radiology* 79 (2014): 181.

26. Wang, Tonghe, et al. "Dosimetric study on learning-based cone-beam CT correction in adaptive radiation therapy." *Medical Dosimetry* 44.4 (2019): e71-e79.

27. Posiewnik, M., and T. Piotrowski. "A review of cone-beam CT applications for adaptive radiotherapy of prostate cancer." *Physica Medica* 59 (2019): 13-21.

28. Gandhi, Meet, Juhi Kamdar, and Manan Shah. "Preprocessing of non-symmetrical images for edge detection." *Augmented Human Research* 5.1 (2020): 1-10.

29. Zhang, Xiaoli, Xiongfei Li, and Yuncong Feng. "A medical image segmentation algorithm based on bi-directional region growing." *Optik* 126.20 (2015): 2398-2404.

30. Singh, J. Fenshia, and V. Magudeeswaran. "Thresholding based method for segmentation of MRI brain images." *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. IEEE, 2017.

31. Seo, Hyunseok, et al. "Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications." *Medical physics* 47.5 (2020): e148-e167.

32. Zaffino, Paolo, et al. "Multi atlas based segmentation: should we prefer the best atlas group over the group of best atlases?." *Physics in Medicine & Biology* 63.12 (2018): 12NT01.

33. Tajbakhsh, Nima, et al. "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation." *Medical Image Analysis* 63 (2020): 101693.

34. Cai, Lei, Jingyang Gao, and Di Zhao. "A review of the application of deep learning in medical image classification and segmentation." *Annals of translational medicine* 8.11 (2020).

35. Zhou, Tongxue, Su Ruan, and Stéphane Canu. "A review: Deep learning for medical image segmentation using multi-modality fusion." *Array* 3 (2019): 100004.

36. Valdes, Gilmer, et al. "Clinical decision support of radiotherapy treatment planning: A data-driven machine learning strategy for patient-specific dosimetric decision making." *Radiotherapy and Oncology* 125.3 (2017): 392-397.

37. Chan, Maria F., Alon Witztum, and Gilmer Valdes. "Integration of AI and machine learning in radiotherapy QA." *Frontiers in artificial intelligence* 3 (2020): 76.

38. Brouwer CL, Steenbakkers RJ, Langendijk JA, Sijtsema NM. Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? *Radiotherapy and oncology.* 2015;115(3):285-294.

39. Dial C, Weiss E, Siebers JV, Hugo GD. Benefits of adaptive radiation therapy in lung cancer as a function of replanning frequency. *Medical physics.* 2016;43(4):1787-1794.

40. Lim K, Stewart J, Kelly V, et al. Dosimetrically triggered adaptive intensity modulated radiation therapy for cervical cancer. *International Journal of Radiation Oncology* Biology* Physics.* 2014;90(1):147-154.

41. van Kranen S, Hamming-Vrieze O, Wolf A, Damen E, van Herk M, Sonke JJ. Head and Neck Margin Reduction With Adaptive Radiation Therapy: Robustness of Treatment Plans Against Anatomy Changes [published online ahead of print 2016/09/30]. *Int J Radiat Oncol Biol Phys.* 2016;96(3):653-660.

# Chapter 2

# Technical Background

This chapter will provide a broad overview of several topics that are crucial for understanding subsequent research chapters of the dissertation. In particular, the introduction will include a general description of radiobiology and radiation therapy where intensity-modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT) will be briefly characterized. The subsequent sections will discuss important concepts related to deformable image registration, adaptive radiation therapy, machine learning as well as deep learning.

## 2.1 Foundations of Radiation Therapy

### 2.1.1 Radiobiology

Radiobiology is a scientific discipline that focuses on the impact of radiation on biological tissue. In the context of radiation therapy, radiobiology attempts to describe the effects of high energy, ionizing radiation on the patient's healthy tissues and the treated volume containing cancerous cells. Advances in radiobiological research allow for a better understanding of the mechanisms behind cellular death and mutations induced by ionizing

radiation as well as help to develop robust dose fractionation schemes used in modern radiation therapy[1].

The relationship between tumor control and the radiation tolerance of healthy tissues is crucial for the successful delivery of radiation therapy. Radiobiological research has shown that both the tumor control probability (TCP) and normal tissue complication probability (NTCP) with respect to the total dose delivered to the patient can be modeled by sigmoidal curves as shown in Figure 2.1. The objective of radiation therapy is to maximize the width of the therapeutic window i.e., maximize the probability of the tumor control while minimizing NTCP. Figure 2.1 also shows that an increase in the per fraction dose (from 2 Gy to 2.5 Gy) narrows the therapeutic window by decreasing the maximum tolerable dose for normal healthy tissues making this approach potentially less beneficial for critical structures[2,3].

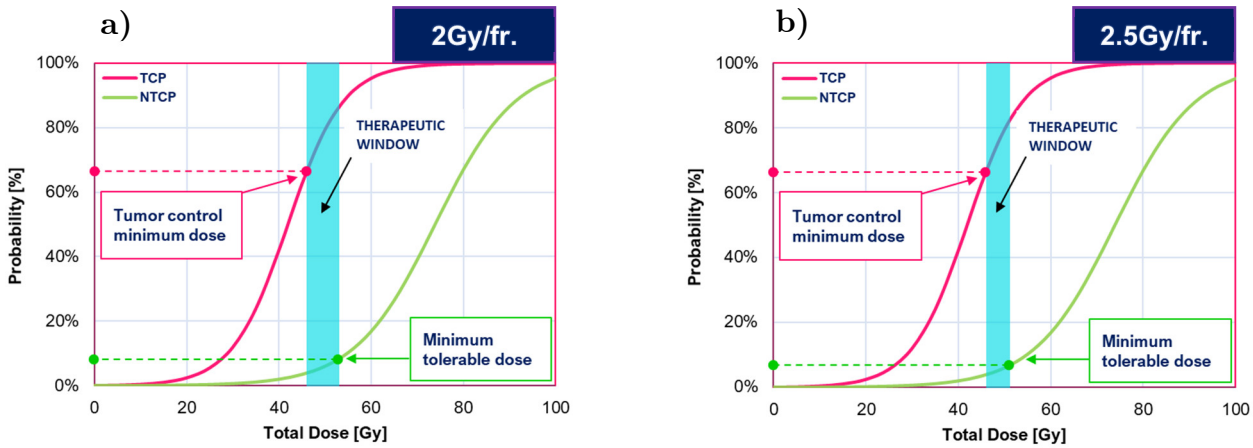



**Figure 2. 1.** Tumor control probability (TCP) and normal tissue complication probability (NTCP) curves for **a)** 2Gy per fraction and **b)** 2.5Gy per fraction treatment[2].

### 2.1.2 Target Volume Definition

The target volume is delineated with GTV, CTV, and PTV contours, the meaning of which is briefly explained in Figure 2.2. The GTV stands for 'gross tumor volume' and

segments the macroscopic distribution of cancer cells as seen on imaging scans or determined through physical examination. In cases where the tumor cannot be properly visualized with Computed Tomography imaging, high contrast Magnetic Resonance Imaging (MRI) is used to facilitate the contouring process. The CTV is the 'clinical target volume' (GTV plus any estimated microscopic extension of disease). The delineation of the CTV depends on the type of cancer, documented patterns of disease spread, and can be guided by functional imaging. The PTV is the 'planning target volume' and includes the CTV plus a margin that incorporates patient setup errors and target motion and is specific to the patient's immobilization and the image guidance technique used.

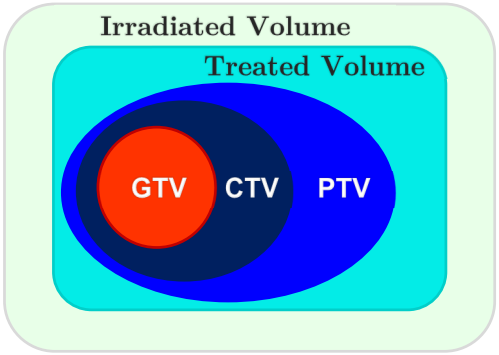



**Figure 2.2.** Gross tumor volume (GTV) delineates the volume(s) of a known tumor; clinical target volume (CTV) is the GTV plus the volume of suspected microscopic tumor infiltration; planning target volume (PTV) represents the volume containing the CTV/GTV with an additional margin to account for setup variations and organ and patient motion[4]. Treated volume is the volume covered by an isodose surface and defined by a radiation oncologist as suitable for achieving an objective of the treatment, while irradiated volume is a tissue absorbing a radiation dose in the magnitude that is deemed significant with respect to the normal tissue tolerance.

### 2.1.3 Dose Volume Histogram

Volume calculations performed by 3D treatment planning systems provide a large amount of information based on the 3D dose calculations. The three dimensional dose distributions need to be reviewed by medical physicists and radiation oncologists to ensure their appropriateness for treatment. This information can be difficult to interpret and evaluate when it is displayed as isodose curves on transverse, sagittal and coronal planes. For the purpose of plan analysis, it is much simpler to condense the 3D dose distribution data to a 2D graph, which displays the radiation accumulated within a specifically defined volume of interest of a structure (such as a target structure that encompasses the known or suspected cancerous cells, or an organ-at risk that one wants to minimize dose to). Such a graphical representation is called a dose volume histogram (DVH). The dose volume histogram, and the numerous dose-volume metrics that can be derived from it, is one of the most important tools available when quantitatively evaluating a radiation therapy treatment plan. This section will explain in detail the DVH concept.

Let's consider a volume that has been delivered a radiation dose, so that its distribution is non-uniform within that volume. In order to create a DVH based on that distribution, the considered volume is first divided into a three-dimensional grid of voxels, and the dose within each individual volume element is considered uniform. Ideally the voxels are small (single millimetre range) in order to provide accurate sampling of dose information throughout the 3D structures. Within an individual structure, the dose of the encapsulated voxels is then binned (by dose magnitude) as a histogram without considering their location. A chart that shows the number of voxels in each dose bin on the y-axis as a function of bin dose range on the x-axis is called a differential DVH (Figure 2.3a show an example of this).

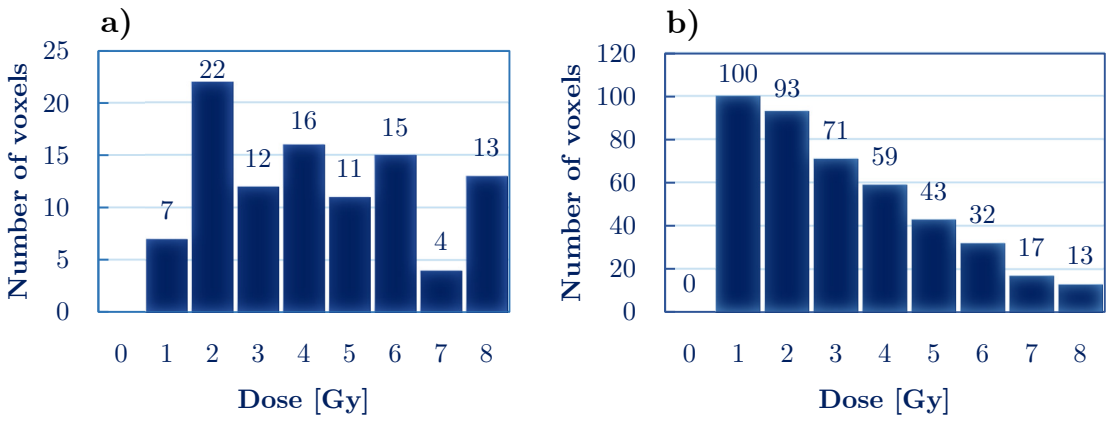



**Figure 2.3. a)** Differential dose volume histogram. The number of voxels on the y-axis is in practice expressed in $cm^3$ of the volume (as an absolute measure) or in the percentage of volume (as a relative measure) occupied by a specific number of voxels. **b)** Cumulative dose volume histogram[5]. The total number of voxels in this illustrative example is 100.

It should be noted that the size of the dose bin defines the height of each histogram column of the differential DVH. For example if the bin widths were to be widened, the height of the histogram bin would increase due to the larger number of voxels falling within a bin range. Therefore it is worth noting that even though the dose-volume histogram data remain unchanged, the exact shape of the differential DVH depends on the width of the dose bin.

Although a differential DVH is an important concept, a cumulative DVH is more often used in clinical applications. A cumulative DVH is a plot in which each dose bin defines the absolute or the percentage volume on the y-axis that receives a dose *greater or equal to* the dose indicated on the x-axis. As shown in the Figure 2.3b the value at any dose bin is calculated by the summation of the number of voxels of the corresponding differential DVH to the right of that dose bin. Therefore, it can be observed that the relative volume of the first dose bin to the left is 100% (i.e. total volume). This means

that the total volume receives at least zero dose. However, the volume of the last dose bin to the right receives the maximum dose because no more voxels receive a dose higher than that.

As mentioned above, specific dose-volume metrics can be extracted from the DVH curves. For example, maximum and minimum point doses ($D_{min}$, $D_{max}$), the percentage of volume receiving greater than or equal to a specific absolute dose (e.g. $V_{30Gy}$ is the volume receiving at least 30Gy) or percentage of a prescription dose (e.g. $V_{20\%}$ is the volume receiving at least 20% of the prescription dose). Another commonly encountered dose metric nomenclature is the absolute dose received by at least a specified percentage of the volume of considered structure (e.g. $D_{95\%}$ is the dose in Gy or cGy, also commonly expressed as a percentage of prescription dose, received by 95% of the structure). Dose-volume metrics defined in these ways are often used for defining treatment planning objectives that need to be met for the treatment plan to be approved by the radiation oncologist. Additionally, DVH are also used for direct quantitative comparison of different competing treatment plans that are being considered to treat a patient[5,6].

### 2.1.4 Intensity Modulated Radiation Therapy

Intensity-modulated radiation therapy (IMRT) is an irradiation technique that utilizes many radiation beams with inhomogeneous intensity distributions in order to create a homogenous dose distribution inside the treated volume while minimizing the amount of dose delivered to surrounding healthy tissues. As shown in Figure 2.4, a uniform dose distribution within the target volume is achieved by the superposition of intensity-modulated radiation beams with different angles of incidence that overlap in the target

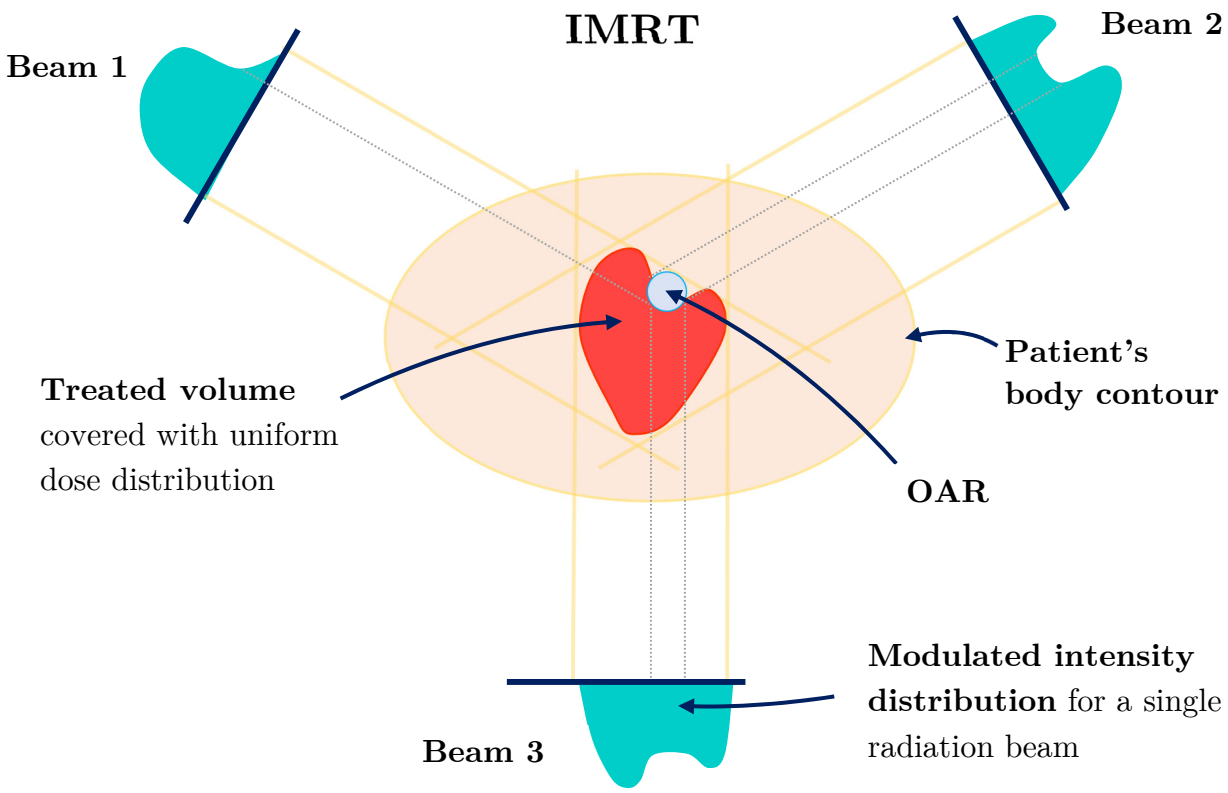



**Figure 2. 4.** The basic principle of operation for intensity-modulated radiation therapy (IMRT).

region in the patient's anatomy. The beam intensity and its shape are adjusted by a multi-leaf collimator (MLC) – a device mounted to the head of a linear accelerator as illustrated in Figure 2.5[7-9].

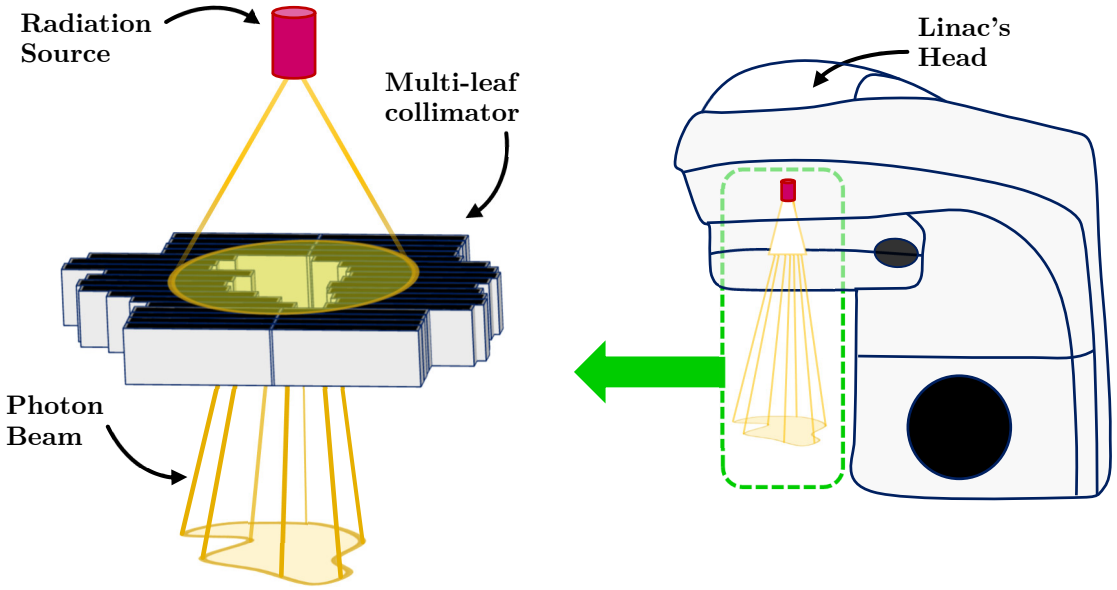



**Figure 2. 5.** Multi-leaf collimator (MLC) attached to the head of the linac. Leaf 3.2.movement modulates the intensity of the radiation beam as well as conforms the shape of the beam to the shape of the treated volume.

### 2.1.5 Volumetric Modulated Radiation Therapy

Similar to IMRT, volumetric modulated arc therapy (VMAT) delivers the therapeutic dose using intensity-modulated photon beams. However, compared to IMRT where each intensity-modulated beam is delivered at fixed gantry angles (ie. the gantry is not moving while the beam is on), VMAT usually uses one or two arcs that continuously deliver the radiation dose while the gantry of the linear accelerator rotates around the patient, as shown in Figure 2.6. During the rotation, both the dose rate and the gantry speed can vary. The increased efficiency of the VMAT technique compared to IMRT allows it to complete the dose delivery in a significantly shorter amount of time, without compromising the treatment quality[9,10,11].

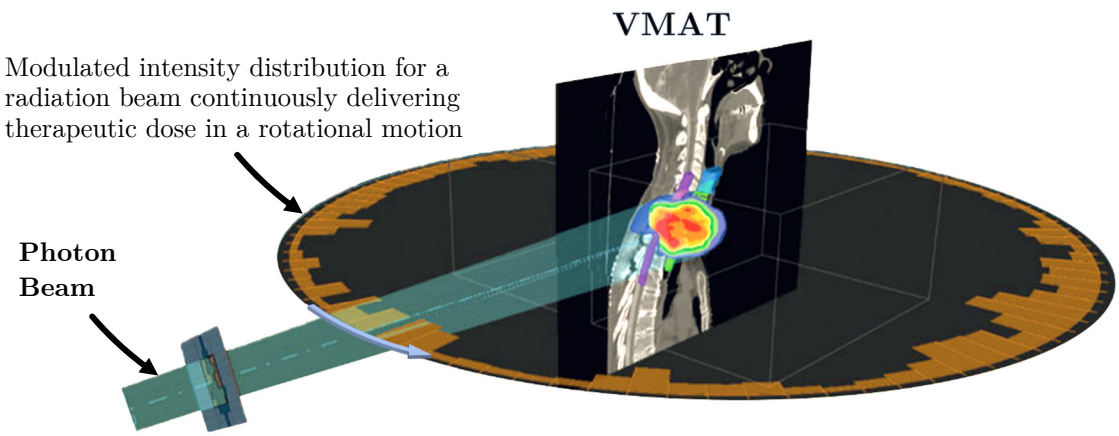



**Figure 2. 6.** *The basic principle of operation for volumetric modulated arc therapy (VMAT)[12].

### 2.1.6 IMRT Plan Optimization

IMRT optimization is the process of generating a clinically optimal plan for an IMRT beam delivery to a particular patient. The task of generating an optimal plan generally can be divided into two main parts: (i) construction of an objective function, and (ii) implementation of an optimization algorithm. The objective function, mathematically, is a function that defines a plan's quality in the form of a single number and is composed of many individual terms that describe the desirable qualities the plan should have. The optimization algorithm seeks to maximize (or minimize) the objective function, by exploring the potential solution space[13]. The potential solutions are the various fluence patterns that can be delivered by the linac, and these need to be mathematically modeled and used as input into the optimization algorithm. The fluence patterns also have to be

* The figure was inspired by and modified based on the work of Baumann et al.[12] – permission for usage in this thesis was obtained from Springer Nature License.

restricted to those that can be physically deliverable by the radiation treatment unit (i.e. respecting limitations such as maximum leaf velocity, dose rates, etc.)

The subsequent paragraphs will describe the general types of optimization constraints, the different forms of objective function terms associated with particular objectives, and the most common optimization algorithms used for IMRT optimization.

### Hard and Soft Constraints

A constraint that the optimization algorithm must follow rigorously is called a 'hard constraint'. For instance, consider the objective that the maximum dose delivered to the spinal cord should not be higher than 45 Gy. If this condition is selected as a hard constraint, the final solution *must* satisfy it. However, a 'soft constraint' is one that does not have to be strictly followed during the optimization calculations, typically being penalized in some manner, but allowed. The penalties often depend on the degree of such a soft constraint violation and need to be mathematically incorporated into the objective function[14,15].

### Objective Function

The objective function contains terms that quantitatively compare the dosimetric qualities of the current plan being considered (modeled based on the fluence patterns under consideration) with the same qualities of the desired dose distribution. Each individual term corresponds to a mathematical representation of a specific aspect of dosimetric quality of the desired plan. Commonly, the objective function is a summation of the squared differences between the dosimetric quantity in the current plan iteration (i.e. the current plan under consideration) and the desired dosimetric quantity. Selection of the

form of the objective function to use is an important consideration in IMRT optimization. The two common types of objective functions are (i) those based on dose objectives that rely on the physical dose distributions in the defined structures only, and (ii) those based on radiobiological objectives, which depend on the physical dose distributions as well as the radiobiological dose responses of the individual organs/structures[15].

*Optimization Algorithms*

After the type of objective function (physical or radiobiological) and relevant objective parameters/terms (i.e. target and OARs structures, types and specific values of the constraints) have been selected, each treatment beam is modeled with 'beamlets'. The beamlets are simply discretized, small beams sampling the intensity of the incident fluence pattern to be delivered, and this intensity can be modified higher or lower during the optimization process as the algorithm attempts to satisfy the defined dose objectives. Next, the search for an optimal IMRT solution may start. This is usually an iterative process as summarized by the framework illustrated in Figure 2.7.

The most commonly encountered optimization methods in IMRT treatment planning can be divided into two groups: deterministic algorithms and stochastic algorithms[13]. Although it is beyond the scope of this introduction it is noted that the "Steepest Decent" and "Newton's Method" approaches belong to the deterministic algorithms and the Simulated Annealing method is an example of a stochastic algorithm.

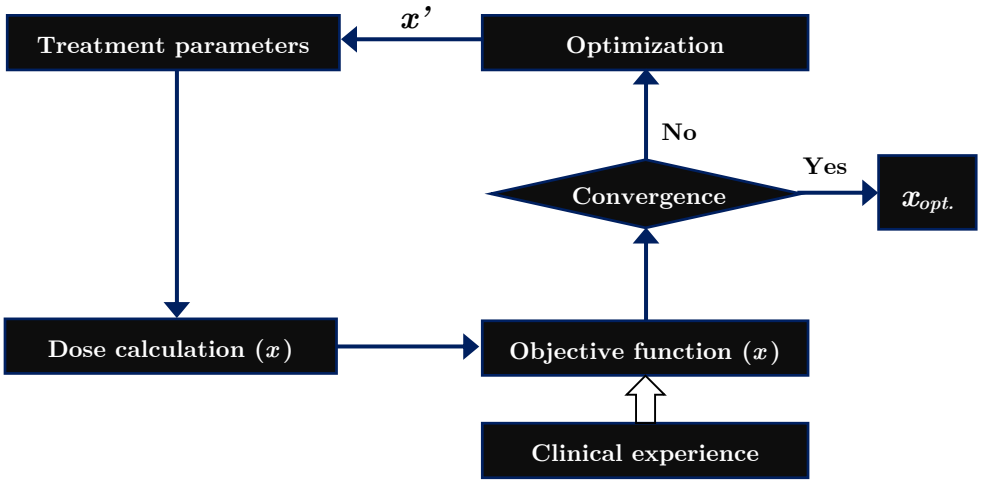



**Figure 2.7.** The framework of iterative IMRT optimization. Based on a set of initial treatment parameters $x$ (here defining the weights/intensities of the particular beamlets) a 3D-dose distribution is determined. Next, the objective function is used to evaluate the quality of the proposed treatment plan with the consideration of predetermined treatment goals. If the quality of the treatment plan is an improvement over the previous iteration, the current value $x$ becomes the optimum $x_{\text{opt}}$. However, in the case the quality of the treatment plan worsens (ie. the objective function worsens), then the optimization algorithm selects new fluence amplitudes $x$' for the subsequent iteration step. The process of IMRT plan optimization continues until a solution of $x_{opt.}$ is found by minimization of the objective function[13].

### 2.1.7 VMAT Plan Optimization

For VMAT delivery, the radiation beam continuously delivers the treatment while the gantry rotates around the patient, often utilizing modulated gantry speed, varying dose rate, and simultaneously moving MLC leaves. Since the beam intensity is constantly modulated with many parameters changing simultaneously, VMAT is considered a technically more advanced form of IMRT. For that reason, much of the information provided in the previous section is relevant to VMAT as well. This section will specifically

focus on the description of arc discretization and control points, deliverability constraints and a basic VMAT optimization model.

*Arc Discretization and Control Points*

Most VMAT optimization methods are based on the discretization of an arc into a finite set of control points (CP), so that each CP represents a single beam direction (representing a small portion of the entire arc, typically about 2 degrees of gantry rotation), with a defined gantry speed, dose rate, and beam aperture shape. Aperture shape, dose rate, and gantry speed specified for the control point, determine the discrete set of the required delivery parameters of the VMAT treatment in the continuous gantry rotation as the gantry passes through the control point[16].

*Deliverability Constraints*

The apertures and the gantry speed for consecutive control points have to be physically achievable. This means that when the gantry rotates from control point $k$ to control point $k+1$, the leaves of the MLC need to shift from the configuration represented by aperture $A_k$ (for CP $k$) to the configuration represented by aperture $A_{k+1}$ (for CP $k$+1). Therefore, the time the gantry spends traveling between control points $k$ and $k$+1 needs to be large enough to enable the desired aperture change, taking into account the limited speed of the MLC leaves. In addition to speed limitations, the configurations of all possible apertures that can be delivered by a given MLC need to be taken into account.

Other physical limitations that need to be accounted for are the upper and the lower limits of the achievable dose rate and the upper and lower bounds of gantry speed. Another restriction is the rate at which the gantry speed can be changed (i.e. the maximum

acceleration/deceleration) and is usually expressed per control point or per specified degree of gantry orientation.

### VMAT Optimization Model

To simplify the problem, the dose delivered to each voxel over the course of radiation treatment can be calculated with the assumption that the aperture ($\mathbf{A}$), dose rate ($\mathbf{r}$), and gantry speed ($\mathbf{s}$) do not change between control points. In the case where the objective function can be described by $\mathcal{F}(\mathbf{z}, \mathbf{r}, \mathbf{s}, \mathbf{A})$, and assuming that minimum value of $\mathcal{F}$ would correspond to the optimal solution (optimal treatment plan) the VMAT optimization problem together with the relevant constraints can be formulated as minimization of $\mathcal{F}(\mathbf{z}, \mathbf{r}, \mathbf{s}, \mathbf{A})$ where $\mathcal{F}(\mathbf{z}, \mathbf{r}, \mathbf{s}, \mathbf{A})$ is usually a function of the delivered dose $\mathbf{z}$, which depends on deliverable parameters $\mathbf{r}$, $\mathbf{s}$, and $\mathbf{A}$. As can be observed from this formulation, in comparison to IMRT, the VMAT optimization is more complex due to the increased number of variables and delivery constraints and thus the amount of data that needs to be processed in the optimization calculations is larger. For instance, a typical IMRT treatment plan contains 5-9 beams each at a unique gantry angle. However, in the case of the VMAT, the number of control points per treatment plan is frequently on the order of a few hundred[16].

## 2.2 Deformable Image Registration

### 2.2.1 Image Registration Process

As briefly mentioned in the rationale section (1.4.1), image registration is the process of establishing the spatial relationship/correspondence between two or more images of the same dimensionality (2D or 3D), with the goal of mapping one image (moving or floating image) to the other (fixed or reference image). The registration of two images is generally a two-step process. In the first step, the images are pre-processed with simple cropping, intensity-based image filtering, rigid transformation including shift, scale, and rotate operations, or may even be manually matched in order to increase the initial geometric and voxel intensity-based similarity. In this step, various image attributes such as contours or landmarks can also be specified and used to guide image registration. The second step involves an iterative registration loop in which the optimization algorithm, with regularization terms and specified stopping criteria, is adjusting the parameters of the transformation in order to minimize (or maximize) the value of the similarity metric. The registration loop is completed by reaching either the maximum number of iterations, the maximum registration time, or when there is no further improvement in the similarity metric. After completion, the transformation that maps a moving image to a reference image is generated. The transformed moving image (i.e., moving image after the application of the transformation) is output as well[17]. Figure 2.8. shows the diagram of the image registration process including all the components described in more detail in the subsequent sections (2.2.2 – 2.2.5).

### 2.2.2 Transformations

There are several types of transformations used in image registration that vary in their complexity. The simplest transformation, called Rigid, can be represented by a translation or rotation matrix, each having three (for 3D image) degrees of freedom. Additional flexibility can be introduced by shearing, scaling, and reflection operations that result in an Affine transformation. Rigid and affine registrations, although often useful for initial alignment or simple image matching tasks, are fairly limited in their ability to map medical images, mainly due to the global nature of their transformations[18].

The application of image registration in the medical imaging domain requires the spatial transformation to be significantly more flexible. Such transformations are frequently represented by a three-dimensional, deformation vector field (DVF) that can perform any non-linear transformation by manipulating an image at the voxel level (i.e. the position of each voxel can be changed by a corresponding vector). In practice, however, the level of DVF flexibility needs to be controlled by regularization terms during the optimization step of image registration.

Constraints imposed on the vector field are designed to ensure that the transformation resulting from deformable image registration will reflect realistic deformations of the human anatomy. The class of image registration that applies non-linear DVF for image transformation is called *deformable* image registration (DIR)[18].

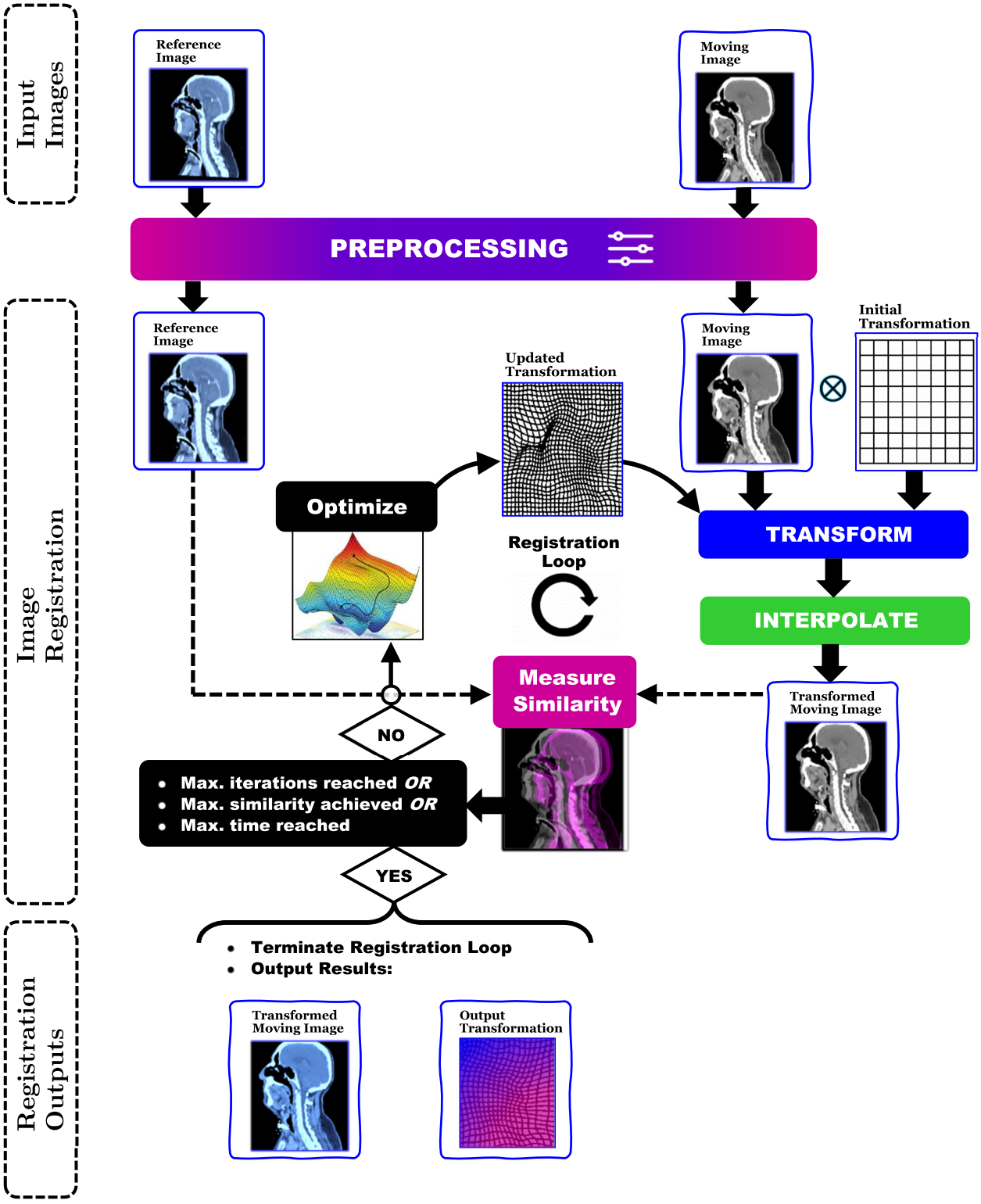



**Figure 2. 8.** Schematic illustration of major steps in the image registration process. A parallel workflow of both the moving and reference images starts at the top of the diagram.

In general the transformation model representing DIR can be categorized as parametric or non-parametric. The former produces a DVF as a linear combination of their basis functions, while the latter computes transformation vectors of all points in the fixed image[19].

### 2.2.3 Similarity Metrics

There are two main groups of similarity metrics used in image registration: feature-based and voxel intensity-based. The first type relies on the Euclidean distance between the same features associated with the image such as volumes, surfaces, or points. The major disadvantage of this approach is that the feature definition is very time-consuming and must be conducted manually, typically by a physician who has in-depth knowledge of the relevant anatomical region. Therefore, in deformable registration of medical images, intensity-based image similarity metrics are more commonly implemented. For moving and fixed images that were acquired by the same imaging modality, sum of squared differences, cross-correlation or normalized cross-correlation are the metrics most often used. For multi-modality image registration where intensities of pixels belonging to the same part of anatomy differ among registered images, mutual information or normalized mutual information metrics are applied[18,20].

### 2.2.4 Optimization

The main objective of the optimizer used in deformable image registration is to maximize the alignment (measured with a similarity metric) between registered images by manipulating the parameters of the vector field. Optimizers find the optimal solution

iteratively (i.e. continuously changing parameters of the transformation until the similarity metric converges or other loop-stopping criteria are met).

The search for the optimal solution is achieved by finding a search direction and an efficient exploration of the parameter space. The optimization algorithms suitable for this task within the medical image registration domain are most commonly gradient descent-based techniques that compute the gradient of the similarity metric function relative to the parameters of the vector field (along the individual dimensions of each vector).

In practice, the optimization process is often performed at multiple image resolution levels by first down-sampling both the moving image and the reference image. Next, both images are registered from the lower to the higher resolutions. Specifically, the transformation that is optimized at the lower resolution is subsequently used as the initial solution for the optimization at the higher resolution. Compared to a single resolution, the multiresolution approach was found to make the optimization process more robust with respect to being trapped in local optimums (e.g. a local minimum or maximum of the objective function over a small search space, which is not the global minimum or maximum over the entire possible search space) and also more time efficient[17,20].

### 2.2.5 Interpolation

In the process of image registration, when the floating image is being aligned with the fixed image, voxels of the floating image are mapped to the target image. Since the domains of both images are discrete but the mapping result is not discrete, the points of the floating image are mapped to locations on the fixed image that are not necessarily on its regular voxel grid.

Therefore, the intensity values at the exact points of the fixed image grid need to be estimated by interpolation. Currently, there are many interpolation methods such as linear, bilinear, trilinear, nearest neighbor, or cubic, among others. The interpolation quality as well as computational efficiency vary between methods and usually constitutes a common quality-to-time trade-off[21].

### 2.2.6 Validation

In medical imaging applications, there are two main methods of validating the accuracy of registration: visual (qualitative) and numerical (quantitative) methods. Within a clinical environment, the initial assessment performed by a physician is usually visual and can reveal major anatomical misalignments between registered images. This form of evaluation is conducted based on the inspection of anatomical landmarks, image and segmentation overlays, checkerboard filter, dynamic magnifying window, and/or examination of absolute image differences. The qualitative evaluation can also include the analysis of the vector field on various image projections as well as an inspection of the absolute magnitude map of the vector field.

The quantitative evaluations most commonly used include the calculation of geometric distances between anatomical landmarks and contours, computation of segmentation overlay metrics, and in the case of monomodal registration (for images acquired by the same imaging modality) the sum-of-square differences relative to the image intensities. The interpretation of any metric based on the similarity of voxel intensities should however be conducted with caution because these metrics provide a single number without the reference to the location of misalignment. When it comes to the vector field, the

numerical assessment is usually based on the inverse consistency error (ICE) and determinant of the Jacobian matrix. Both these metrics are further discussed in Chapter 3[17,18].

### 2.2.7 Challenges of CT to CBCT Image Registration in Radiotherapy

Due to the common use of CBCT imaging in adaptive radiation therapy, the main challenge for the image registration (one of the crucial steps in any ART process), is the poorer image quality of the CBCT scans (fixed images) as well as the inconsistency in their voxel intensities relative to the CT images (moving images). In particular, the inconsistency in voxel intensities between image sets (for example in multimodality image pairs) means that the same type of tissue may not be represented by the same voxel intensity. The poor quality of CBCT images is also the reason why it is more beneficial to register pCT to CBCT and not CBCT to pCT. For example, due to correct representation of attenuation coefficients of the patients' tissue, the patient dose calculation will be more accurate on deformed pCT rather than on deformed CBCT.

Moreover, there are a few other issues that affect the robustness and accuracy of deformable image registration. One of these issues is the presence of artifacts appearing in the images due to dental fillings, metal prosthesis, stents, etc. present in the patient's anatomy. Artifacts can also be seen in images due to the patient or individual organ motion during the image acquisition[17,18].

Large changes in the patient's anatomy due to weight loss, response to therapy, or bladder filling, require large local deformations and can also cause difficulties in the image registration. For example, too large or too small gradient step may cause the optimization

process to converge very slowly, converge to a local minima or not converge at all. Other anatomical issues include that of sliding motion that sometimes exists between two adjacent tissues/organs, for example, commonly encountered between the lungs and the thoracic cage. Beyond large local deformations, the sliding motion issue requires an estimate of a 'quasi-discontinued DVF' (challenging to estimate locally in alternating regions of a vector field with very large and very small deformations)[17,18].

Other image registration challenges may also include scenarios where certain parts of the patient's anatomy are visible on one image but are not visible on the other image. For example, appearing and disappearing fluid/gas-filled spaces, or an organ shift beyond the field of view[17,22].

Also, non-uniqueness is an issue encountered within those areas of the image that have uniform-intensity. The problem is that the voxel belonging to the homogenous region in one image can be mapped to many voxels in the other image, potentially causing a non-physical deformation[17].

Another possible difficulty is non-linear deformation of rigid structures such as bones[23,24,25]. Visual inspection of the registration results and/or adding relevant constraints to registration algorithm can effectively manage the potential negative impact of this problem in the clinical environment.

### 2.2.8 Clinical Applications

Deformable image registration has found multiple applications in radiation therapy including:

i) multi-modality image fusion for better localization and contouring of target structures, as well as monitoring local tumor response to radiation dose[26-29]; ii) atlas and multi-atlas segmentation of organs at risk[30-34]; iii) contour propagation between imaging data acquired at different time points[35-39]; iv) dose accumulation, estimation and monitoring for adaptive treatment re-planning and quality assurance purposes[40-42].

## 2.3 Adaptive Radiation Therapy Strategies

### 2.3.1 Offline Techniques

The offline ART approach is implemented by replanning the treatment during the time between delivered dose fractions (i.e. one or two days), utilizing most of the standard treatment planning processes but executed in a shorter timeframe. One of the benefits of offline ART is that not only can the daily CBCT be used for replanning but also, if higher image quality or functional imaging data is required, conventional CT, MRI or even PET scanners can be used as well. One of the disadvantages of offline ART is that if anatomical variations occur within a fraction or frequently between fractions, then offline ART may not be fully capable of responding to such changes[43-47]. This is because in offline ART, the CBCT acquired in the previous fraction is used to adapt the treatment plan that is intended to be delivered in the next fraction, therefore if any significant changes in the patient anatomy appear after fraction $n$ but before fraction $n$+1, these changes will not be taken into account by the offline approach.

Figure 2.9. shows an offline adaptive radiation therapy workflow that starts from the initial simulation image acquisition. The imaging data are used for initial target and OAR contouring performed by a radiation oncologist on the planning computed tomography scans (pCT). Once the pCT is fully segmented, radiation therapy planners create the treatment plan containing photon beam arrangements and optimized dose distribution. After passing the verification and approval process, the treatment plan is then sent electronically to the delivery console at the linac. Before the radiation treatment can be delivered to the patient, his/her position is verified (patient set-up) by comparing pCT to the daily on-treatment CBCT (or alternatively using orthogonal planar images). Rigid transformations (couch shifts and three-axis rotations) can be implemented to maximize position congruence. If the remaining difference in alignment between these two image sets is within tolerance limits (match in Figure 2.9), the initial treatment plan is delivered. However, if the patient's anatomy significantly deviates from the initial anatomy (i.e. that based on which the treatment plan was prepared), which is a rare occurrence, then the irradiation session would be rescheduled, and/or a new treatment plan created (re-planning) based on the new anatomy and prepared ahead of the next fraction to be delivered. Even if the first fraction dose is delivered without issue, the CBCT data can still be used to adapt the initial treatment plan accounting for changes in the patient anatomy that might have caused only a minor mismatch between pCT and CBCT. As shown in Figure 2.9. the offline plan adaptation takes place after the first and before the second fraction of the treatment. Offline ART involves mapping the pCT to the daily CBCT with deformable image registration and then using the deformed pCT (dCT) to create a new treatment plan.

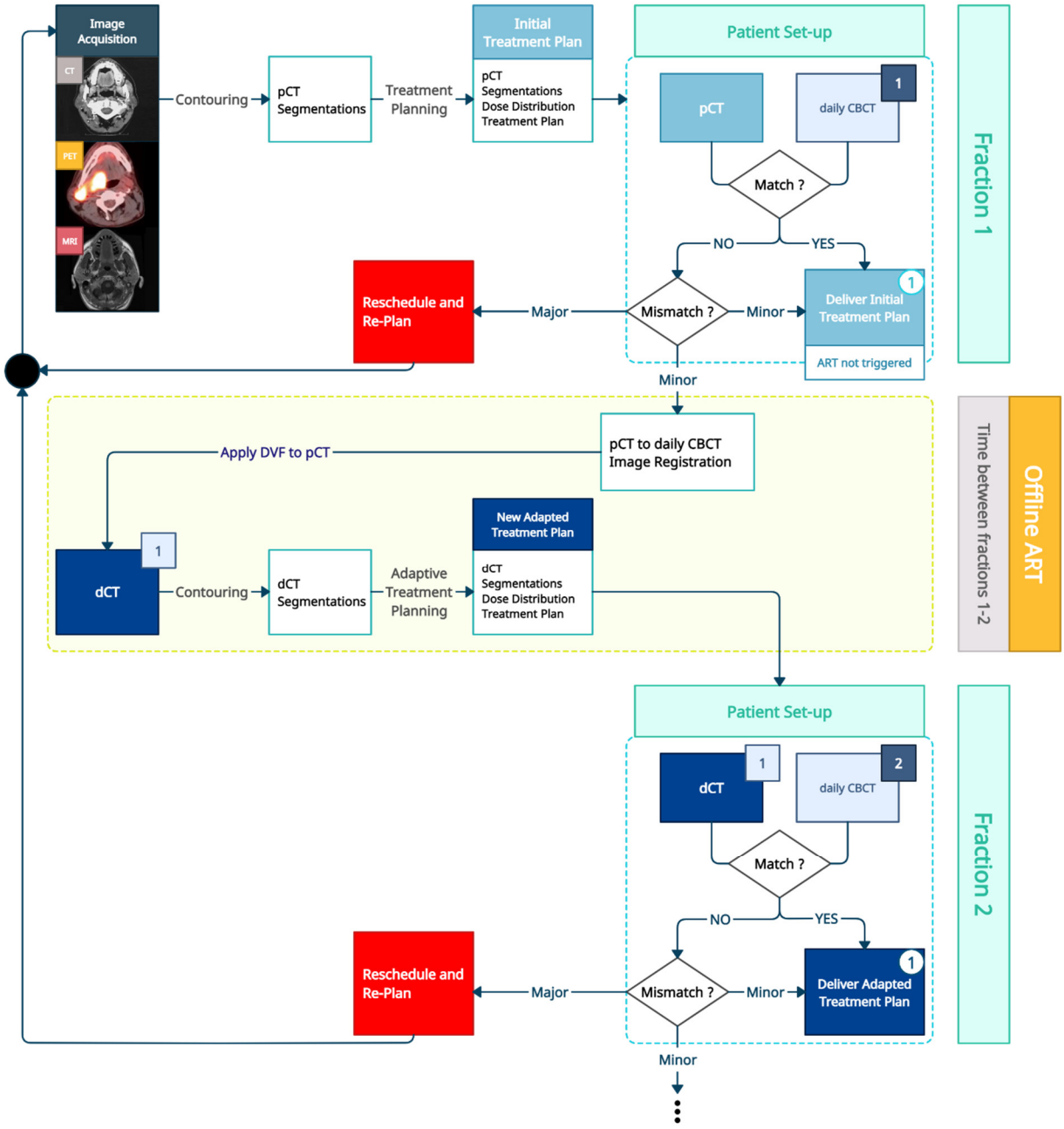



**Figure 2. 9.** A diagram illustrating a workflow of offline adaptive radiation therapy.

Before the next fraction dose can be delivered, the alignment between dCT and the new daily CBCT is examined to ensure correct patient position and alignment of patient's anatomy. Depending on the results of this assessment, further steps are the same as during the first treatment session. The entire offline ART procedure can be repeated until the radiotherapy treatment is completed. However, it is important to note that the offline re-planning does not have to be performed if the patient anatomy maintains consistent size, shape, and positioning over the course of the treatment.

### 2.3.2 Online Techniques

In general, 'online' ART implies that the plan adaptation takes place while the patient is on the treatment couch. The most common form of online ART takes place while the patient is in the treatment position but immediately before the dose fraction is delivered. For example, the patient is setup at the linac, imaged, and then the plan adapted while they wait on the couch, followed by delivery of the new plan. Theoretically, online ART can also take place *during* the fraction delivery, however this full approach would require nearly real-time imaging and replanning and is extremely rare. An increasingly common implementation of on-line ART is termed 'real-time tumour tracking' where the field aperture tracks the tumour location in near real-time as indicated by imaging or other positional tracking mechanisms. In this implementation, the delivered fluence pattern tracks the geometric motion, but no live dosimetric replanning is performed. Because the online ART procedure includes assessing the need for ART, re-planning, and performing quality assurance, the entire process needs to be efficient and requires specialized, well-integrated tools. Despite the complexity and time sensitivity, the major advantage of

online ART is that changes in the patient's anatomy at the time of dose delivery, once discovered, can be accounted for before the continuation of the treatment delivery. However, one of the challenges to both offline and online ART is significant intrafraction change in patient anatomy. As mentioned previously, some of those variations can be accounted for by using real-time motion management methods or real-time tumour tracking. Nevertheless, anatomical changes that occur between the time of in-room imaging and subsequent modified beam delivery can decrease the accuracy of online ART. For example, bladder filling changes or stomach emptying can occur during the time it takes to evaluate, reoptimize, and verify a new online ART treatment plan. Another difficulty in the implementation of online ART solutions in clinical practice is the inability to perform patient-specific quality assurance measurements because the patient remains in the treatment position. For the above reasons, quality assurance procedures need to be adjusted for online ART applications[48-52]. An online adaptive radiation therapy workflow example is shown in Figure 2.10. The description of the major components of this diagram is similar to the offline ART description in the previous section 2.3.1 (for Figure 2.9) except that the replanning is performed immediately on the patient's acquired images for that fraction, and the adapted plan delivered in the same session.

### 2.3.3 Clinical Applications of ART

ART has been applied in the treatment of many types of cancers across a variety of anatomical regions including the prostate, head and neck, brain, bladder, lung, breast, cervix, and pancreas, among which the first two (i.e. prostate, and head and neck) are the most studied in the literature[53,54,55].

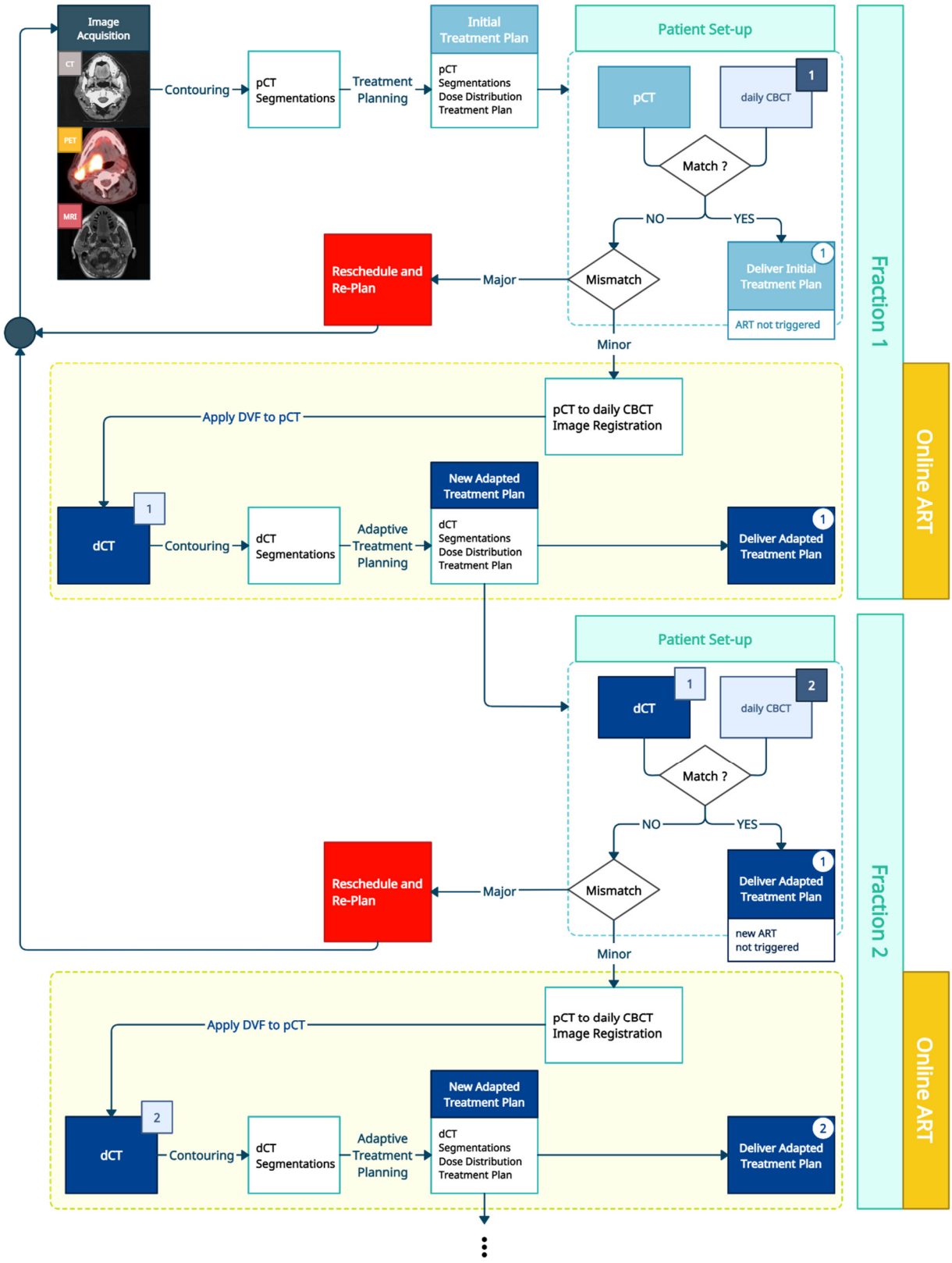



**Figure 2.10.** A diagram illustrating a workflow of online adaptive radiation therapy.

In the radiation therapy of prostate patients, both the position and shape of the prostate target volume can change during treatment, mainly due to the variable filling conditions of the bladder and rectum, organs that lie immediately adjacent to the prostate. The geometric variations can result in undesired and substantial deviations of the delivered dose away from the planned dose[56,57]. For example in a study of 28 prostate cancer patients who underwent radiotherapy treatment, Huang et al.[58] showed that the mean percentage differences (± one standard deviation) in the structure volume and radiation dose were 44% (± 41) and 18% (± 17) for the bladder, 20% (± 21) and 2% (± 2) for the prostate, and 36% (± 29) and 22% (± 15) for the rectum, respectively. Also, Boer et al.[59] investigated the effect of rotations of the prostate gland for five prostate cancer patients who received VMAT treatment and demonstrated that the rotations induced considerable dosimetric uncertainty in prostate irradiation. On average, the clinical target volume minimum dose ($D_{\text{min}}$) decreased by up to 10% in the case of the treatment plans that were delivered without any corrections.

Similarly, head and neck cancer patients are well documented in the literature to significantly benefit from radiation therapy plan adaptations due to substantial anatomy deformations that can occur throughout the course of their treatment. For example, Brouwer et al.[60], in the review of 51 studies on ART for head and neck cancer patients, reported volume changes and associated dose differences (planned versus delivered dose) for organs at risk and target volumes. They found that for parotid glands, on average, there has been a 26 ± 11% decrease in reported volume change which caused a 2.2 ± 2.6

Gy increase in radiation dose relative to the dose calculated on the planning CT (up to 10.4 Gy on average for patients with stage III-IV nasopharyngeal carcinoma). Zhao et al.[61] found the highest dose increase in $D_{1\%}$ (dose to 1% of the organ's volume) of the spinal cord and the brainstem of 5.6 and 2.5 Gy on average, respectively, for the entire treatment course. In a study of 19 patients Cheng et al.[62] reported that if no replanning was performed, the tolerance dose of 54 Gy for the maximum dose to the brainstem was exceeded after the 15$^{th}$ and 25$^{th}$ fraction (out of 33) for 11% and 16% of the patients, respectively. Furthermore, for 11% of the patients, the maximum dose to the spinal cord was higher than the tolerance dose of 45 Gy after the 15$^{th}$ fraction. Thus, based on the comprehensive literature review of Brouwer et al.[60] it is clear that ART has great potential for improving the dosimetry of the radiation therapy treatment of head and neck cancer patients.

Brain tumor treatment such as glioblastoma can also significantly benefit from ART. The main issue with this kind of tumor is the precise identification of cancerous cells. Therefore, MR imaging is often used to observe the radiation response of the tumor, which then can be used to modify radiation treatment. MRI techniques may also identify the areas of the tumor that would respond to drugs compared to areas that would not, and therefore may be irradiated in different ways[63-66]. However it is noted that at the moment it is not usually practical to acquire MRI images of the patient anatomy during the course of radiotherapy treatment although this may change with a broader adoption of recently available commercial MRI-linacs.

For bladder cancer, the possibility of large organ motion and differences in bladder filling make it necessary to apply margins of up to 2 cm, even when drinking protocols and Foley catheters are used in the treatment planning and delivery workflow. This margin is significant as, for example for a CTV of 130 $cm^3$, a 2-cm margin results in a PTV of 600 $cm^3$. Adaptive planning to conform to the daily bladder-filling variation is valuable in reducing these relatively large PTV margins[67].

Breast conservation surgery causes the development of a seroma at the surgical site that may change its shape significantly after the surgery and over the course of radiation treatment, as it is a region filled with fluid located within a deformable organ[68]. Additionally, daily anatomical changes of up to 2 cm may appear in the location and volume of a lumpectomy surgical site[69], justifying the application of ART for treatment techniques that target the seroma (ie. partial breast irradiation) instead of the whole breast.

Significant tumor shrinkage and OAR motion around the cervix make this site a good candidate for the application of ART. In the first 3 to 4 weeks of radiation treatment, tumor regression of up to 60% to 80% of the original GTV has been reported[70].

Large magnitudes in the shift of the pancreas location have been used to justify the application of ART for that treatment site as well. There can be as much as 2 cm of variation of the pancreas and local anatomy from day to day[71]. The pancreas is also closely

located next to many other abdominal OARs, and in cases with microscopic disease invasion, the impact on those OARs becomes more acute.

In summary, ART approaches are of significant interest in nearly all cancer sites and have been demonstrated to be valuable for many, as outlined above.

## 2.4 Machine Learning

### 2.4.1 Introduction

The definition of machine learning (ML) varies widely and depends on how this broad discipline is perceived by researchers with different and often combined scientific, engineering, art, and business backgrounds. However, there are a few definitions that were formulated by early adopters of machine learning methods and are still frequently quoted in the available literature. In 1959, Arthur Samuel described machine learning as "the field of study that gives computers the ability to learn without being explicitly programmed", also concluding that *"programming computers to learn from experience should eventually eliminate the need for much of this detailed programming effort"* [72]. Compared to Samuel's computer science-oriented view, Tom Mitchell, in 1997, described ML in his book more from the engineering perspective as: *"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E"* [73].

Both of these, and many other definitions[74-77] share a common characteristic of ML as a collection of algorithms that allow a computer program to first analyze and find meaningful relationships in historical data (i.e. to learn from past experiences) and then

use that knowledge to build a mathematical model that would be able to complete a specific task. Ultimately, a program should also be able to automatically optimize an ML model based on new incoming data in order to keep improving its performance (i.e. continuously improve and effectively learn from new experiences).

The main advantage of machine learning is that it makes few to no assumptions about the underlaying data allowing the algorithms to automatically learn the relationships between variables describing the data. The rapid growth of the ML discipline has resulted in the development of solutions to a wide variety of practical problems found in nearly every sector and industry of modern civilization. Despite the great benefits that ML brings to the world of science and technology, it still has several limitations. These mainly include insufficient quantity and quality of the data itself, non-representative data, time-consuming data processing, feature engineering, and model overfitting (or underfitting) of the training data[77].

The subsequent sections will describe the general types of common machine learning models (section 2.4.2. and 2.4.3.), the basic concepts of machine learning model development (section 2.4.4.), and several commonly used machine learning algorithms (sections 2.4.5. – 2.4.10.)

### 2.4.2 Supervised Learning

Based on the available data and the level of supervision required for model training, machine learning is in general divided into two categories: supervised learning and unsupervised learning. In practice, models belonging to those categories are often combined in order to solve complex problems or achieve better performance. This section

will briefly describe supervised learning while section 2.4.3 will outline unsupervised learning.

Supervised learning is a type of ML in which the training data contains response (target) variables that label individual data samples or instances. The objective of supervised learning is to create the model based on the available data samples so that the model is able to predict a response variable given a new, unlabeled data instance (observation). Depending on the type of the response variable itself, the supervised learning is further subdivided into classification and regression problems[76,78-80]. The most common supervised learning algorithms include k-Nearest Neighbors, Linear Regression, Elastic Net, Logistic Regression, Support Vector Machines (SVMs), Decision Trees, and Random Forests[77]. More detailed descriptions of these algorithms are found in sections 2.4.5-2.4.9.

In classification problems, the response variable is categorical (discrete) and classifies each data sample into a particular class. For example, in the simplest binary classification problem, the data could contain the medical records of a large number of cancer patients treated with radiation therapy, chemotherapy, surgery, or some combination of the three modalities. The response variable could categorize those patients into two groups with a survival rate above and below a certain number of years. In this problem, the purpose of a ML binary classifier would be to create an accurate mapping between input data (medical records) and the target variable so that when a new unclassified observation is fed into the model, the model would classify this observation into one of two classes, usually by assigning a probability score to each class. An ML problem containing more than two classes is called a multi-classification problem.

In regression problems the response variable is continuous and the objective of the ML regressor is to predict the value of the output variable. Similar to classification problems, the prediction is made based on the available observations that are characterized by several independent or predictor variables called features. An example of a regression problem would be the prediction of radiation response of the tumor and/or healthy tissues, or the number of new radiation therapy patients in a given period of time[74].

### 2.4.3 Unsupervised Learning

Unsupervised learning is a type of ML in which the training data does not contain target variables that would label specific observations (i.e. the model input data are unlabeled). This is because the objective of unsupervised learning is to identify underlying patterns in the data in a way that would allow the ML model to group or reduce the dimensionality of the data. Therefore, unsupervised learning is further subcategorized into clustering and dimensionality reduction problems. The most common algorithms applied for clustering problems include K-Means and DBSCAN (Density-Based Spatial Clustering of Applications with Noise), while for dimensionality reduction tasks, PCA (Principal Component Analysis) and Kernel PCA[77] are frequently used.

The clustering methods focus on dividing the data into groups so that the observation belonging to a particular group is more similar to other observations from that group than to the observation from another group (the similarity measure is selected within a specific algorithm). An example of clustering would be to group acceptable and unacceptable radiation therapy treatment plans by applying a clustering algorithm to the set of

treatment plan features[73]. Another common task is anomaly detection[77], for instance the detection of sudden and abnormally large increases in a certain disease incidence.

The dimensionality reduction techniques aim to transform an original structure of the data into a lower-dimensional representation while preserving the properties of the primary data structure. These ML methods are often used to visualize complex, high-dimensional datasets and to pre-process data features for supervised learning problems[76].

For this dissertation, the most relevant type of machine learning problem is supervised learning as it was utilized in the research project described in Chapter 6. Therefore, the next section will highlight the main steps involved in the training of supervised learning algorithms.

### 2.4.4 Model Development

Although the detailed process of training a supervised machine learning model can vary widely depending on the specific problem, available data, selected algorithm, and model validation method, there are a few important steps that are generally followed when creating an ML model.

The process starts with gathering labeled data and splitting them into training, validation, and test sets. The training data are used by the ML algorithm to learn patterns in the data and map the model inputs to the model outputs. The validation samples are utilized for tuning hyperparameters of the model, while the test portion of the data is used to assess the model performance.

The next step after gathering the data is called 'feature engineering' which is the selection and postprocessing of learning features. This part of ML model development is very important because identifying the predictive features, as well as removal of faulty data samples, can facilitate the learning process and potentially result in improved performance of the model. Also, feature engineering is the stage where the application of the researcher's expertise is the most impactful.

Once the data are gathered and feature engineering is completed the process of training the model can begin. During the model training, its free parameters are adjusted to create a specific mapping function. Additionally, for a different set of hyperparameters, the parameter search process would generate the number of model instances (models with different sets of hyperparameters) each with a different performance on the validation set. The best performing model instance is then selected and evaluated on the testing data. After obtaining satisfactory results in the ML training, the model together with the data processing pipeline can be deployed in production, monitored and, if necessary, re-trained or otherwise improved[76].

The next sections will briefly introduce the basic concepts of the most common supervised learning models and specifically those that were used in the work described in Chapter 6 including Linear Regression, Elastic Net, Logistic Regression, Support Vector Machines, Decision Trees, and Random Forest.

### 2.4.5 Linear Regression

Linear regression is a very simple machine learning algorithm that maps input features to the response variable using a linear function:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \cdots + \ \beta_m x_m$$

(Eq.2-1)

where:

$\hat{y}$ – response variable

$x_m$ – the $m$-th independent variable (feature); $x_0 = 1$

$\beta_m$ – model parameter of the $m$-th variable ($\beta_0$ – is a model bias) or $m$-th regression coefficient

$m$ – number of features

Equation 2-1 can also be simplified to vector form:

$$\hat{y} = \beta^T \boldsymbol{X}$$

(Eq.2-2)

in which $T$ denotes a row vector of regression coefficients (the transpose of vector $\beta$).

The training of the linear model is conducted through the calculation of weights $\beta_m$ associated with the input features $x_m$ in a way that minimizes the mean square error cost function:

$$C_{Linear}(\beta) = \frac{1}{n} \sum_{i=1}^{n} (\beta^T \boldsymbol{X}_i - \ y_i)^{\ 2}$$

(Eq.2-3)

where n is the number of individual samples in the data. To prevent a machine learning model from overfitting, the reduction of model complexity is often considered. In the case

of linear regression, it can be achieved by eliminating a number of model weights (i.e. setting them to zero). The disadvantage of this approach however is a loss of information about the potential impact the removed variables could have had on the model performance. Therefore, in practice, instead of removing the regression coefficients associated with the model features, regularization terms are added to the objective function $C(\beta)$. The purpose of the regularization terms is to penalize regression coefficients by limiting their values. The commonly used regularization methods for the linear regression model are ridge and lasso regularization, described below.

The objective function in Eq.2-3 with added ridge regularization can be expressed as:

$$R(\beta)_{Ridge} = C(\beta) + \underbrace{\alpha \frac{1}{2} \sum_{j=1}^{m} {\beta_j}^2}_{\substack{\text{Ridge} \\ \text{Regularization}}} \tag{Eq.2-4}$$

The parameter $\alpha$ reflects the level of model regularization so that for $\alpha = 0$, Eq.2-4 becomes an ordinary linear regression (Eq.2-1), while for $\alpha >> 0$, the regression coefficients approach zero. M is a number of

The equation for a cost function that includes lasso (Least Absolute Shrinkage Selection Operator) regularization is:

$$R(\beta)_{Lasso} = C(\beta) + \underbrace{\alpha \sum_{j=1}^{m} |\beta_j|}_{\substack{\text{Lasso} \\ \text{Regularization}}} \tag{Eq.2-5}$$

It can be noticed that the lasso regularization is expressed as a sum of absolute values of regression coefficients and is referred to as $L_1$ penalty or $L_1$ norm, while ridge regularization penalized the sum of squared regression coefficients and is referred to as $L_2$ penalty or $L_2$ norm. Due to that difference, large values of alpha in the lasso term will result in many $\beta$ coefficients being set equal to zero. In contrast, the ridge regularization will not zero regression coefficients even in the case of large alpha. This important characteristic of lasso regression allows it to eliminate the least important features of the model and, during ML model training, use only those with relevant predictive power[77,79,81,82].

### 2.4.6 Elastic Net

Elastic Net is a model that combines a linear regression with both ridge and lasso regularization and can be described in the following equation:

$$C_{Elastic}(\beta) = \frac{1}{n}\sum_{i=1}^{n}(\beta^T \boldsymbol{X}_i - y_i)^2 + \frac{1-r}{2}\alpha\sum_{j=1}^{m}{\beta_j}^2 + r\alpha\sum_{j=1}^{m}|\beta_j| \quad \text{(Eq.2-6)}$$

where parameter $r$ controls the balance of ridge and lasso contributions to the regularization of a linear regression model so that for $r$=0 and $r$=1, Elastic Net becomes equivalent to Eq. 2-4 and Eq. 2-5, respectively[77,82].

### 2.4.7 Logistic Regression

Logistic regression is one of the most commonly used models for a binary classification problem. Compared to linear regression which most frequently predicts the value of the

continuous variable, the logistic regression model classifies the data by assigning a probability to each class.

The logistic regression model can be described in the form of the logarithm of the odds (i.e. logit transformation):

$$\text{Log}\left(\frac{p}{1-p}\right) = t = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \cdots \ + \ \beta_m x_m \tag{Eq.2-7}$$

where $p$ is a probabilistic logistic function equal to:

$$p(t) = \frac{1}{1+e^{-t}} \tag{Eq.2-8}$$

and $t$ denotes the right side of Eq.2-7.

The logistic regression model can be considered a linear regression model that has been transformed with the sigmoid function. When it comes to machine learning applications, the model predictions are often expressed using probability $p$ due to the convenience of its range lying between 0 and 1, instead of using odds $p/(1-p)$ with a range lying between 0 and $+\infty$.

Assuming that the value of probability $p$ is calculated, the model prediction $\hat{y}$ for binary classification tasks can be represented as:

$$\hat{y} = \begin{cases} 0 \ if \ p < 0.5 \\ 1 \ if \ p \ \geq 0.5 \end{cases} \tag{Eq.2-9}$$

where $\hat{y} = 0$ and $\hat{y} = 1$ indicate two distinct classes of data samples.

In order to train a logistic regression model, the following logarithmic cost function averaged over all the training samples is used:

$$C_{Logistic}(\beta) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(p_i) + \ (1-y_i)\log\ (1-p_i)] \tag{Eq.2-10}$$

This form of the cost function is particularly useful for the logistic regression model, due to its intuitive interpretation. That is, $C_{Logistic}(\beta)$ will be approximately equal to 0 (or 1) in the case the probability $p$ is close to 0 (or 1), classifying a sample into one of two classes for which $\hat{y} = 0$ or $\hat{y} = 1$[77,79,81,83].

### 2.4.8 Support Vector Machines

Support Vector Machines (SVM) is a flexible and powerful model, very often used in machine learning applications. It is also mathematically more complex compared to models described in the three previous sections. Nevertheless, this section will introduce the main concepts of SVM for a binary classification task.

Assuming that the training data are partitioned into two classes, the purpose of SVM is to find an oriented hyperplane so that the data on each side of that plane belong to one distinctive class and are labeled as $y_i = +1$ or $y_i = -1$ ($i$ denotes individual data sample). Most importantly, the SVM algorithm selects the hyperplane that maximizes the distance between itself and the data points located on each side of the hyperplane. The selection of the hyperplane is mainly affected by the data points lying closest to hyperplane, called support vectors.

If the data points in the binary classification problem are denoted as $x_i (i = 1, \ldots, n)$, their weights as $w_i (i = 1, \ldots, n)$, and their corresponding labels as $y_i = \pm 1$ then the basic function describing the model is:

$$f(\vec{\boldsymbol{x}}) = sign\ (\vec{\boldsymbol{w}} \cdot \vec{\boldsymbol{x}} + b\ ) \tag{Eq.2-11}$$

so that the closest points on one side of the hyperplane will be defined by $\vec{\boldsymbol{w}} \cdot \vec{\boldsymbol{x}} + b = 1$ while the points on the other side by $\vec{\boldsymbol{w}} \cdot \vec{\boldsymbol{x}} + b = -1$. Two hyperplanes passing through the closest points are called canonical hyperplanes. The SVM hyperplane itself would be described as $\vec{\boldsymbol{w}} \cdot \vec{\boldsymbol{x}} + b = 0$. Although it is beyond the scope of this introduction, it can be shown that the margins between the two canonical hyperplanes and the SVM hyperplane are equal to $1/\|\vec{\boldsymbol{w}}\|$.

Considering the above, the objective of the SVM is to minimize this margin by minimizing:

$$\frac{1}{2}\|\vec{\boldsymbol{w}}\|^{\mathbf{2}} \tag{Eq.2-12}$$

subject to the constraints:

$$y_i(\vec{\boldsymbol{w}} \cdot \vec{\boldsymbol{x}}_i + b) \geq 1 \quad \forall i \tag{Eq.2-13}$$

Another important consideration for support vector machines is the fact that for many practical problems the training data are not linearly separable in the input space (for example circular distribution of data or data samples sharing a similar range of features). Therefore, the SVM algorithm is frequently used with kernels (for example Polynomial,

Gaussian, or Sigmoid kernels) that map the data points from the input space to an alternative space of different dimensionality where the hyperplane separating the data may exist[77,79,81,84-88].

### 2.4.9 Decision Trees and Random Forest

Currently, decision trees are relatively complex and flexible machine learning algorithms that can be used for both classification and regression problems. Decision tree predictions are based on a series of decision-making rules that, after model visualization, can be easily interpreted.

The algorithm forms a tree-like structure from the training data by first identifying one of the data features (the root of the tree) and using it to split the data into subsets. Then each subset is further split based on another (or the same) feature. The process is repeated until the tree structure reaches its leaf (terminal) nodes. The longest distance between the root of the tree and its leaf is called the depth of the tree. Its maximal value is one of the decision tree hyperparameters.

Splitting the tree node into subsequent nodes is an important aspect of growing the decision tree structure and is commonly based on the calculation of the Gini Index:

$$G_i = 1 - \sum_{k=0}^{n} p_{i,k}^2 \qquad \text{(Eq.2-14)}$$

where $i$ is a node of the tree, $n$ in a number of classes, $k$ is a particular class and $p_{ik}$ is the probability that for a node $i$ a given sample belongs to a class $k$ (in the classification problem).

Specifically, the feature and its rule-based threshold used for node split are selected in a way that minimizes the value of the Gini Index.

Random forest (RF) is a model built upon many decorrelated decision trees, where the predictions are made by the majority voting rule. In particular, the RF algorithm starts by drawing a random sample of size $N$ from the training data. Then, for that sample, RF builds a decision tree using randomly selected $m$ features ($m<M$; $M$ - total number of features available). The process repeats until a predetermined maximum number of decision trees is created. For a binary classification task, a new data sample is classified to a particular class if the prediction for that class comes from the majority of decision trees. The purpose of introducing a random forest classifier that uses multiple high-variance decision trees is to reduce the total model variance with respect to the testing data, thus preventing the model from overfitting[75,77,81,85,88,89].

## 2.5 Deep Learning

### 2.5.1 Introduction

Similar to machine learning, deep learning (DL) focuses on estimating a function that maps the input data characterized by a number of features into the output/response variable(s). The fundamental difference between ML and DL is that in ML the features describing the data are selected and engineered by a human while in DL a design of neural networks allows them to learn relevant features from the data without human involvement. This important property of deep learning makes it suitable for solving very challenging problems with a significantly smaller inclusion of human intelligence. For example, image recognition and segmentation in the field of computer vision, or speech

understanding and text analytics in the area of natural language processing. In regards to terminology, deep learning is considered a subfield of machine learning, while both of those disciplines belong to a broad domain of artificial intelligence.

The following sections will briefly explain the main components and model training principles for artificial neural networks (ANN; section 2.5.2) and convolutional neural networks (CNN; section 2.5.3), as well as the building structure and the application benefits of inception and residual modules in the performance of CNNs (sections 2.5.4-2.5.7).

### 2.5.2 Artificial Neural Networks

The fundamental learning unit of the human brain is a neuron. Its objective is to receive information in the form of electrical impulses originating from other neurons. Then the information is processed within the neuron, and the outputs are transmitted to other neurons. Figure 2.11a illustrates this process in a simplified fashion. A single neuron collects several impulses with its dendrites. Based on how frequently each dendrite is used, it is strengthened or weakened. Therefore, the strength of each dendrite connection impacts the contributions of incoming signals to the neuron's output. All the neuron's inputs are summed in the cell body, propagated along the axon, and transmitted to the other neurons through the connections at the synaptic terminals.

In ANNs, the equivalent of a biological neuron is an artificial neuron (Figure 2.11b) that accepts the inputs $x_1, x_2, x_3, \cdots, x_n$ through weighted connections $w_1, w_2, w_3, \cdots, w_n$. Specifically, each input $x_n$ is weighted by $w_n$, summed together with the associated bias

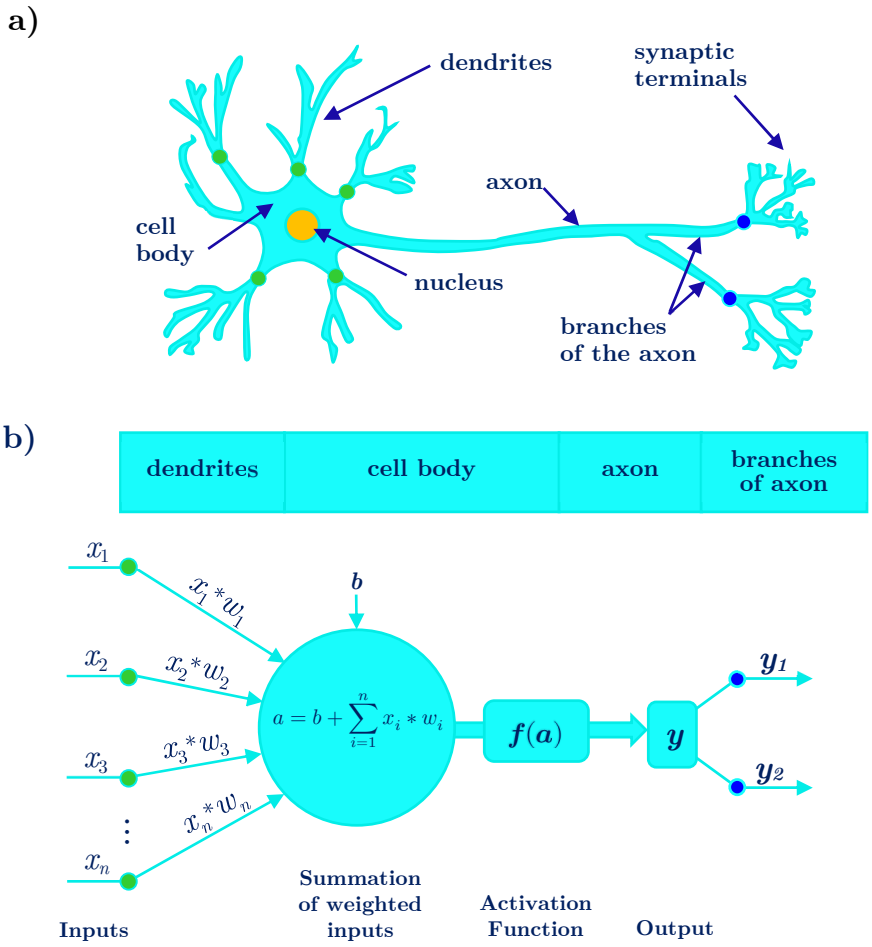



**Figure 2.11.** The structure of the biological neuron **a)**, and the artificial neuron **b)**. Green and blue dots symbolize input and output signals, respectively.

$b$, and then transformed by the activation function $f$. This formulation allows the artificial neuron to be modeled by the following function:

$$y = f(b + \sum_{i=1}^{n} x_i * w_i) \qquad \text{(Eq. 2-15)}$$

where $y$ is the neuron response.

Equation 2-15 shows that the neuron's output is determined by the activation function. Commonly used activation functions are Heaviside (step function), ReLu (rectified linear unit), Sigmoid, and Hyperbolic Tangent as shown in Figure 2.12.

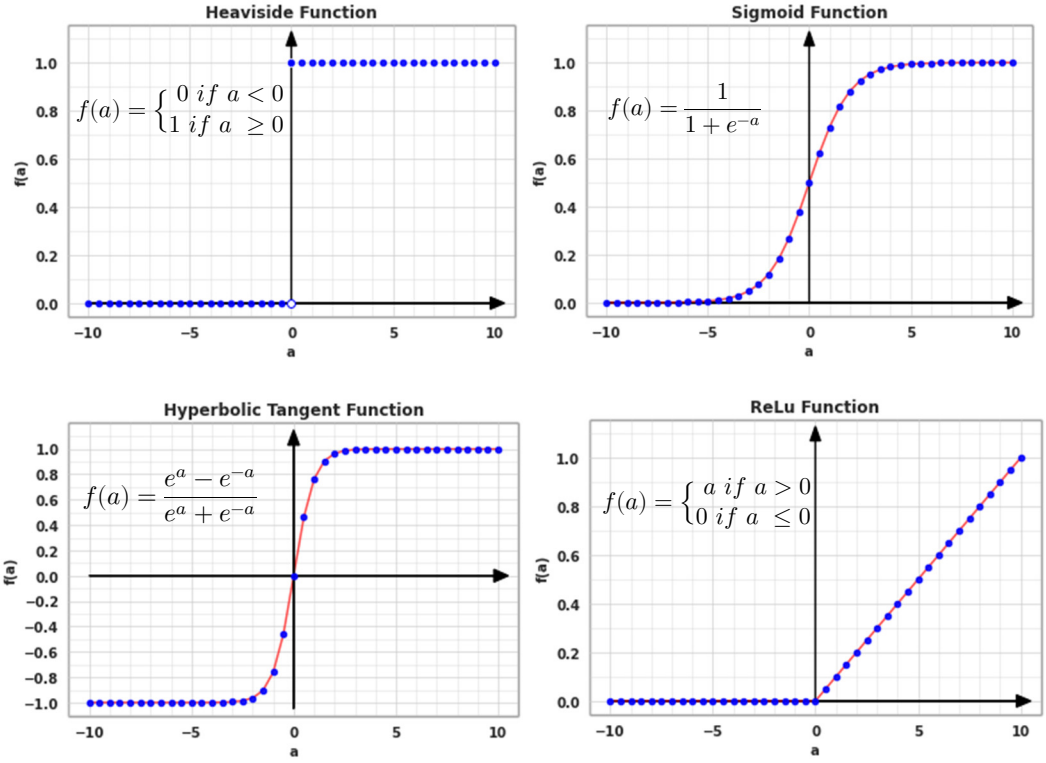



**Figure 2.12.** Activation functions commonly used in artificial neurons. For simplicity, the plots assume the bias $b = 0$.

Although a single artificial neuron can perform a basic input-to-output mapping, modeling of complex relationships requires an application of multiple, interconnected neurons that form an artificial neural network.

A typical ANN consists of an input layer, hidden layers, and an output layer. The input layer receives available information. The hidden layers are built by multiple artificial neurons (also called nodes of the network) that extract patterns from the input data. The output layer is composed of neurons that generate the ANN outputs based on the computations performed in the previous layers.

The simplest ANN architecture is called a single-layer feedforward network (Figure 2.13). It consists of one input layer, represented by an input vector, and one neural layer that is also an output layer represented by an output vector. Therefore, the number of neurons in this simple network is equal to the number of ANN outputs. Each layer (here only one) has a weight matrix, a bias vector, as well as the output vector constructed from the weights, biases, and outputs of individual neurons. Another characteristic of this network is that the flow of information is unidirectional forward from the input layer to the output layer. According to the convention illustrated in Figure 2.13, a single node $y_k$ of the output layer can be modeled by the following function:

$$y_k = f\left(b_k + \sum_{i=1}^{n} x_i * w_{ik}\right)$$

(Eq. 2-16)

The same equation can also be written for all $K$ nodes in the output layer. A slight improvement in flexibility over the single-layer feedforward network is a shallow network, as illustrated in Figure 2.14. It is also the simplest type of ANN architecture containing the three separate major components that are also present in the larger, more complex networks i.e., input layer, one hidden layer (in large ANNs there is more than one hidden

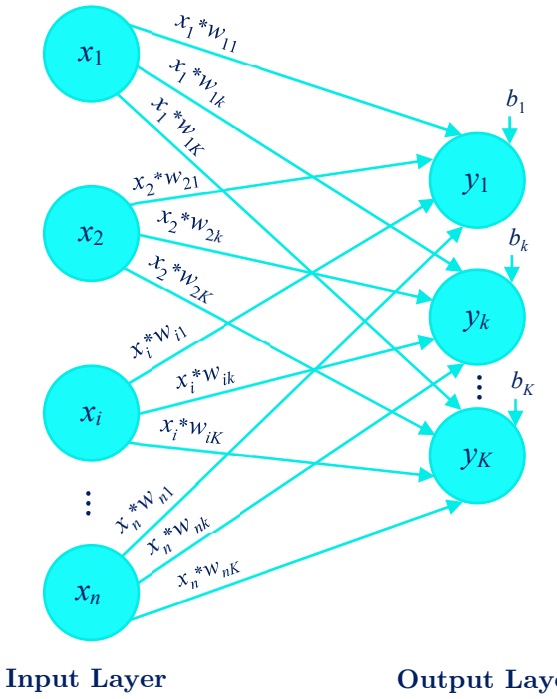



**Figure 2. 13.** Single-layer feedforward network.

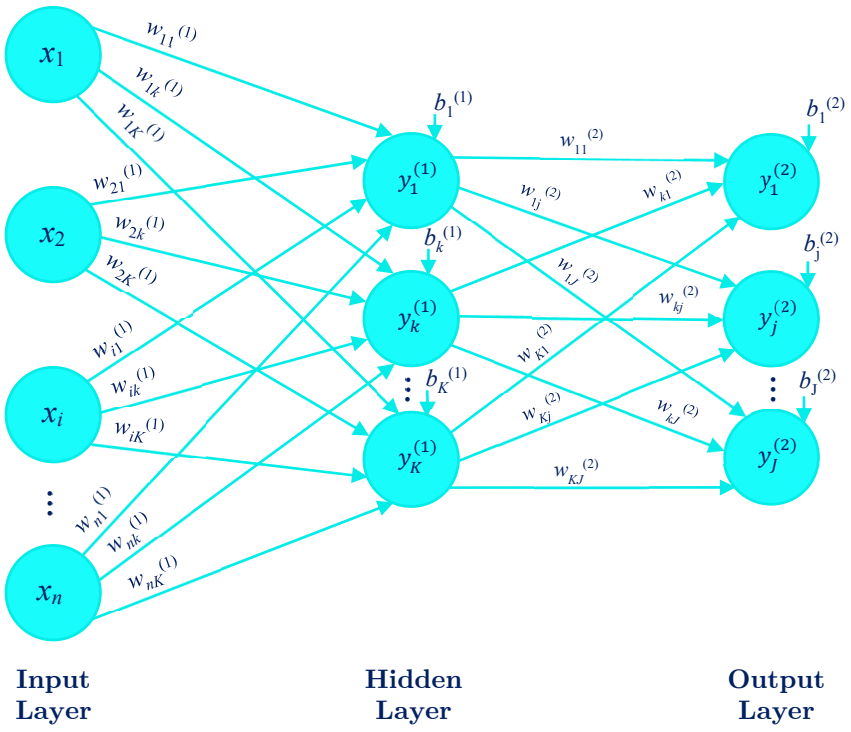



**Figure 2. 14.** Shallow neural network.

layer), and output layer. According to the convention from Figure 2.14, a single node $y_j$ of an output layer can be modeled by the function described in Eq. 2-17.

$$y_j^{(2)} = f\left(b_j^{(2)} + \sum_{k=1}^{K} f\underbrace{\left(b_k^{(1)} + \sum_{i=1}^{n} x_i * w_{ik}^{(1)}\right)}_{y_k^{(1)}} * w_{kj}^{(2)}\right) \tag{Eq.2-17}$$

where *i, k and j* are *i*-th, *k*-th and *j*-th nodes in the output layer; superscripts (1) and (2) denote hidden and output layers while $n$, $K$ and $J$ describe the total number of nodes in the input, hidden and output layers respectively.

Although both ANNs mentioned above constitute an advancement relative to the single neuron, most realistic science and engineering problems require the architecture and the topology of the network to be more complex. A high level of ANN flexibility is achieved by increasing the number of hidden layers, as well as the number of neurons in each hidden layer. In terms of terminology, if the ANN has more than one hidden layer, it is called a deep neural network (DNN). The mathematical description of a network with $L$ number of layers (hidden layers plus output layer) containing $N$ nodes and $L-1$ layers of $m$ nodes can be expressed as a formula for a value of a single $N$-th node in the layer $L$:

$$y_N^{(L)} = \left\{ f \sum\nolimits_m w_{Nm}^{(L)} * \left[ \cdots \left[ f\left( \sum_{k=1}^{K} w_{kj}^{(2)} * \left[ f\left( \sum_{i=1}^{n} w_{ik}^{(1)} * x_i + b_k^{(1)} \right) \right] + b_j^{(2)} \right) \right] \cdots \right]_m + b_N^{(L)} \right\} \tag{Eq.2-18}$$

It can be noticed that this equation is a generalization of equations 2-16 and 2-17.

The training of deep neural networks is performed through a backpropagation algorithm that iteratively adjusts biases and weights associated with each artificial neuron and its connections, respectively. The objective of training a DNN is to minimize the error between network outputs and true labels[77,85,88-92], true labels being the known 'correct' outputs for a given set of inputs.

### 2.5.3 Convolutional Neural Networks

Convolutional neural networks (CNNs) were designed to process 2D grid-structure inputs that exhibit strong spatial dependencies in all their regions. The common example of this type of data are two-dimensional images. CNNs are very effective in performing challenging tasks of image classification, object detection, and image segmentation for both 2D and 3D imaging data sets (where the 3D imaging set can be thought of as a stack of 2D images). This section will briefly introduce the main building blocks of modern 2D convolutional neural networks.

A typical CNN architecture consists of four main segments: an input layer, convolutional and pooling layers, fully connected layers, and an output layer as illustrated in Figure 2.15. The input layer is simply a 2D image. A 1-channel, gray-scale image (ie. one intensity per pixel) will be assumed in this discussion; where a channel is a matrix of pixels that build an image. Compared to gray-scale image, a color image has 3 channels – i.e. red, green, and blue intensities for each pixel. In terms of dimensions, a 50x50 pixel gray scale image has a dimension of 50x50x1 (one channel) whereas a similar sized color image would have a dimension of 50x50x3 (three channels). In the first convolutional layer, the input image is subjected to a convolution operation with multiple filters. The number of filters

determines the depth of a single convolutional layer. Because each filter is applied to the image, the output of the convolution is a series of feature maps. Then each feature map passes through the ReLu function (or other activation function) and is subsequently downsampled by the pooling layer (most commonly a maximum or average pooling) in order to capture high level features and to reduce the number of trainable parameters which improves training efficiency. Down sampling continues until a first fully connected layer can be formed. The convolutional neural network is usually composed of more than one convolution layer, therefore the steps of convolution, ReLu activation, and pooling operations are repeated in the following CNN layers.

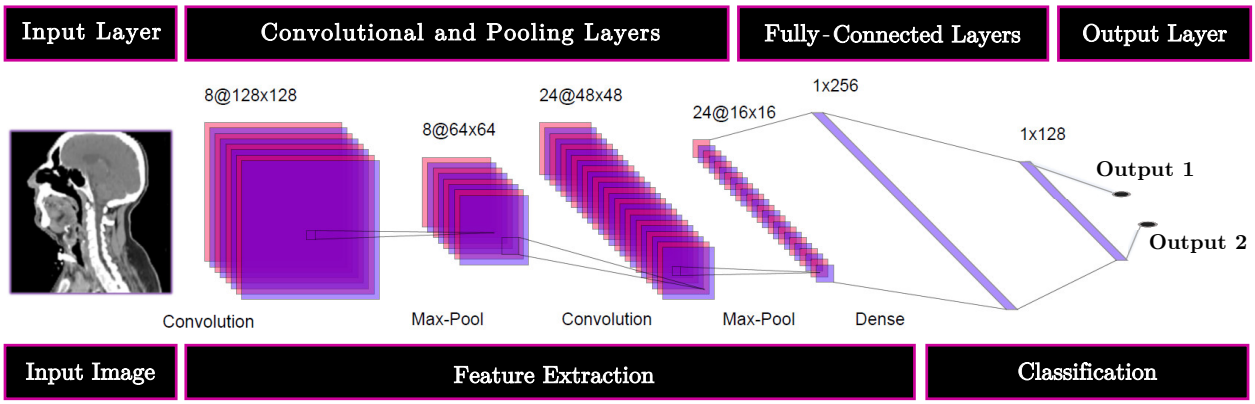



**Figure 2. 15.** A simplified diagram of a convolutional neural network for an image classification task. The notation '8@128x128' means that after the convolution is completed there are 8 feature maps of 128x128 pixel size.

After the input image passes through all the convolutional layers, the resulting feature maps are converted (by pooling operation) into a few fully-connected layers of artificial neurons that lead to the output layer. The output layer delivers predictions in the same way as the ANN approach described in the previous section. The CNN is trained by

iteratively adjusting the weights of the convolutional filters and the weights in the fully-connected layers with the objective to minimize the prediction error[90-92].

### 2.5.4 Inception Module

Inception modules were first introduced in the GoogleLeNet architecture and contributed to the achievement of state-of-the-art results of that network in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2014. The main attribute of the inception modules is their ability to create very deep convolutional neural networks while using significantly fewer learning parameters. For example, GoogleLeNet with approximately 6 million parameters achieved better performance during the 2014 ILSVRC than another large network called AlexNet with 60 million parameters.

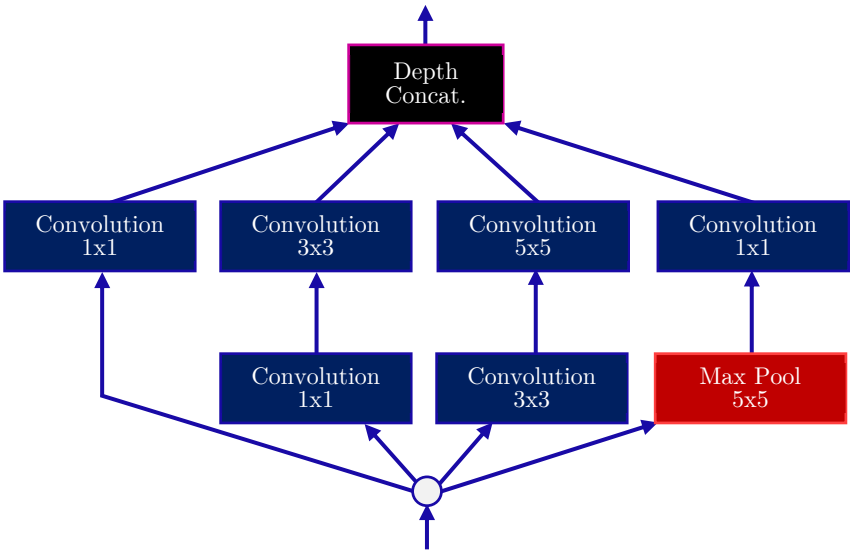



**Figure 2. 3.** Inception module of GoogLeNet architecture[93].

As shown in Figure 2.16. the inception module consists of two convolution layers. The most impactful is the second layer containing the convolutional filters of various sizes allowing the network to find patterns at different dimension levels in all the subregions of the image. Additionally, the same size of four feature maps generated by the top convolutions (1x1, 3x3, 5x5, and 1x1) makes it possible to concatenate these maps in the depth concatenation layer. The purpose of the first two 1 x 1 convolutions is to identify the patterns along the depth dimensions (for 3-channel images) and reduce the dimensionality of the feature maps. Conceptually, the inception module is a more sophisticated version of a single convolution layer and was proven to discover more complex patterns in the imaging data leading to more efficient training and better overall performance of the CNN[77,91,93,94].

### 2.5.5 Residual Module

Similar to inception modules, residual blocks were first applied to convolutional networks during the 2015 ILSVRC challenge and surpassed inception networks in performance. A CNN containing residual (or 'skip') connections is called a Residual Network or ResNet for short. The idea behind developing ResNet was to continue the trend of building an even deeper neural network with fewer and fewer parameters compared to previously introduced CNNs.

The most important characteristic of skip connections, as shown in Figure 2.17 is that the signal entering the initial convolutional layer is simultaneously fed into the output of the layer located deeper in the network. This facilitates the signal propagation throughout the network allowing for more time-efficient training and thus building deeper CNNs (for

a given training time) such as ResNet which is essentially a stack of multiple residual units. Despite the application of skip connection, the hidden layers of residual module are used for learning a residual function $f$(x) as shown in Figure 2.17 which is more efficient than learning the output function $h$(x) directly[77,88,91,94,95]. Moreover, when the training of the non-residual network is initialized, its weights are close to zero, therefore the output of the network will be close to zero as well. However in the case of the residual network, the initialization stage of the training will result in the outputs being equal to the inputs, therefore the network will first model the identity function only. Frequently, the target function h(x) is close to the identity function, which was shown to improve the training speed significantly[77].

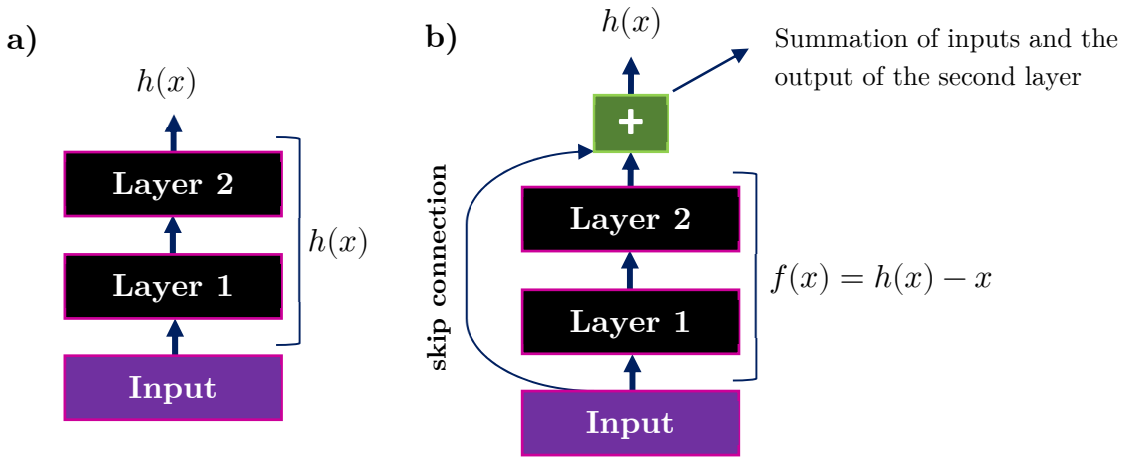



**Figure 2. 4. a)** sequential layer connection; **b)** residual unit with skip connection[77].

### 2.5.6 Inception-ResNet-v2 Architecture

The Inception-ResNet units combine the flexibility of the inception modules in discovering complex image patterns with the residual blocks containing a skip connection that allow

for fast signal propagation throughout the network. The goal is to create a high-performing and time-efficient CNN architecture. The Inception-Resnet-v2 (version two; compared to version one it has a larger number of layers and convolutional filters) network briefly outlined in this section was used for the feature extraction of the segmentation network utilized within the research project described in Chapter 5 of this thesis.

The example of a single Inception-ResNet module is illustrated in Figure 2.18a while Figure 2.18b shows a full Inception-Resnet-v2 network containing 20 such units. The units differ between each other in the number and sizes (dimensions) of convolutional filters. The network also contains stem, reduction, average pooling, dropout, and softmax output layers. A stem layer is an initial layer of the network that is responsible for identifying local features that are then used by deeper layers to learn global features. Reduction layers reduce the dimensionality of feature maps using only inception modules. The average pooling layer also decreases the dimensions of the feature maps, but in contrast to the reduction layer, a pooling layer calculates the average value of a predetermined number of nearest-neighbor pixel intensities. Also, the size of the convolutional filters was selected in the way that would optimize the training speed and capture meaningful image features. The authors of the original paper[94], where this network was proposed (and optimized for image classification task), mentioned that the memory consumption of layers with larger activation size consumed a disproportionately high amount of GPU memory. The dropout layer reduces model overfitting by passing only 80% (i.e. a 0.8 dropout rate) of randomly selected inputs to the next layer (the remaining 20% is ignored and associated weights are equal to zero). Softmax layer is one-dimensional array of fully connected artificial neurons with softmax activation functions, that generate the outputs in the form of continuous probability values[94].

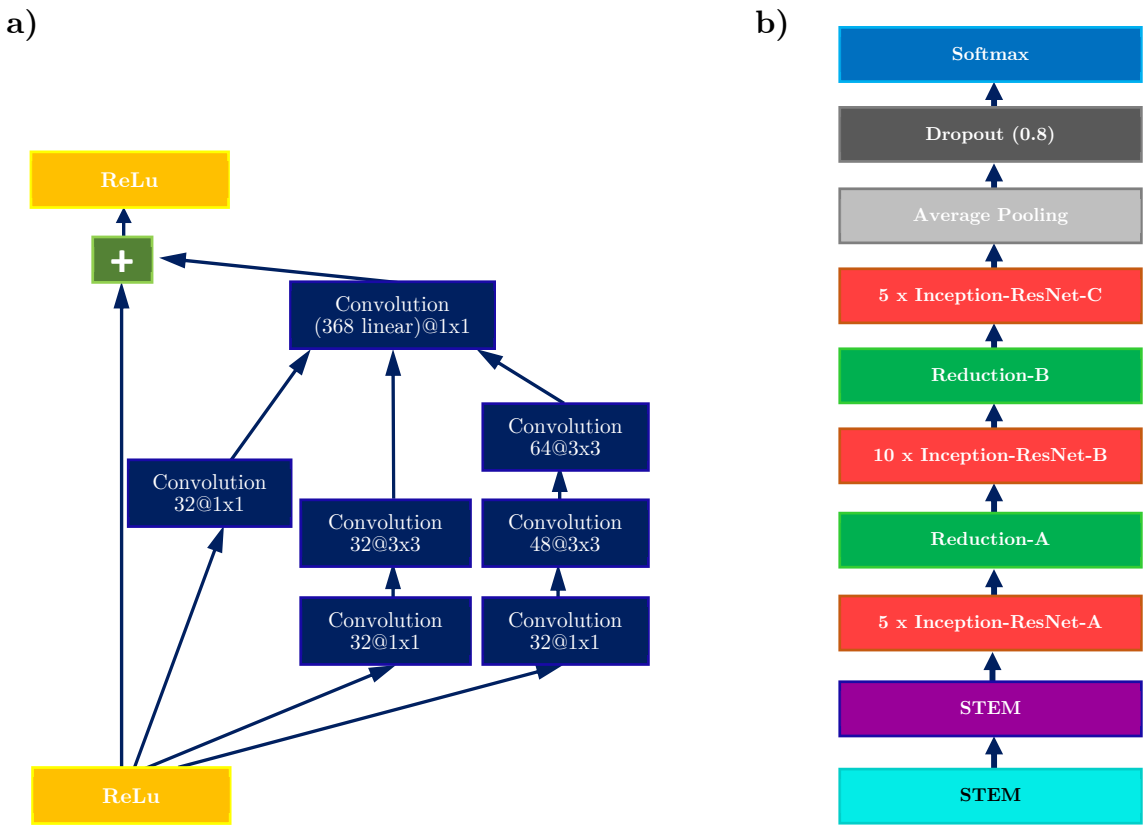



**Figure 2. 5. a)** Inception-Resnet-A unit; **b)** structural components of Inception-ResNet-v2 network[94].

### 2.5.7 Deep Learning Based Image Segmentation

This section will briefly describe the basic principles of a U-Net[96] network commonly used for automated segmentation of medical images[97,98]. As shown in Figure 2.19, U-net takes a two-dimensional image as an input and passes it through several convolutional layers of the Encoder that sequentially down samples the feature maps of that input image. Then the Decoder up samples these feature maps, passes them through a few convolutional layers and concatenates with the feature maps previously created in the Encoder. The last

layer the feature maps are passed through is a Softmax activation layer, after which the network generates its output. In this example it is a binary image representing a contoured structure on a single CT scan.

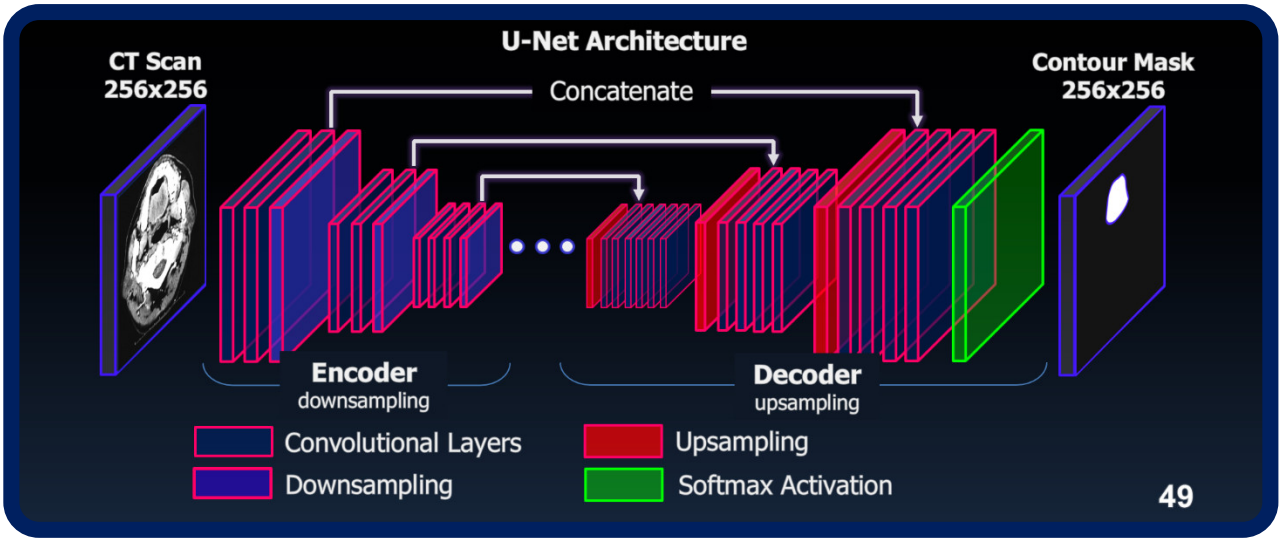



**Figure 2.19.** Schematic illustration of U-net network used for automated segmentation of CT scans.

This network can be improved by replacing simple CNN layers of the Encoder with the Inception-Resnet modules[99] that consist of 7 different types of convolutional layers so that each layer is built with a different number and size of filters, with some of them working in parallel as illustrated in Figure 2.20. Additionally, each Inception-Resnet module has a residual connection. It should be noted that the diagram in Figure 2.20 is a simplification of a U-net Inception-ResNet-v2 network in that it shows only one Inception-Resnet module (Figure 2.18 illustrates a more detailed architecture of the Inception-Resnet-v2 network).

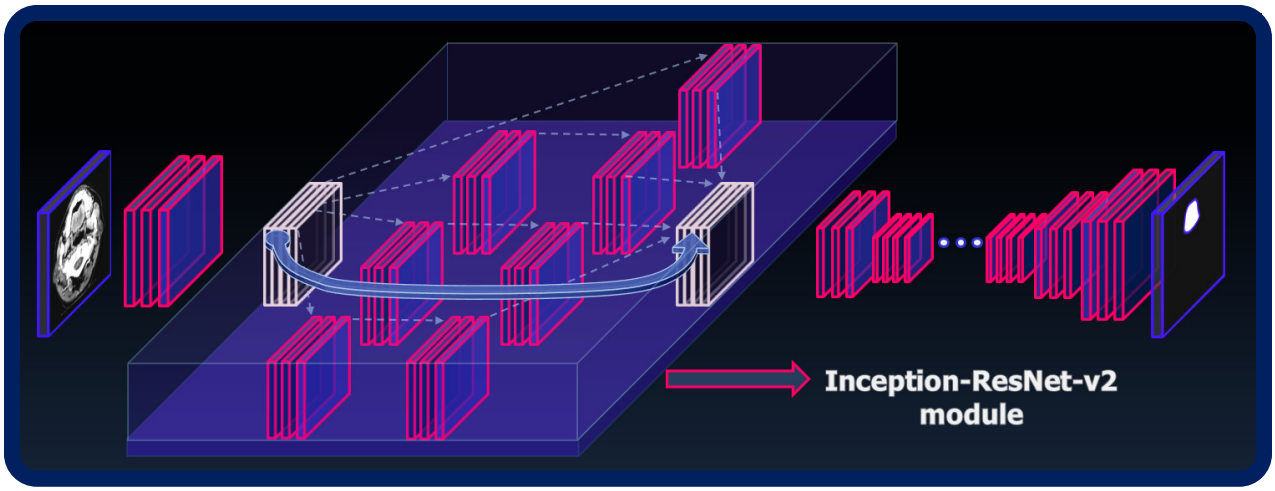



**Figure 2.20.** Simplified architecture of U-Net Inception-ResNet-v2 segmentation network.

## 2.6 References

1. Saha GB. *Physics and radiobiology of nuclear medicine.* Springer Science & Business Media; 2012.

2. Beyzadeoglu M, Ozyigit G, Ebruli C. *Basic radiation oncology.* Springer Science & Business Media; 2010.

3. Chang DS, Lasley FD, Das IJ, Mendonca MS, Dynlacht JR. *Basic radiotherapy physics and biology.* Springer; 2014.

4. Levitt SH, Purdy JA, Perez CA, Vijayakumar S. *Technical basis of radiation therapy.* Springer; 2012.

5. Perez, C. A., Brady, L. W. (2008). Perez and Brady's Principles and Practice of Radiation Oncology. United Kingdom: Wolters Kluwer Health/Lippincott Williams & Wilkins.

6. Rosenwald, J. (2007). Handbook of Radiotherapy Physics: Theory and Practice. United States: Taylor & Francis.

7. Bortfeld T, Neve W, Schmidt-Ullrich R, Wazer DE. *Image-guided IMRT.* Springer; 2006.

8. Gunderson LL, Tepper JE. *Clinical Radiation Oncology E-Book.* Elsevier Health Sciences; 2015.

9. Nishimura Y, Komaki R. *Intensity-Modulated Radiation Therapy: Clinical Evidence and Techniques.* Springer Japan; 2015.

10. Ballhausen H, Li M, Ganswindt U, Belka C. Shorter treatment times reduce the impact of intra-fractional motion. *Strahlentherapie und Onkologie.* 2018;194(7):664-674.

11. Unkelbach J, Bortfeld T, Craft D, et al. Optimization approaches to volumetric modulated arc therapy planning. *Medical physics.* 2015;42(3):1367-1377.

12. Baumann M, Krause M, Overgaard J, et al. Radiation oncology in the era of precision medicine. *Nature Reviews Cancer.* 2016;16(4):234.

13. Bortfeld T, Schmidt-Ullrich R, De Neve W and Wazer DE. *Image-Guided IMRT.* Springer Berlin Heidelberg, 2006.

14. Nishimura Y and Komaki R. *Intensity-Modulated Radiation Therapy: Clinical Evidence and Techniques*. Springer Japan, 2015.

15. Van Dyk J and Van Dyk J. *The Modern Technology of Radiation Oncology: A Compendium for Medical Physicists and Radiation Oncologists*. Medical Physics Pub., 2005.

16. Jia X and Jiang SB. *Graphics Processing Unit-Based High Performance Computing in Radiation Therapy*. CRC Press, 2015.

17. Rigaud B, Simon A, Castelli J, et al. Deformable image registration for radiation therapy: principle, methods, applications and evaluation. *Acta Oncologica.* 2019;58(9):1225-1237.

18. Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. *Medical physics.* 2017;44(7):e43-e76.

19 Oh, Seungjong, and Siyong Kim. "Deformable image registration in radiation therapy." *Radiation oncology journal* 35.2 (2017): 101.

20. Oliveira FP, Tavares JMR. Medical image registration: a review. *Computer methods in biomechanics and biomedical engineering.* 2014;17(2):73-93.

21. El-Baz AS, U RA, Laine AF, Suri JS. *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies: Volume II.* Springer New York; 2011.

22. Crum WR, Hartkens T, Hill D. Non-rigid image registration: theory and practice. *The British journal of radiology.* 2004;77(suppl_2):S140-S153.

23. Yip, Stephen, Timothy Perk, and Robert Jeraj. "Development and evaluation of an articulated registration algorithm for human skeleton registration." *Physics in Medicine & Biology* 59.6 (2014): 1485.

24. Reaungamornrat, S., et al. "Deformable image registration with local rigidity constraints for cone-beam CT-guided spine surgery." *Physics in Medicine & Biology* 59.14 (2014): 3761.

25. König, Lars, et al. "Deformable image registration for adaptive radiotherapy with guaranteed local rigidity constraints." *Radiation Oncology* 11.1 (2016): 1-9.

26. Awan MJ, Siddiqui F, Schwartz D, Yuan J, Machtay M, Yao M. Application of positron emission tomography/computed tomography in radiation treatment planning for head and neck cancers. *World journal of radiology.* 2015;7(11):382.

27. van Egmond SL, Piscaer V, Janssen LM, et al. Influence of FDG-PET on primary nodal target volume definition for head and neck carcinomas. *Acta Oncologica.* 2016;55(9-10):1099-1106.

28. Sultana S, Song DY, Lee J. A deformable multimodal image registration using PET/CT and TRUS for intraoperative focal prostate brachytherapy. Paper presented at: Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling2019.

29. Kai Y, Arimura H, Toya R, et al. Comparison of rigid and deformable image registration for nasopharyngeal carcinoma radiotherapy planning with diagnostic position PET/CT. *Japanese Journal of Radiology.* 2020;38(3):256-264.

30. Acosta O, Dowling J, Drean G, Simon A, de Crevoisier R, Haigron P. Multi-atlas-based segmentation of pelvic structures from CT scans for planning in

prostate cancer radiotherapy. In: *Abdomen and Thoracic Imaging.* Springer; 2014:623-656.

31. Hoang Duc AK, Eminowicz G, Mendes R, et al. Validation of clinical acceptability of an atlas-based segmentation algorithm for the delineation of organs at risk in head and neck cancer. *Medical physics.* 2015;42(9):5027-5034.

32. Tao C-J, Yi J-L, Chen N-Y, et al. Multi-subject atlas-based auto-segmentation reduces interobserver variation and improves dosimetric parameter consistency for organs at risk in nasopharyngeal carcinoma: a multi-institution clinical study. *Radiotherapy and Oncology.* 2015;115(3):407-411.

33. Ciardo D, Gerardi MA, Vigorito S, et al. Atlas-based segmentation in breast cancer radiotherapy: evaluation of specific and generic-purpose atlases. *The Breast.* 2017;32:44-52.

34. Schipaanboord B, Boukerroui D, Peressutti D, et al. An evaluation of atlas selection methods for atlas-based automatic segmentation in radiotherapy treatment planning. *IEEE transactions on medical imaging.* 2019;38(11):2654-2664.

35. Siddiqui F, Kumarasiri A, Kim J, Liu C, Chetty I. Deformable Image Registration-Based Automatic CT-to-CT Contour Propagation for Head-and-Neck Cancer Radiation Therapy: Imaging. *International Journal of Radiation Oncology • Biology • Physics.* 2014;88(2):503.

36. Chapman CH, Polan D, Vineberg K, et al. Deformable image registration–based contour propagation yields clinically acceptable plans for MRI-based cervical cancer brachytherapy planning. *Brachytherapy.* 2018;17(2):360-367.

37. Loi G, Fusella M, Lanzi E, et al. Performance of commercially available deformable image registration platforms for contour propagation using patient-based computational phantoms: A multi-institutional study. *Medical physics.* 2018;45(2):748-757.

38. Motegi K, Tachibana H, Motegi A, Hotta K, Baba H, Akimoto T. Usefulness of hybrid deformable image registration algorithms in prostate radiation therapy. *Journal of applied clinical medical physics.* 2019;20(1):229-236.

39. Hay LK, Paterson C, McLoone P, et al. Analysis of dose using CBCT and synthetic CT during head and neck radiotherapy: A single centre feasibility study. *Technical Innovations & Patient Support in Radiation Oncology.* 2020;14:21-29.

40. Rigaud B, Simon A, Castelli J, et al. Evaluation of deformable image registration methods for dose monitoring in head and neck radiotherapy. *BioMed research international.* 2015;2015.

41. Veiga C, Lourenço AM, Mouinuddin S, et al. Toward adaptive radiotherapy for head and neck patients: uncertainties in dose warping due to the choice of deformable registration algorithm. *Medical physics.* 2015;42(2):760-769.

40. Azcona JD, Huesa-Berral C, Moreno-Jiménez M, Barbés B, Aristu JJ, Burguete J. A novel concept to include uncertainties in the evaluation of stereotactic body radiation therapy after 4D dose accumulation using deformable image registration. *Medical physics.* 2019;46(10):4346-4355.

41. Lowther NJ, Marsh SH, Louwe RJ. Quantifying the dose accumulation uncertainty after deformable image registration in head-and-neck radiotherapy. *Radiotherapy and Oncology.* 2020;143:117-125.

42. Yuan Z, Rong Y, Benedict SH, Daly ME, Qiu J, Yamamoto T. "Dose of the day" based on cone beam computed tomography and deformable image registration for lung cancer radiotherapy. *Journal of applied clinical medical physics.* 2020;21(1):88-94.

43. Green OL, Henke LE, Hugo GD. Practical clinical workflows for online and offline adaptive radiation therapy. Paper presented at: Seminars in radiation oncology2019.

44. Nigay E, Bonsall H, Meyer B, Hunzeker A, Lenards N. Offline adaptive radiation therapy in the treatment of prostate cancer: a case study. *Medical Dosimetry.* 2019;44(1):1-6.

45. Reilly M, Kavanaugh J, Green O, Mutic S. Quantitative and Dosimetric Evaluation of Offline Adaptive Radiation Therapy Toward Establishing a Decision Support Framework for Evaluating the Necessity for Real-Time Adaptation. *International Journal of Radiation Oncology • Biology • Physics.* 2016;96(2):S226.

46. Yang C, Liu F, Ahunbay E, et al. Combined online and offline adaptive radiation therapy: A dosimetric feasibility study. *Practical radiation oncology.* 2014;4(1):e75-e83.

47. Zhang B, Lee S-W, Chen S, et al. Action levels on dose and anatomic variation for adaptive radiation therapy using daily offline plan evaluation: preliminary results. *Practical radiation oncology.* 2019;9(1):49-54.

48. El-Bared N, Portelance L, Spieler BO, et al. Dosimetric benefits and practical pitfalls of daily online adaptive MRI-guided stereotactic radiation therapy for pancreatic cancer. *Practical radiation oncology.* 2019;9(1):e46-e54.

49. Kibrom AZ, Knight KA. Adaptive radiation therapy for bladder cancer: a review of adaptive techniques used in clinical practice. *Journal of medical radiation sciences.* 2015;62(4):277-285.

50. Lamb J, Cao M, Kishan A, et al. Online Adaptive Radiation Therapy: Implementation of a New Process of Care [published online ahead of print 2017/11/07]. *Cureus.* 2017;9(8):e1618.

51. Lim-Reinders S, Keller BM, Al-Ward S, Sahgal A, Kim A. Online adaptive radiation therapy. *International Journal of Radiation Oncology* Biology* Physics.* 2017;99(4):994-1003.

52. van Kranen S, Hamming-Vrieze O, Wolf A, Damen E, van Herk M, Sonke J-J. Head and neck margin reduction with adaptive radiation therapy: robustness of treatment plans against anatomy changes. *International Journal of Radiation Oncology* Biology* Physics.* 2016;96(3):653-660.

53. Glide-Hurst, Carri K., et al. "Adaptive radiation therapy (ART) strategies and technical considerations: A state of the ART review from NRG Oncology." *International Journal of Radiation Oncology* Biology* Physics* (2020).

54. Heukelom, Jolien, and Clifton David Fuller. "Head and neck cancer adaptive radiation therapy (ART): conceptual considerations for the informed clinician." *Seminars in radiation oncology.* Vol. 29. No. 3. WB Saunders, 2019.

55. Green, Olga L., Lauren E. Henke, and Geoffrey D. Hugo. "Practical clinical workflows for online and offline adaptive radiation therapy." *Seminars in radiation oncology.* Vol. 29. No. 3. WB Saunders, 2019.

56. Chen Z, Yang Z, Wang J, Hu W. Dosimetric impact of different bladder and rectum filling during prostate cancer radiotherapy. *Radiation Oncology.* 2016;11(1):103.

57. Wahl M, Descovich M, Shugard E, et al. Interfraction anatomical variability can lead to significantly increased rectal dose for patients undergoing stereotactic body radiotherapy for prostate cancer. *Technology in Cancer Research & Treatment.* 2017;16(2):178-187.

58. Huang T-C, Chou K-T, Yang S-N, Chang C-K, Liang J-A, Zhang G. Fractionated changes in prostate cancer radiotherapy using cone-beam computed tomography. *Medical Dosimetry.* 2015;40(3):222-225.

59. de Boer J, Wolf AL, Szeto YZ, van Herk M, Sonke J-J. Dynamic collimator angle adjustments during volumetric modulated arc therapy to account for prostate rotations. *International Journal of Radiation Oncology* Biology* Physics.* 2015;91(5):1009-1016.

60. Brouwer CL, Steenbakkers RJ, Langendijk JA, Sijtsema NM. Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? *Radiotherapy and oncology.* 2015;115(3):285-294.

61. Zhao L, Wan Q, Zhou Y, Deng X, Xie C, Wu S. The role of replanning in fractionated intensity modulated radiotherapy for nasopharyngeal carcinoma. *Radiotherapy and Oncology.* 2011;98(1):23-27.

62. Cheng HC, Wu VW, Ngan RK, et al. A prospective study on volumetric and dosimetric changes during intensity-modulated radiotherapy for nasopharyngeal carcinoma patients. *Radiotherapy and Oncology.* 2012;104(3):317-323.

63. Panet-Raymond V, Ansbacher W, Zavgorodni S, et al. Coplanar versus noncoplanar intensity-modulated radiation therapy (IMRT) and volumetric-modulated arc therapy (VMAT) treatment planning for fronto-temporal high-grade glioma. *Journal of applied clinical medical physics.* 2012;13(4):44-53.

64. Kupelian P, Sonke J-J. Magnetic resonance–guided adaptive radiotherapy: a solution to the future. Paper presented at: Seminars in radiation oncology. 2014.

65. Cao Y, Tseng C-L, Balter JM, Teng F, Parmar HA, Sahgal A. MR-guided radiation therapy: transformative technology and its role in the central nervous system. *Neuro-oncology.* 2017;19(suppl_2):ii16-ii29.

66. Darázs B, Ruskó L, Vegvary Z, et al. Adaptive radiation therapy for high grade brain tumors: impact on the dose distribution and disease outcome. *International Journal of Radiation Oncology • Biology • Physics.* 2017;99(2):E79.

67. Ponsky LE, Fuller DB, Meier RM, Ma C. *Robotic Radiosurgery Treating Prostate Cancer and Related Genitourinary Applications.* Springer Berlin Heidelberg; 2011.

68. Hurkmans CW, Dijckmans I, Reijnen M, van der Leer J, van Vliet-Vroegindeweij C, van der Sangen M. Adaptive radiation therapy for breast IMRT-simultaneously integrated boost: Three-year clinical experience. *Radiotherapy and Oncology.* 2012;103(2):183-187.

69. Li X, Ahunbay E, Godley A, Morrow N, Wilson J, White J. An online replanning technique for breast adaptive radiation therapy. *International Journal of Radiation Oncology • Biology • Physics.* 2009;75(3):S71.

70. Tanderup K, Georg D, Pötter R, Kirisits C, Grau C, Lindegaard JC. Adaptive management of cervical cancer radiotherapy. Paper presented at: Seminars in radiation oncology2010.

71. Li XA, Liu F, Tai A, et al. Development of an online adaptive solution to account for inter-and intra-fractional variations. *Radiotherapy and Oncology.* 2011;100(3):370-374.

72. Samuel AL. Some studies in machine learning using the game of checkers. *IBM Journal of research and development.* 1959;3(3):210-229. 65

73. Mitchell T. Machine Learning, McGraw-Hill Higher Education. *New York.* 1997.

74. El Naqa I, Li R, Murphy MJ. *Machine learning in radiation oncology: theory and applications.* Springer; 2015.

75. Dadson N, Pinheiro L, Royer J. Decision Making with Machine Learning in Our Modern, Data-Rich Health-Care Industry. In: *Decision Making in a World of Comparative Effectiveness Research.* Springer; 2017:277-289.

76. Mohri M, Rostamizadeh A, Talwalkar A. *Foundations of Machine Learning, second edition.* MIT Press; 2018.

77. Géron A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.* O'Reilly Media; 2019.

78. Cochran JJ. *Informs analytics body of knowledge.* John Wiley & Sons; 2018.

79. Lee WM. *Python Machine Learning.* Wiley; 2019.

80. Cui S, Tseng HH, Pakela J, Ten Haken RK, El Naqa I. Introduction to machine and deep learning for medical physicists. *Medical Physics.* 2020;47(5):e127-e147.

81. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics.* 2002;35(5-6):352-359.

82. Hastie T, Tibshirani R, Wainwright M. *Statistical learning with sparsity: the lasso and generalizations.* CRC press; 2015.

83. Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied logistic regression.* Vol 398: John Wiley & Sons; 2013.

84. Smola BSCJCBAJ, Soentpiet R, Schölkopf B, et al. *Advances in Kernel Methods: Support Vector Learning.* MIT Press; 1999.

85. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition.* Springer New York; 2009.

86. Campbell C, Ying Y. *Learning with Support Vector Machines.* Morgan & Claypool; 2011.

87. Berry MW, Mohamed A, Yap BW. *Supervised and Unsupervised Learning for Data Science.* Springer; 2019.

88. Ranschaert ER, Morozov S, Algra PR. *Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks.* Springer International Publishing; 2019.

89. da Silva IN, Spatti DH, Flauzino RA, Liboni LHB, dos Reis Alves SF. *Artificial Neural Networks: A Practical Course.* Springer International Publishing; 2016.

90. Buduma N, Locascio N. *Fundamentals of deep learning: Designing next-generation machine intelligence algorithms.* " O'Reilly Media, Inc."; 2017.

91. Aggarwal CC. *Neural Networks and Deep Learning: A Textbook.* Springer International Publishing; 2018.

92. Shrestha A, Mahmood A. Review of deep learning algorithms and architectures. *IEEE Access.* 2019;7:53040-53065.

93. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.

94. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. Paper presented at: Proceedings of the AAAI Conference on Artificial Intelligence; 2017.

95. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.

96. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.

97. Ghosh, Swarnendu, et al. "Understanding deep learning techniques for image segmentation." *ACM Computing Surveys (CSUR)* 52.4 (2019): 1-35.

98. Hesamian, Mohammad Hesam, et al. "Deep learning techniques for medical image segmentation: achievements and challenges." *Journal of digital imaging* 32.4 (2019): 582-596.

99. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. Paper presented at: Proceedings of the AAAI Conference on Artificial Intelligence; 2017.

# Chapter 3

# Evaluation of CT to CBCT non-linear Dense Anatomical Block Matching registration for prostate patients

Deformable image registration (DIR) is a rapidly developing discipline in the field of medical imaging that has found numerous applications in modern radiation therapy. To be used in the clinical environment, DIR requires an accurate and robust algorithm supported by careful evaluation. The purpose of the study described in this chapter was to evaluate the performance of the non-linear Dense Anatomical Block Matching (DABM) algorithm for CT-CBCT image registration of prostate cancer patients. This work demonstrated that for all the patients and anatomical structures considered here, both the accuracy and the consistency of the DABM algorithm are considerably better than several other commonly-implemented registration methods. Generated deformation vector fields (DVFs) have a well-preserved topology and small inverse consistency errors (ICEs). Presented findings showed that DABM is a promising alternative to the existing common strategies for CT-CBCT image registration and its application in the adaptive radiation therapy of the pelvic region. This chapter was published in the peer-reviewed Journal of Biomedical Physics and Engineering Express.

## 3.1 Introduction

Deformable image registration (DIR) is a rapidly developing discipline in the field of medical imaging. Its goal is to establish the spatial correspondence between two images. In the DIR process, the image to be registered i.e. the 'moving' or 'floating' image is deformed by a non-linear transformation so that it matches the image of interest i.e. the 'fixed' or 'target' image[1].

Due to its ability to map the patient's anatomy at one point in time to the anatomy at another point in time, deformable image registration has found numerous applications in modern radiation therapy[2] such as contour propagation[3], dose accumulation[4], automatic image segmentation[5], multimodality image fusion[6] and analysis of organ motion[7]. Currently, DIR is also widely used in Adaptive Radiation Therapy (ART) studies[8]. For example, Li et al.[9] and Yu et al.[10] utilized deformable image registration to propagate the contours between patients' daily image data sets, perform dose accumulation, and based on that, accounted for anatomy changes by adapting the original treatment plan in established ART frameworks. Veiga et al.[11] conducted studies using deformable registration on computed tomography (CT) to cone-beam CT (CBCT) image data sets for "dose of the day" calculations and examined uncertainties in dose warping due to the choice of deformable registration algorithm[12], showing its potential benefits in adaptive radiation therapy. More recent papers also describe the influence of the image registration methods on the adaptive radiotherapy in a selected case of prostate IMRT[13] and DIR applications in dose accumulation purposes[14].

The ultimate goal of these DIR applications in radiation therapy is to improve the clinical outcome of treatment, make it more efficient, and more tailored to the individual patient. Achieving this goal, however, remains a challenge because medical image registration is often not as accurate as desired due to the limited performance of DIR algorithms, poor quality of images to be registered, or the characteristics of the patient's anatomical changes[15].

Considering the important role in the clinical decision-making process they may serve, it is clear that DIR methods need careful evaluation prior to clinical implementation. Assessment procedures, both on phantoms and patient data, focus mainly on the comparison of contours and images between deformed and target data sets using a variety of quantitative metrics. Contour comparisons usually include the Dice Coefficient[16], Hausdorff Distance, and Average Surface Distance[17] while for image comparison the Normalized Cross Correlation[18], Normalized Mutual Information[18,] and Mean Squared Error[17] are commonly utilized. Such numerical analyses are often supported by clinical experts who not only visually evaluate the results of the image registration[19] but also assist in quantitative analyses, for example by identifying anatomical landmarks that can be analyzed after DIR[20]. Apart from images and contours, the deformation vector field (DVF) estimated during the image registration process should be evaluated as well. For instance, Varadhan et al.[17] and Veiga et al.[12] used the Jacobian Determinant metric to characterize the topology preservation of DVF to ensure that the deformed image does not contain any unphysical anatomy deformations. For quality assurance of dose accumulation applications, the Inverse Consistency Error (ICE) describing the inverse of the vector field has also been examined frequently[12].

An important factor that affects the performance of DIR is the quality of images to be registered. In the case of CT-CBCT image registration, CBCT as the target image introduces a significant challenge due to the presence of image artifacts and reduced contrast of soft tissues[21]. Numerous methods have been introduced to improve the image quality of CBCT. For instance, Lou et al. proposed a viscous fluid model that incorporates an intensity correction for CBCT image enhancement[22], while Thing et al.[23] and Watson et al.[24] suggested efficient Monte Carlo simulations for CBCT scatter correction, which improves CBCT contrast. Zhang et al.[25] focused on artifact reduction using optimization-based reconstruction. Despite those efforts, the authors indicate that there is still room for improvement in correcting CBCT images.

Another challenge for deformable CT-CBCT image registration is inconsistency in the intensity values of CT and CBCT images, which means that corresponding voxels in CBCT and CT may not have the same intensity. Zhen et al[26] accounted for this problem by applying a variant of the Demons DIR algorithm with an added intensity correction term for CT-CBCT registration. Meanwhile, Yu et al.[10] used gradient-based free-form deformation (GFFD), whereas Park et al.[27] iteratively corrected CBCT intensities by local histogram matching. Unfortunately, all these methods, together with the approach of image quality enhancement of the CBCT itself mentioned in the previous paragraph, add extra time and complexity to the image registration process, which can negatively impact on clinical adoption. However, CBCT is a commonly used tool for radiotherapy setup verification[28,29] and there remains great potential for increased utilization of CT-CBCT image registration in Adaptive Radiation Therapy[12,30]. Therefore, there is a strong need

for further improvement of this type of registration, especially in the quantitative evaluation of specific deformable image registration algorithms.

The purpose of this work was to evaluate the performance of the non-linear Dense Anatomical Block Matching (DABM) algorithm for CT-CBCT image registration of five prostate cancer patients. The DABM method was developed by Garcia et al.[31] and Commowick et al.[32] and was initially utilized in their studies for registration of thoracic CT and dynamic cervical MRI images respectively. Later on, Huger et al.[33] applied the algorithm to CT-CBCT registration of head-and-neck patients. However, there is a lack of work quantitatively evaluating the performance of the DABM algorithm for CT-CBCT image registration for other clinical sites. We compare the performance of the DABM algorithm with other widely used registration methods including Affine and Bspline algorithms.

## 3.2 Material and Methods

### 3.2.1 Acquisition of Patients' Data

Five prostate cancer patients that underwent intensity-modulated radiation therapy (IMRT) were selected for this study. 74 Gy of radiation dose was prescribed for each patient and delivered in daily 2 Gy doses over 37 fractions. In this work, pre-treatment planning CT (pCT) scans and the CBCT data set acquired in the middle of the treatment course ('mid-treatment CBCT'; one CBCT per patient) were used for analysis. Gross tumor volume (GTV), bladder, and rectum were segmented by an experienced radiation oncologist on both image data sets while the PTV was created by automatic expansion of GTV.

The mid-treatment CBCT scans were selected as they provide the deformable registration algorithms with realistically challenging registration problems. This justification is supported by previous CT-CBCT and CT-CT DIR studies in which the target images of the patient's anatomy were acquired in the middle of the radiation therapy treatment and then evaluation of image registration algorithms was conducted [19,34,35].

The pCT images were obtained with a spatial resolution of 0.98 mm x 0.98 mm per pixel and 2.0 mm slice thickness, using the helical mode of a Toshiba Aquilion LB CT scanner with the following scan parameters: 120 kVp, 398 mAs, and FOV of 50 cm. CBCT images with a resolution of 1.17 mm x 1.17 mm per pixel and 2.5 mm slice thickness were acquired by a Varian On-Board Imager® (OBI) System with the following scan parameters: 125 kVp, 80 mA, 8 ms, and field-of-view (FOV) of 45 cm.

### 3.2.2 Registration Algorithms

CT-CBCT image registrations using Affine, Bspline, and Dense Anatomical Block Matching algorithms were performed for five patients. Each registration was preceded by Rigid alignment. The parameters of each particular algorithm were chosen on an empirical basis to provide us with a reasonable trade-off between registration accuracy and computational time efficiency. A more detailed explanation for parameter selection is described in the following paragraphs.

The fraction of voxels of the fixed image that was used for Rigid and Affine registration was set to 0.4% and 1% respectively (registration filter uses random sampling on fixed image). The maximum number of iterations was set at 1500 for both algorithms. Cost function used was Mutual Information. Obtained transformations were applied to the

floating image through linear interpolation. We observed that increasing the fraction of voxels and/or the maximum number of iterations increased the computational time but had an insignificant effect on the accuracy of the registration. A decrease in the value of these parameters resulted in incorrect registrations that were easily detected by visual inspection. Especially, Rigid alignment revealed undesirable shifts in axial, coronal, and sagittal planes depending on the specific patient geometry. Rotations, although present, were not significant. Both the Affine and the Rigid registrations were performed based on the implementation in open-source 3D Slicer Software.

The Bspline registration method used in this study is modeled as a weighted sum of Bspline basis functions, placed on a uniform control point grid where free-form deformations (FFDs) based on Bspline interpolation of local deformations are calculated. FFDs are then utilized to transform the floating image by manipulating the control points of the grid[36,37]. In our multi-resolution (two resolution levels) registrations, we used an adaptive stochastic gradient descent optimizer (max. 1000 iterations) as it demonstrates low sensitivity to the settings of the user-defined parameters and does not require predetermination of step size[38]. Mutual information was set as a similarity metric because it is appropriate for multimodality deformable image registration[39]. Elastix software developed at Image Sciences Institute, University Medical Center Utrecht was used to perform the registration[40].

The Dense Anatomical Block Matching algorithm used in this study calculates spatially sparse, locally optimal Rigid transformations between floating and reference images using a multi-resolution, pyramidal bock matching approach and numerical optimizer called BOBYQA[41]. Next, based on these local transformations it derives a dense, regularized

vector field. These two steps are iterated until the final diffeomorphic transformation is defined[32]. To improve the robustness of the registration, the block-matching is coupled with an outlier rejection scheme to remove incorrect pairings that may occur due to several factors such as noise or lesions[31]. The most relevant parameters of the algorithm were presented in Table 3.1 below.

**Table 3. 1.** Parameters of the DABM algorithm

| **Parameter** | **Value** |
|---|---|
| Similarity metric | Squared correlation |
| Block size | 5x5x5 voxels |
| Regularization | Baloo aggregator* with all $\sigma$ =2.5 (extrapolation, elastic and outlier sigma) |
| Pyramid levels | 4 |
| Optimal block search optimizer | BOBYQA |
| Max. block match iterations | 7 |
| Max. local optimizer iterations | 100 |

*Baloo aggregator is one of the modules of Dense Anatomical Block Matching Algorithm implementation that performs the following operations: i) interpolation of a dense correction field; ii) computation of outlier-free transformation correction; iii) composition of transformation correction with the current transformation that might be regularized using an elastic-like regularization.[31]

The time to complete the registration was mostly influenced by the choice of pyramid levels. Nevertheless, the enhancement in the accuracy of the registration was negligible as the pyramid levels were increased above four. A similar correlation between image

registration quality and the computational time was observed for block size, search optimizer, block match iterations, and local optimizer iterations. The last critical user-set parameter was the 'elastic sigma' used in the regularization step of the image registration process. We found that even a slight increase in the value of $\sigma$ from 2.5 to 3.0 caused the vector field to be overly flexible. As a result, the deformed image contained unrealistic (irregular and large) deformations in many regions of the patient's anatomy, especially those regions that were most visible near the external body contour. MedInria software[42] developed at the French National Institute for Research in Computer Science and Control (Inria, Paris, France) was used to perform the registration.

### 3.2.3 Registration Framework

In the case of all five patients, their pCT scans (floating images) were registered to the CBCT data sets (reference images) using the two deformable and one affine image registration algorithms described in section 3.2.2 above. Each registration was preceded by an initial Rigid alignment of the image sets. As a result of image registration, the deformed pCT ('dCT') and vector fields were obtained. Next, structures (PTV, GTV, bladder, and rectum) delineated on the pCT by an experienced radiation oncologist were deformed using the acquired vector fields and propagated to the CBCT images. Subsequently, a quantitative evaluation of all image registration methods was conducted.

### 3.2.4 Measurements of Image Registration Accuracy

The accuracy of image registration was evaluated by the application of several standard metrics. Dice Score (DSC), Hausdorff Distances (HD; average and 95th percentile), and Center of the Mass Shift (COM) were used in the case of the structure comparison -

deformed contours from pCT (GTV, PTV (by GTV expansion), bladder, and rectum) were compared to contours delineated directly on the CBCT by a physician i.e. the reference contours. Deformed to CBCT contours were also evaluated by a physician to analyze their clinical accuracy. The topology of vector fields obtained from the deformable image registration was accessed by the Jacobian determinant metric. The time to calculate each registration was recorded as well. It is noted that although the PTV was created by an expansion, it is a reliable and consistent contour for comparison purposes.

For two evaluated deformable registration algorithms, we inverted its vector field using the Insight Segmentation and Registration Toolkit (ITK)[43] implementation of the Modified Newton Method[44] and calculated the associated Inverse Consistency Error (ICE). This quantifies the performance of the chosen DIR technique for dose accumulation – a critical task in any Adaptive Radiation Therapy framework. To evaluate the symmetry of the deformable image registration methods we also calculated ICE using CBCT-CT registration.

All mentioned methods of image registration evaluation are described in more detail below.

*Dice Score*

Dice Score (DSC) measures the mean volume overlap between two structures and was calculated based on the ITK implementation[45] as DSC=2*$|V_1 \cap V_2|/(|V_1| + |V_2|)$; where $V_1$ and $V_2$ are the volumes of individual structures defined by the deformed and reference contours respectively.

*Hausdorff Distances*

For contour comparison purposes we also calculated the undirected averaged Hausdorff Distance ($HD_{avg}$) and undirected 95th percentile of Hausdorff Distance ($HD_{95}$). Both $HD_{avg}$ and $HD_{95}$ were calculated in SlicerRT software[46] using the Plastimatch implementation[47].

*Center of the mass shift*

For all the structures considered in this work, the center of the mass shift (COM; the geometrical center of the structure contours) was calculated to indicate the global displacement between the reference and deformed contours[19]. As mentioned by Hardcastle et al., evaluation of COM, especially for the GTV, is useful as it affects the position of the isocenter in the case of radiation treatment replanning[48].

*Physician Evaluation*

A physician edited the contours deformed from pCT to CBCT to ensure that they accurately reflected the patient's anatomy contained in the CBCT data sets. We assumed that the less time that was needed to correct the structures, the better they reflected the daily patient anatomy before the correction, and thus the better the performance of a particular image registration method. This analysis was made for the GTV, bladder, and rectum. The PTV was not considered, as in clinical practice it is generated automatically based on margin expansion around the GTV or CTV (clinical target volume). Times were measured by the physician to an accuracy of 1s per structure (with no distractions while segmenting each structure). The physician was very familiar with the contouring software

used. During contour correction, the physician was not aware of which contours were deformed by which registration algorithm.

Also, it was expected that corrected contours that segment the same anatomical structures should be the same regardless of which algorithm was applied in their deformation initially. However, it is well-known that in the case of segmenting the same anatomical region a few times even by the same physician, the contours will differ. That difference between considered contours is called intra-physician variability and its analysis, using DSC and $HD_{avg}$ metrics, was included in this study as well.

### 3.2.5 Time Efficiency of Image Registration

As with spatial accuracy, calculation efficiency is a crucial factor in accessing the utility of registration algorithms in the clinical environment. All image alignment computations were GPU-based and performed on a workstation equipped with a 4-core, 1.70GHz processor and 8 GB of RAM. Times required to complete each registration were measured with 1s accuracy.

### 3.2.6 Evaluation of Deformation Vector Field

*Jacobian Determinant*

The Jacobian determinant ($|J|$) metric analyzes the topology of the vector field generated by the deformable image registration algorithm. In particular, a value of Jacobian determinant below zero indicates the appearance of unrealistic deformations in the patient's anatomy[17] that can be interpreted as the lack of the proper neighborhood relationship and connectivity of anatomical structures[49]. Therefore, a desirable, well-

preserved deformation vector field should have $|J|>0$, reflecting realistic local compression and expansion of considered volumes respectively, while $|J|<0$ represents an unreal folding of structures. It is noted that the Jacobian takes an input vector field only, which does not carry information about deformability properties of the underlying anatomical structures (i.e. it is not aware of differences between bone and soft tissue).

*Inversion of a deformation vector field*

The forward transformations generated by the Dense Anatomical Block Matching algorithm were inverted using the Modified Newton Method (MNM) as implemented in VTK (Visualization Toolkit). MNM is more reliable and calculates the inverted transformation faster than the original Newton's method. The modification is based on the distance between the transformed point and the forward transformation of the point that is inverted. If after any iteration this distance increases, it is then minimized by the application of a quadratic approximation and used in the next iteration step. A more detailed explanation of that method is given by Gobbi and Peters[44].

*Inverse Consistency Error*

The inverse consistency error (ICE) describes the accuracy of inverted transformations. In our work, we calculated ICE using [50]:

$$\mathrm{ICE} = \frac{1}{2}\left(\left\|\vec{x} - \left(\vec{T} \circ \vec{T}^{-1}\right)(\vec{x})\right\| + \left\|\vec{x} - \left(\vec{T}^{-1} \circ \vec{T}\right)(\vec{x})\right\|\right) \qquad \text{(Eq.3-1)}$$

where $\vec{T}$ and $\vec{T}^{-1}$ are forward and inverted transformations respectively. In this form, ICE is a mean distance between the original point $x$ in the floating image and its position in this image after mapping to the reference image and subsequent mapping back to the original, floating image. The inverted transformation $\vec{T}^{-1}$ was determined by two approaches. In the first approach $\vec{T}^{-1}$ was estimated using MNM described in the previous section (inverted forward transformation; $\text{ICE}^{(\text{inv})}$). In the second method $\vec{T}^{-1}$ was obtained through reverse CBCT-CT registration (backward transformation; $\text{ICE}^{(\text{b})}$).

The first approach of determining $\vec{T}^{-1}$ was selected mainly due to the clinical relevance. Namely, in Adaptive Radiation Therapy it is common to first register the pCT (planning CT) to the daily CBCT, then perform a dose calculation (depending on the specific scenario using the adapted and/or original plan) on the deformed image. Once the desired dose distribution is calculated on the deformed image, it needs to be mapped back to the reference image space (in this case, the planning CT) for final analysis. The process of mapping the dose back to the reference image needs to be performed using an inverse transformation (as in detail described for example by[51]). The second approach of obtaining $\vec{T}^{-1}$ is commonly encountered in the literature and some consider that the ICE calculated in this fashion more accurately reflects the level of image registration symmetry [52].

### 3.2.7 Statistical Analysis

The results of the accuracy metrics of the image registration algorithms were analyzed using the arithmetic mean values and associated standard error of the mean (hereafter called simply standard error; $SE$) defined as $SE = SD\sqrt{n}$, where $SD$ is the standard deviation and $n$ is the sample size (ie. here the number of the patients). We choose to

calculate the standard error instead of the standard deviation as it reflects the uncertainty in the mean and its dependency on the sample size. The second factor is of major importance as we used a relatively small sample size in our study (5 patients). The statistical significance of the data was measured with the $p$-value based on the paired t-test and 95% confidence level. The results related to the deformation vector fields were analyzed using the mean, standard deviation, and median values.

## 3.3 Results

### 3.3.1 Measurements of Image Registration Accuracy

Figure 3.1 demonstrates the four metrics used to evaluate the accuracy of the image registration algorithms investigated in this study. Each metric was calculated for PTV, GTV, bladder, and rectum and averaged over all five patients.

In the case of Rigid, Affine, and Bspline registration algorithms, the Dice similarity coefficient yielded similar values of 0.85, 0.8, 0.75, and 0.7 for PTV, bladder, GTV, and rectum respectively. Application of the DABM algorithm improved the DSC to around 0.9 for bladder and rectum, 0.85 for GTV, and 0.8 for the rectum.

The shift in the center of the mass between reference contours and those deformed by the three baseline registration methods was from 6 mm to 8 mm for the rectum and from 4 mm to 6 mm for the remaining structures. Using DABM we obtained improved COM shift values of about 3.0 mm, 2.5 mm, 2.0 mm, and 4.5 mm for PTV, bladder, GTV, and rectum, respectively.

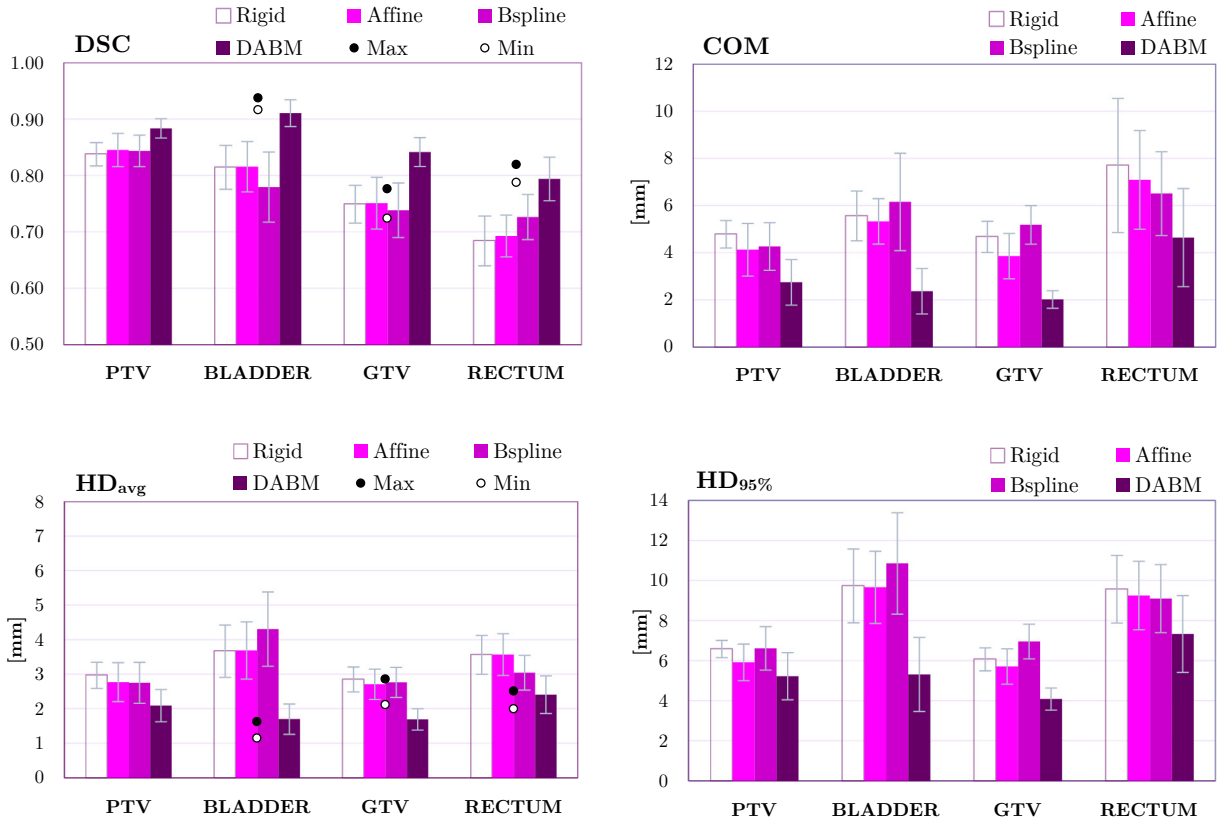



**Figure 3. 1.** Dice Score(DSC) mean Hausdorff Distance ($HD_{avg}$), the 95th percentile of Hausdorff Distance ($HD_{95}$), and center of the mass shift (COM) metrics for Rigid, Affine, Bspline, and Dense Anatomical Block Matching image registration methods for four contours – PTV, bladder, GTV, and rectum. All the values are averaged over five patients and error bars are determined by the standard error. The Max and Min values in the graphs for DSC and $HD_{avg}$ represent the maximum and the minimum differences between contours corrected by the physician.

Mean Hausdorff distances for the baseline registrations were from around 3mm for PTV and GTV to 3-4 mm for bladder and rectum. For DABM registration we observed improved $HD_{avg}$ values of about 2 mm for all structures considered.

The 95$^{th}$ percentiles of Hausdorff distance for reference algorithms were mostly from 6 mm to 7 mm in the case of PTV and GTV while for bladder and rectum $HD_{95\%}$ were approximately 10 mm and 9 mm. For this metric, utilizing the DABM method, we observed improvement to roughly 5 mm for PTV and GTV, 4 mm for bladder, and 7 mm for the rectum.

Figure 3.1 for DSC and $HD_{avg}$ also shows the results of the intra-physician variability by providing the minimum and maximum values of DSC and $HD_{avg}$ metrics as an indication of minimum and maximum differences between contours corrected by the physician. The minimum and maximum values of DSC for bladder, GTV, and rectum were respectively 0.918 and 0.939, 0.725 and 0.777, 0.789 and 0.820. For $HD_{avg}$ those values were equal to 1.146 mm and 1.627 mm, 2.118 mm and 2.859 mm, 1.995 mm, and 2.512 mm for bladder, GTV, and rectum respectively.

Overall, DABM outperforms the other registration methods in every metric and for every anatomical structure which, in more detail, is shown in Figure 3.2 that compares the performance of the Rigid, Affine, and Bspline methods to the performance of the DABM algorithm based on the percentage differences in evaluation matrices. The Dice Score for DABM was improved by around 10% for most of the structures, while other similarity metrics indicated approximately 30-60% (COM), 10-50% ($HD_{95}$), and 25-60% ($HD_{avg}$) difference in favour of Block Matching registration.

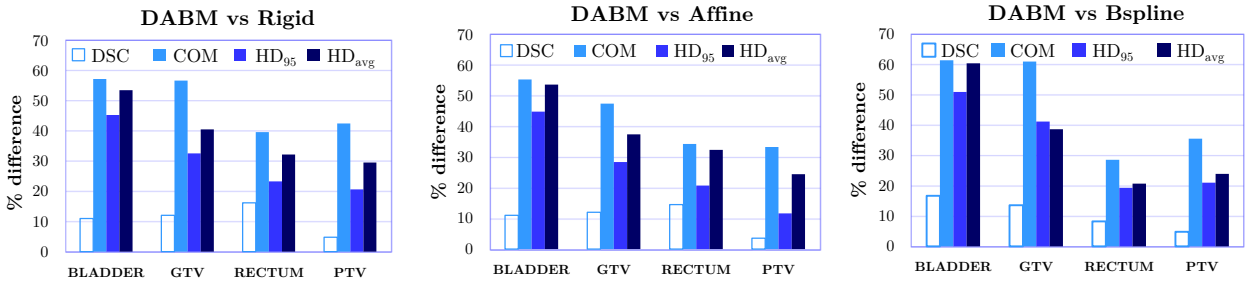



**Figure 3. 2.** Percentage differences between the averaged (over five patients) values of DIR evaluation metrics. Calculations were made for three pairs of registration algorithms and four anatomical structures – bladder, GTV, rectum, and PTV. Positive values of the bars indicate that the value of the given metric is superior for the DABM algorithm by the associated percentage difference.

The charts in Figure 3.2 also demonstrate that structures benefiting the most from the application of the DABM algorithm were bladder, GTV, rectum, and PTV in descending order. Interestingly, if we look at the bladder and GTV in the 'DABM vs Rigid' and 'DABM vs Bspline' graphs it is noticed that the percentage difference for those two structures is larger for the second graph. This suggests that the Bspline registration was less accurate than Rigid. The explanation of this unexpected observation is included in the discussion section. The results for all three comparisons (Figure 3.2) were statistically significant ($p$<0.005).

Apart from higher relative accuracy, the performance of the DABM registration is also more consistent across the patient data sets studied here, compared to the Rigid, Affine, and Bspline algorithms. Table 3.2 shows standard errors (*SE*) averaged over all the patients and anatomical structures. For DSC, COM, and $HD_{avg}$ metrics, the errors of DAMB are mostly smaller than the other registration methods. Results in the case of DSC, COM, and $HD_{avg}$ metrics were statistically significant ($p$<0.05) however no statistical significance was found for $HD_{95}$ ($p$=0.37).

**Table 3. 2.** The standard error (SE) averaged over all patients and anatomical structures.

| Metric | Rigid | Affine | Bspline | DABM |
|---|---|---|---|---|
| DSC | 0.034 | 0.039 | 0.045 | 0.026 |
| COM | 1.286 | 1.284 | 1.419 | 1.096 |
| $HD_{95}$ | 1.133 | 1.326 | 1.545 | 1.374 |
| $HD_{avg}$ | 0.514 | 0.609 | 0.652 | 0.440 |

Similar to the results discussed above, the physician evaluation results also indicate that the accuracy of the DABM technique is superior to the alternate registration methods. As shown in Figure 3.3, correction times varied depending on the anatomical structures and were in the range of 1.3 to 2.2 minutes for DABM, 2.4-3.2 minutes for Bspline, 2.0-3.7 minutes for Affine, and 3.2-5.7 minutes for the Rigid algorithm. The total correction times averaged for each patient data set and calculated as the sum of correction times for

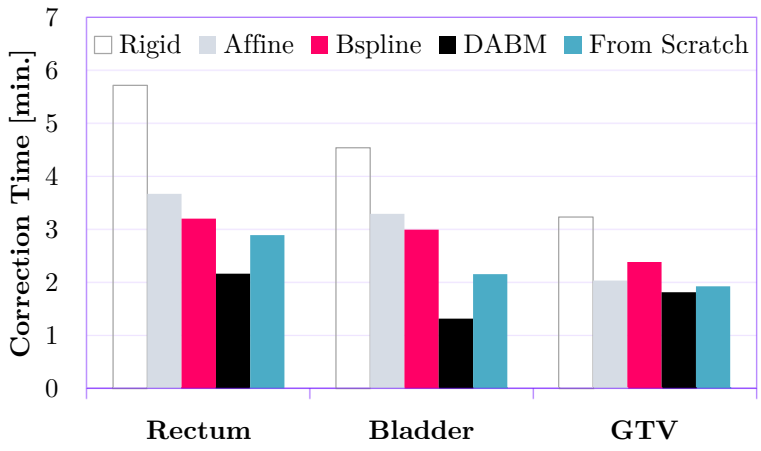



**Figure 3. 3.** Physician correction times for considered registration algorithms and three anatomical structures – rectum, bladder and GTV. All the values are averaged over five patients. Error bars are determined by the standard error. For the purpose of comparison, the time the physician needed to delineate the structures 'from scratch' was included as well.

individual structures were: 5.3 minutes for DABM, 8.6 minutes for Bspline, 9.0 minutes for Affine, and 13.5 minutes for Rigid registration. The physician needed a total of 7.0 minutes to segment the rectum, bladder, and GTV 'from scratch'. The most time-consuming segmentation was found for the rectum, then for bladder and GTV.

### 3.3.2 Time Efficiency of Image Registration

Registration times varied significantly among analyzed algorithms. Measurements indicate that methods with a high level of complexity, i.e. Bspline and DABM, needed more time to complete image alignment than the simpler Rigid and Affine algorithms. Also, as expected the quality of the performance of the registration algorithms was inversely related to their required computational time. The most accurate method, DABM, required 6.6±0.6 minutes, compared to 2.8±0.2 minutes for Bspline, 1.5±0.2 minutes for Rigid, and 0.23±0.02 minutes for the Affine registration.

### 3.3.3 Evaluation of Deformation Vector Fields

The Jacobian determinant and Inverse Consistency Error were calculated within the 'masked' patent's volume i.e. within the external or 'skin' contour as the deformations present outside the patient's body are not of clinical significance. Figure 3.4a shows that both DABM and Bspline methods generated smooth, well-preserved vector fields without unphysical deformations – Jacobian determinants were positive only. Furthermore, the DABM algorithm is more flexible than Bspline as on average its $|J|$ deviates from 1 by a larger magnitude. Figure 3.4b demonstrates the Jacobian determinant maps for an example patient, showing that vector fields generated by the DABM registration contain deformations distributed more locally as opposed to the Bspline method in which

deformations cover larger anatomical regions. This feature was observed for all patients. The highest values of the Jacobian determinant (approximately 2) were found near the external contours of the posterior edge of the patients.

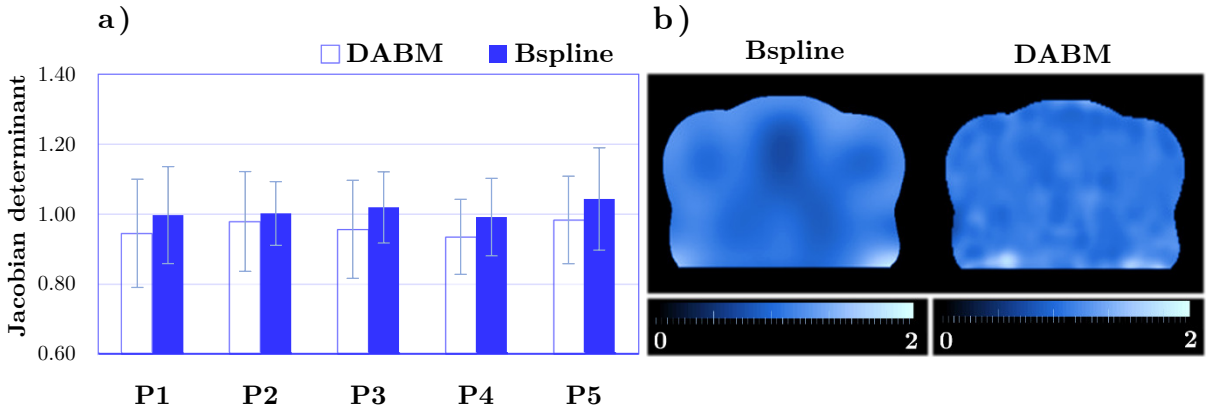



**Figure 3. 4. a)** Jacobian determinants of deformation vector fields generated by Dense Anatomical Block Matching and Bspline algorithms for five patients, averaged over all the voxels in their volume. Error bars are equal to 2 standard deviations; **b)** Examples of Jacobian determinant maps for DABM and Bspline registration methods from patient 2.

### 3.3.4 Inverse Consistency Error

*ICE based on forward and inverted forward transformations*

Figure 3.5 shows the volumetric distribution of the inverse consistency error for DABM and Bspline registrations, based on forward and inverted forward transformations. The most common values of $\text{ICE}^{(\text{inv})}$ are below 0.005 mm for all five patients and both algorithms while larger values create a tail ending at approximately 0.01 mm for DABM and Bspline. Table 3.3 indicates that mean $\text{ICE}^{(\text{inv})}$ remains in the range from 0.0017±0.0013 mm to 0.0029±0.0019 mm in the case of block matching registration and from 0.0018±0.019 mm to 0.0036±0.0354 mm for the Bspline algorithm. Median values of the inverse consistency error $\text{ICE}^{(\text{inv})}_{\text{median}}$ are reported in Table 3.3 as well and vary from 0.0014±0.0007 mm up to 0.0025±0.0011 mm for Dense Anatomical Block Matching

algorithm and from 0.0010±0.0005 mm to 0.0016±0.0007 mm for the Bspline method. The mean value of $ICE^{(inv)}$ over 5 patients was 0.0024±0.0019 mm for DABM and 0.0054±0.0480 mm for Bspline.

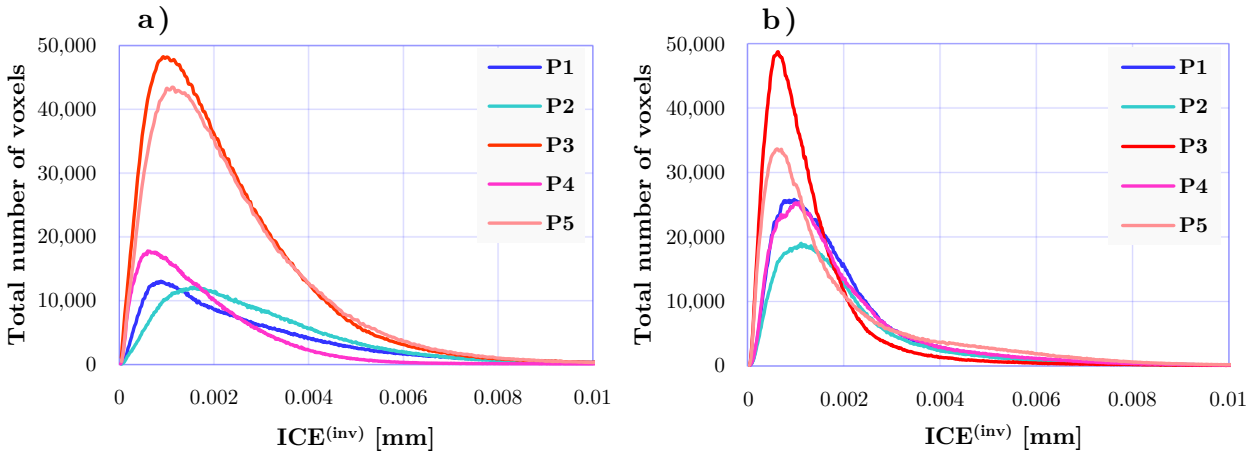



**Figure 3. 5. a)** Distribution of the inverse consistency error for forward vector fields generated by the Dense Anatomical Block Matching Algorithm and **b)** Bspline Algorithm as well as their inverse vector fields estimated by the Modified Newton Method for five patients. The total number of voxels refers to the voxels inside the entire patient's volume.

**Table 3. 3.** Mean and median values of inverse consistency error $ICE^{(inv)}$ for Dense Anatomical Block Matching and Bspline algorithms with standard deviations (SD) and median absolute deviation (MAD).

| **Patient No.** | **DABM** | | **Bspline** | |
|---|---|---|---|---|
| | $ICE^{(inv)}_{mean} \pm$ SD [mm] | $ICE^{(inv)}_{median} \pm$ MAD [mm] | $ICE^{(inv)}_{mean} \pm$ SD [mm] | $ICE^{(inv)}_{median} \pm$ MAD [mm] |
| 1 | 0.0027 ± 0.0022 | 0.0021 ± 0.0012 | 0.0022 ± 0.0131 | 0.0015 ± 0.0007 |
| 2 | 0.0029 ± 0.0019 | 0.0025 ± 0.0011 | 0.0027 ± 0.0181 | 0.0016 ± 0.0007 |
| 3 | 0.0024 ± 0.0024 | 0.0019 ± 0.0009 | 0.0018 ± 0.0190 | 0.0010 ± 0.0005 |
| 4 | 0.0017 ± 0.0013 | 0.0014 ± 0.0007 | 0.0036 ± 0.0354 | 0.0015 ± 0.0007 |
| 5 | 0.0024 ± 0.0017 | 0.0020 ± 0.0009 | 0.0035 ± 0.0346 | 0.0013 ± 0.0008 |
| **Average** | **0.0024 ± 0.0019** | **0.0020 ± 0.0010** | **0.0027 ± 0.0240** | **0.0014 ± 0.0007** |

*ICE based on forward and backward transformations*

Figure 3.6 shows the volumetric distribution of the inverse consistency error for DABM and Bspline registrations, based on forward and backward transformations. The most common values of $ICE^{(inv|b)}$ for all five patients are below 5 mm for DABM and below 8 mm for Bspline while larger values create a tail ending at approximately 10 mm for both algorithms. Table 3.4 indicates that mean $ICE^{(inv|b)}$ remains in the range from 1.75±1.73 mm to 2.46±2.27 mm in the case of block matching registration and from 2.58±1.81 mm to 3.95±1.84 mm for the Bspline algorithm. Median values of the inverse consistency error $ICE^{(inv|b)}_{median}$ are reported in Table 3.4 as well and vary from 1.01±0.67 mm up to 1.56±1.07 mm for Dense Anatomical Block Matching algorithm and from 2.07±0.92 mm to 3.71±1.25 mm for the Bspline method. The mean value of $ICE^{(inv|b)}$ over 5 patients was 2.01±2.03 mm for DABM and 3.25±1.68 mm for Bspline.

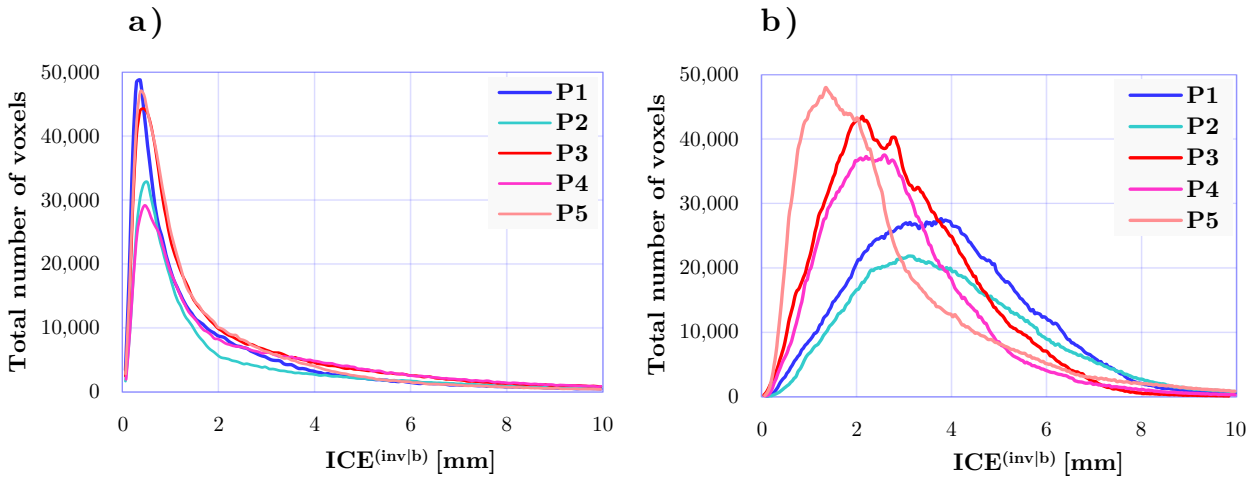



**Figure 3. 6. a)** Distribution of the inverse consistency error for forward vector fields generated by the Dense Anatomical Block Matching Algorithm and **b)** Bspline Algorithm as well as their backward transformation created by reverse (CBCT-CT) registration. The total number of voxels refers to the voxels inside the entire patient's volume.

**Table 3. 4.** Mean and median values of inverse consistency error $ICE^{(inv|b)}$ for Dense Anatomical Block Matching and Bspline algorithms with standard deviations (SD) and median absolute deviation (MAD).

| **Patient No.** | **DABM** | | **Bspline** | |
|---|---|---|---|---|
| | $ICE^{(inv\|b)}_{mean} \pm$ SD [mm] | $ICE^{(inv\|b)}_{median} \pm$ MAD [mm] | $ICE^{(inv\|b)}_{mean} \pm$ SD [mm] | $ICE^{(inv\|b)}_{median} \pm$ MAD [mm] |
| 1 | $1.76 \pm 1.88$ | $1.01 \pm 0.67$ | $3.84 \pm 1.73$ | $3.70 \pm 1.20$ |
| 2 | $2.01 \pm 2.13$ | $1.09 \pm 0.67$ | $3.95 \pm 1.84$ | $3.71 \pm 1.25$ |
| 3 | $2.09 \pm 2.12$ | $1.21 \pm 0.80$ | $2.96 \pm 1.50$ | $2.74 \pm 1.00$ |
| 4 | $2.46 \pm 2.27$ | $1.56 \pm 1.07$ | $2.89 \pm 1.51$ | $2.65 \pm 0.91$ |
| 5 | $1.75 \pm 1.73$ | $1.06 \pm 0.66$ | $2.58 \pm 1.81$ | $2.07 \pm 0.92$ |
| **Average** | $\mathbf{2.01 \pm 2.03}$ | $\mathbf{1.19 \pm 0.78}$ | $\mathbf{3.25 \pm 1.68}$ | $\mathbf{2.98 \pm 1.05}$ |

## 3.4 Discussion

This study showed that the Dense Anatomical Block Matching algorithm is more accurate in CT-CBCT registration of pelvic anatomy than Rigid, Affine, and Bspline methods. As mentioned in section 3.3.1 the results of registration accuracy were statistically significant considering all similarity metrics and the anatomical structures. This is consistent with the fact that the DABM algorithm outperformed the other examined registration methods in every metric and for every anatomical structure as presented in Figure 3.1. Additionally, Figure 3.2 reveals more details on the differences between the DABM method and other registration approaches and indicates that those differences were most significant for the bladder. This is likely due to the bladder having the highest contrast

relative to the surrounding tissues both on CT and CBCT images and therefore was less challenging to register for the DABM algorithm despite poor initial alignment after the Rigid registration. Moreover, since the bladder is the structure with the largest percentage differences relative to the other structures, it benefited the most from image registration compared to the other structures. However, the absolute values of COM, $HD_{avg,}$ and $HD_{95}$ metrics indicate relatively high accuracy for the GTV which is mostly due to the GTV being of lower volume than the other structures. Another reason why $HD_{avg}$ and $HD_{95}$ have lower (more desirable) values for GTV than for other structures might be associated with the fact that both the movements and the deformations of the GTV are very limited compared to the bladder and rectum. This highlights that, when analyzing results of image registration algorithms, one may want to consider both the percentage and absolute changes in the metrics.

Since the numerical metrics used here are commonly employed in image registration evaluation, our results can be compared to those in the literature. In the study of Bspline (ITK implementation) CT-CBCT registration performed on synthetic image data sets of the pelvic region, Varadhan et al. [17] obtained Dice Scores of 0.91, 0.95, and 0.89 as well as average surface distances of 1.03, 0.42, and 0.8 mm for GTV, bladder, and rectum respectively. For the same structures, Woerner et al.[53] reported DSC of approximately 0.8 and mean surface distance slightly above 2 mm, based on deformed (CT-CBCT Bspline registration using Velocity Advanced Imaging 2.8.1; Varian Medical Systems, Palo Alto, California) and reference contours of five prostate cancer patients. In addition, similar to our findings, the authors also found that image registration was most beneficial for the bladder. For the remaining structures, the limited benefit was reported. In the same study,

Woerner et al.[53] also explored the inter-observer variability and found that RO1 (Radiation Oncologist 1) to RO2 differences measured as mean values of Dice Score and Hausdorff distances for bladder, rectum, GTV and seminal vesicles were 0.90±0.06 and 0.99±0.44 mm. In our study, mean values of Dice Score and Hausdorff distances for the DABM algorithm in the case of bladder, rectum, and GTV(without seminal vesicles) were 0.85±0.05 and 1.94±0.34 mm which approximately fall in the range of RO1 to RO2 differences and shows the potential clinical significance of DABM.

A desirable feature of image registration algorithms is robustness in terms of consistent performance. Since the result of the image, matching can be highly dependent on the magnitude of deformations of the patient's anatomy in the target image, which varies from patient to patient, there is a strong need to establish confidence in the DIR outputs. As presented in Table 3.2 the performance of the DABM algorithm is more consistent compared to the other registration methods - standard error for the block matching approach is, in the majority of cases, smaller than for the other algorithms by 11% to 42%. A definitive reason for the improved consistency of the DABM registration has not been identified within this work however it is most probably due to an outlier rejection process that was implemented in the block matching algorithm by Garcia et al.[31]. The authors explain that this correction feature reduces the error associated with voxel pairing (in homogenous regions) estimated by DABM through comparing original pairings available in a sparse displacement field to those in interpolated displacements. The other image registration algorithms that we have evaluated within this study lack any similar error reduction features.

Figure 3.2, apart from providing details on the performance of examined registration algorithms, indicates that in the case of the bladder (three out of five patients) and the GTV (two out of five patients) Rigid registration is more accurate than the Bspline algorithm. We believe that this is due to the specific anatomy of those patients as well as the flexibility of the Bspline method. We have noticed that the deformation of the bladder was determined mainly by the deformations applied to the other anatomical structures and the external body contour, especially the front of the abdomen. In other words, Bspline was not able to deform the bladder as independently as expected. That is the result of the insufficient flexibility of the algorithm in that region of the patient's body. Consequently, the registration accuracy of the GTV located in close proximity to the bladder was affected as well in a similar fashion. Despite this observation, we decided to keep the regularization term of the Bspline unchanged to limit the unphysical deformations in the other anatomy regions. This superiority of the simple Rigid registration over a more complex deformable image registration algorithm might be surprising but is not uncommon in the literature. Kirby et al. [54] evaluated 11 deformable image registration algorithms using deformable phantom of pelvic anatomy and reported that the DIR algorithm available in commercial software (MIM Software Inc, Cleveland, USA) is less accurate than a Rigid registration over most of the range of calculated error values. Those results are even more pronounced considering that both the floating and target images were CT scans. Also, Li et al.[55] in a study including 21 Head and Neck patients and 10 intensity-based DIR algorithms found that in soft-tissue regions, deformable registrations (of CT-CBCT) performed comparable or worse than the Rigid method. In particular, for the submandibular gland authors mentioned that Rigid registration outperformed all the DIRs examined.

Physician evaluation confirmed that the DABM algorithm outperformed other registration methods for accuracy. Figure 3.3 shows that the accuracy of registration steadily decreases as the complexity of the algorithm decreases, which is consistent with the results for the metrics presented in Figure 3.1. Also, Figure 3.2 indicates that the most challenging structure to register was the rectum followed by bladder and GTV, likely due to the relatively large deformations and complex shape of the rectum compared to the bladder and GTV. However, the correspondence between values of the metrics and physician assessment should be treated with caution. Thor et al.[56], in a study of five prostate cancer patients, obtained a positive correlation between DSC and physician score for bladder only – they observed that the Dice coefficient decreased with radiation oncologist scoring. For GTVand rectum, no such correlation was found. Hardcastle et al.[57], based on the CT-CBCT image registrations of 17 lung cancer patients, reported weak correspondence between the Dice coefficient and mean Hausdorff distances and physician scores for compared structures. The inconsistency in the literature, as well as results obtained within this study (in our case the quantitative and physician's evaluation were consistent), suggest that the quantitative assessment of DIR accuracy should be followed by a careful evaluation of radiation oncologists. This is especially important in the case where DIR is utilized as a tool, for example in contour propagation or dose accumulation, as image registration errors have a direct impact on those procedures as shown by Veiga et al.[12] and Kumarasiri et al.[19].

Figure 3.3 shows that the time the physician needed to correct contours deformed by the DABM algorithm (5.3 min.) was relatively close to the time that was required to segment structures from scratch (7.0 min.). The reason for this is that when a new structure is

created in the treatment planning system, the physician uses semi-automated segmentation tools such as auto-filling, thresholding brush, and interpolating, which eliminate the need to contour a structure on every slice. Regions of interest can be delineated on every other slice and then interpolated over all the slices. Therefore, although DABM outperforms other algorithms in terms of physician correction time it can as time efficient to contour the GTV from scratch.

The analysis of the intra-physician variability was used to establish margins for clinically acceptable image registration results when it comes to the contour propagation. As shown in Figure 3.1 the DABM algorithm was the only algorithm that was able to deform contours so that they fall within the range of intra-physician variability. For GTV both the DSC and $HD_{avg}$ metrics indicate that block matching algorithm-generated even more satisfying results compared to those resulting from physician correction relative to the reference contours. The difficulties in correcting GTV contours by a physician are not surprising and were a consequence of very poor contrast in that region on the CBCT images.

Numerical results and related discussion of deformable image registration evaluation can often provide very useful information about the accuracy of examined algorithms however for better understanding of those results especially in terms of clinical practice the visual inspection of deformed anatomy is considered very valuable as well. Figure 3.7 shows the deformed anatomy with segmented structures in three perpendicular planes for all the registration algorithms evaluated within this study. The difference between contours that were manually delineated by the expert physician and deformed contours is the smallest

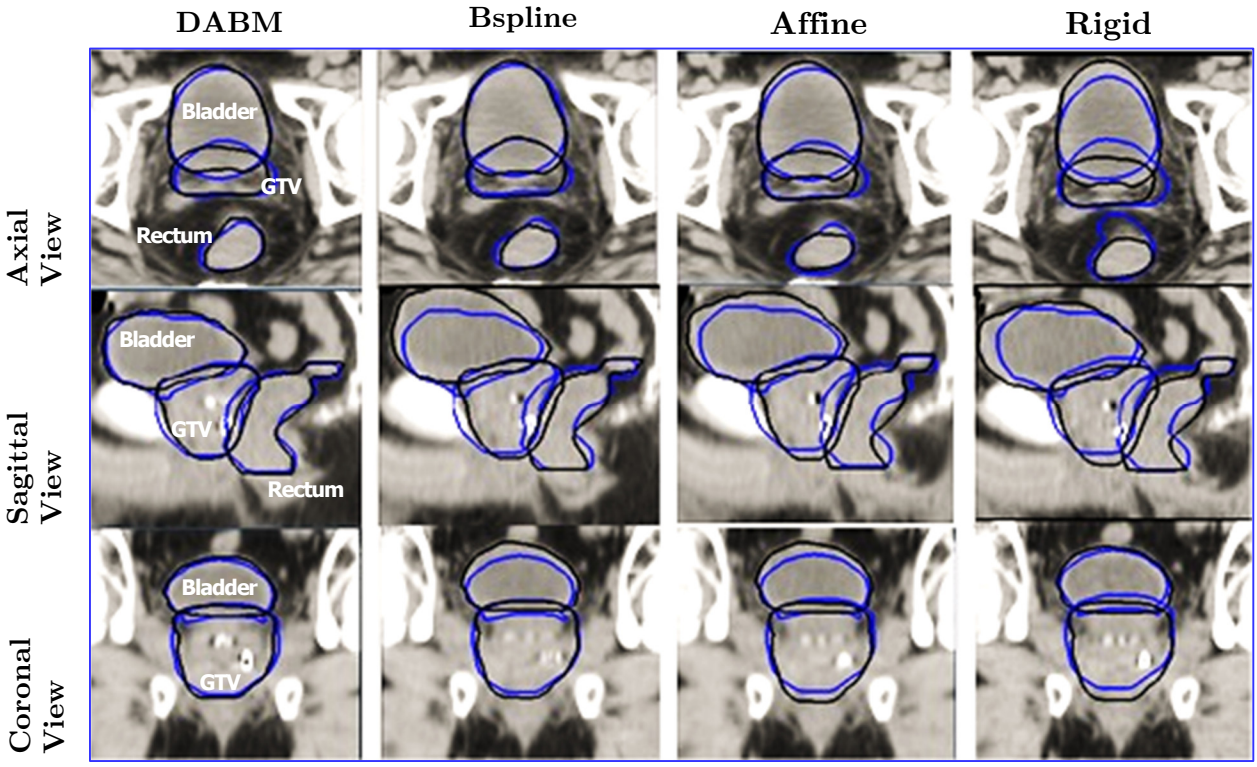



**Figure 3. 7.** Deformed anatomy in three perpendicular planes for all the registration algorithms evaluated within this study. For the purpose of this figure the anatomy of the second patient was used as it best shows the potential of DABM registration to satisfactory align two images in the case of challenging anatomical deformations. Black contours are deformed by the particular registration algorithm and match displayed anatomy. Contours manually segment by the experienced physician on the fixed image are represented by blue contours. In the ideal scenario, the black contours would perfectly match the blue contours.

for DABM registration in the case of all the structures. This quick visual evaluation agrees with all the numerical results. Deformation vector fields that resulted from Bspline and DABM registrations were well-preserved and did not contain any unphysical deformations. Mean values of the Jacobian determinants were in the range from 0.8 to 1.2 and are commonly encountered in the literature for different clinical sites and registration algorithms. For instance, Veiga et al.[12] in their study of dose warping

uncertainties due to registration algorithm performed on five Head and Neck patients obtained Jacobian determinants ranging from 0.53 to 1.57 for the Bspline Free Form Deformation algorithm. Similarly, Kim et al.[58] reported mean Jacobian determinants for bladder, rectum, and GTVin the range of 0.7 to 2.2 utilizing the Bspline algorithm for CT-CBCT image registration based on three prostate cancer patients. As mentioned in section 3.3.3, we found that the largest values of $|J|$ were near the external patient contour at the posterior edge. The reason for that is the proximity of the therapeutic table that had much higher contrast on the CBCT image than the diagnostic table in the pCT image. This contrast difference made it very difficult for the registration algorithms to preserve the deformed image geometry in that region – we observed that the diagnostic table in the deformed image was subjected to non-Rigid deformations despite it being a Rigid object. Although this observation is a registration error, its effect on, for example, dose mapping would most likely be insignificant because the posterior edge of the patients is usually an area of low dose in the radiotherapy of prostate cancer. While cropping the table out seems like a reasonable solution, that approach causes the appearance of image artifacts near the edge of the image after registering CT to CBCT and was not done here.

The inverse consistency error, $\mathrm{ICE}^{(\mathrm{inv})}$, that was calculated using the Modified Newton Method for vector fields generated through the DABM algorithm showed submillimeter values for all the patients studied here ($\mathrm{ICE}^{(\mathrm{inv})}_{\mathrm{mean}} = 0.0024 \pm 0.0019$ mm; Table 3.3). Compared to the available literature these are considered very low (and desirable) values. For instance, Bender et al.[59] based on MVCT scans of head and neck region, reduced the ICE from approximately 0.9 mm to 0.08 mm using in-house developed ICE reduction equation for Fast Symmetric Demons registration which led to the reduction in mean

accumulated dose differences (reference dose vs. mapped dose) to the right and left parotid glands from 0.8 Gy and 0.9 Gy to 0.2 Gy and 0.3 Gy, respectively. Also, Veiga et al.[12] investigated the uncertainties in dose warping due to the choice of the deformable image registration and found that the most suitable algorithm for dose accumulation is the one with the lowest ICE (which in their work was 0.008±0.003mm). The maximum values of the $ICE^{(inv)}$ volume distributions for all five patients studied here are observed to be approximately the same (Figure 3.5a). The most probable reason for that consistency is the fact that the accuracy of the DABM algorithm is based on the forward transformation which was consistent for all five patients. In addition, the Jacobian determinants for the five patients (Figure 3.4) do not differ significantly from each other. Thus, it may be concluded that consistent algorithmic performance including the characteristics of forward transformations, might lead to the consistent ICE that is associated with them. For Bspline registration $ICE^{(inv)}$ may seem to be smaller than for DABM especially when observing Figure 3.5b. The volumetric distribution of inverse consistency error is clearly shifted toward lower values for the Bspline algorithm compared to the DABM. It can be noted however that the standard deviation of this distribution is higher by around one order of magnitude relative to its mean value and compared to the standard deviation calculated for the DABM algorithm. The reason for this is most probably poorer accuracy of Bspline registration as well as larger standard error for all the evaluation metrics as reported in the results section.

The inverse consistency error, $ICE^{(inv|b)}$, obtained by the application of reverse (CBCT-CT) registration was significantly larger than $ICE^{(inv)}$ (few orders of magnitude). The main reason for this discrepancy is the robustness of the Modified Newton Method of inversion

resulting in a smaller value of $ICE^{(inv)}$ and the absence of an inverse consistency penalty term in the objective function of the DABM and Bspline registration algorithms, leading to a large value of $ICE^{(inv|b)}$ as shown in Figure 3.6. Additionally, we made a significant effort in tuning the parameters of those two registration methods to maximize the accuracy of CT-CBCT registrations. At the same time, however, to obtain $T^{-1}$ for the calculation of $ICE^{(inv|b)}$, which required performing CBCT-CT registration, the same parameters were used (necessary to compute a valid ICE). Visual inspection of the CBCT-CT registration results showed that the parameters we found optimal for CT-CBCT registration were not optimal for the CBCT-CT registration. The alignment between deformed CBCT and planning CT were far less satisfactory than the alignment between the deformed planning CT that was mapped back to the original planning CT and the original planning CT (which were nearly indistinguishable by visual inspection). Therefore, we believe that in applications that require symmetric image registration the inverse consistency penalty term should be included in the objective function of a given DIR algorithm. Also, as indicated in Table 3.4, $ICE^{(inv|b)}$ for DABM is smaller than for Bspline for all patients by 38% (1.24 mm) on average. This is most likely due to the diffeomorphic nature of the DABM algorithm. Diffeomorphic registration prevents folding and tearing deformation and maintains the smoothness of the transformation[60]. The Bspline algorithm evaluated in this paper was non-diffeomorphic.

The most clinically relevant limitation of the Dense Anatomical Block Matching algorithm lays in its relatively long computational time. Nearly seven minutes for a single CT-CBCT registration might be acceptable for off-line treatment plan adaptation but would be unsuitable for on-line modifications. Nevertheless, this difficulty might be overcome by

simply increasing the computational power of the workstation or parallelization of the DABM code. As for the physician evaluation of the tested image registration algorithms, our study is limited to utilizing only one radiation oncologist. Future work may expand the number of radiation oncologists involved to investigate inter-physician contouring variability.

Another limitation of our work is the limited number of patients. We realize that performing the study on a larger group of patients would improve the reliability of the study and we do intend to further explore the DABM algorithm on an increased number of patients. At the same time, we intended to share our results using the data that we currently have available because we have noticed a very consistent and satisfactory performance of DABM and consider our study as a very promising step for further work. For that reason, we believe that although the number of patients was limited, the study will bring value to and spark an interest among researchers who are searching for a registration algorithm that would be a great candidate for the challenging CT-CBCT registration in the pelvic region. Additionally, in reviewing the relevant literature in the field of deformable image registration we found numerous studies that use a similarly limited number of patients for both the evaluation of DIR algorithms[61-63] and their clinical application[56,64,65], to introduce a new tool or concept.

## 3.5 Conclusion

In this study, four image registration algorithms – Rigid, Affine, Bspline, and Dense Anatomical Block Matching, were evaluated on CT and CBCT volumetric data sets for five prostate cancer patients. Our work demonstrated that DABM outperformed other

registration methods that were evaluated within this study in terms of accuracy and consistency while generating vector fields with well-preserved topology and small inverse consistency errors. Comparison to the available literature shows that these evaluation results for Dense Anatomical Block Matching algorithm are generally at a clinically acceptable level. This makes DABM a very promising alternative to the existing registration methods when it comes to challenging CT-CBCT deformable image registration and its application in adaptive radiation therapy of the pelvic region.

## 3.6 References

1. Sotiras A, Davatzikos C, Paragios N. Deformable medical image registration: A survey. *IEEE transactions on medical imaging.* 2013;32(7):1153-1190.

2. Jingu K. Use of deformable image registration for radiotherapy applications. *J Radiol Radiat Ther.* 2014;2(2):1042-1050.

3. Li X, Zhang YY, Shi YH, Zhou LH, Zhen X. Evaluation of deformable image registration for contour propagation between CT and cone-beam CT images in adaptive head and neck radiotherapy. *Technology and health care : official journal of the European Society for Engineering and Medicine.* 2016;24 Suppl 2:S747-755.

4. Lee DS, Woo JY, Kim JW, Seong J. Re-Irradiation of Hepatocellular Carcinoma: Clinical Applicability of Deformable Image Registration. *Yonsei medical journal.* 2016;57(1):41-49.

5. Dean JA, Welsh LC, McQuaid D, et al. Assessment of fully-automated atlas-based segmentation of novel oral mucosal surface organ-at-risk. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2016;119(1):166-171.

6. Chen S, Quan H, Qin A, Yee S, Yan D. MR image-based synthetic CT for IMRT prostate treatment planning and CBCT image-guided localization. *Journal of Applied Clinical Medical Physics.* 2016;17(3).

7. Tahir BA, Swift AJ, Marshall H, et al. A method for quantitative analysis of regional lung ventilation using deformable image registration of CT and hybrid hyperpolarized gas/1H MRI. *Phys Med Biol.* 2014;59(23):7267-7277.

8. Simon A, Nassef M, Rigaud B, et al. Roles of Deformable Image Registration in adaptive RT: From Contour propagation to dose monitoring. Paper presented at: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)2015.

9. Alongi F, Cozzi L, Arcangeli S, et al. Linac based SBRT for prostate cancer in 5 fractions with VMAT and flattening filter free beams: preliminary report of a phase II study. 2013;8(1):171.

10. Yu G, Liang Y, Yang G, et al. Accelerated gradient-based free form deformable registration for online adaptive radiotherapy. *Phys Med Biol.* 2015;60(7):2765-2783.

11. Veiga C, McClelland J, Moinuddin S, et al. Toward adaptive radiotherapy for head and neck patients: Feasibility study on using CT-to-CBCT deformable registration for "dose of the day" calculations. *Med Phys.* 2014;41(3):031703.

12. Veiga C, Lourenço AM, Mouinuddin S, et al. Toward adaptive radiotherapy for head and neck patients: Uncertainties in dose warping due to the choice of deformable registration algorithm. *Medical physics.* 2015;42(2):760-769.

13. Berenguer R, de la Vara V, Lopez-Honrubia V, et al. The influence of the image registration method on the adaptive radiotherapy. A proof of the principle in a selected case of prostate IMRT. *Physica Medica.* 2018;45:93-98.

14. Oh S, Kim S. Deformable image registration in radiation therapy. *Radiation oncology journal.* 2017;35(2):101.

15. Schnabel JA, Heinrich MP, Papiez BW, Brady SJ. Advances and challenges in deformable image registration: From image fusion to complex motion modelling. *Med Image Anal.* 2016. doi: 10.1016/j.media.2016.06.031.

16. Zukauskaite R, Brink C, Hansen CR, et al. Open source deformable image registration system for treatment planning and recurrence CT scans : Validation in the head and neck region. *Strahlentherapie und Onkologie : Organ der Deutschen Rontgengesellschaft [et al].* 2016;192(8):545-551.

17. Varadhan R, Karangelis G, Krishnan K, Hui S. A framework for deformable image registration validation in radiotherapy clinical applications. *Journal of applied clinical medical physics/American College of Medical Physics.* 2013;14(1):4066.

18. Han SC, Lee SS, Kim M-S, et al. Evaluation of various deformable image registrations for point and volume variations. *Journal of the Korean Physical Society.* 2015;67(1):218-223.

19. Kumarasiri A, Siddiqui F, Liu C, et al. Deformable image registration based automatic CT-to-CT contour propagation for head and neck adaptive radiotherapy in the routine clinical setting. *Med Phys.* 2014;41(12):121712.

20. Nielsen MS, Ostergaard LR, Carl J. A new method to validate thoracic CT-CT deformable image registration using auto-segmented 3D anatomical landmarks. *Acta Oncol.* 2015;54(9):1515-1520.

21. Srinivasan K, Mohammadi M, Shepherd J. Cone beam computed tomography for adaptive radiotherapy treatment planning. *Journal of Medical and Biological Engineering.* 2014;34(4):377-385.

22. Lou Y, Niu T, Jia X, Vela PA, Zhu L, Tannenbaum AR. Joint CT/CBCT deformable registration and CBCT enhancement for cancer radiotherapy. *Med Image Anal.* 2013;17(3):387-400.

23. Thing RS, Bernchou U, Mainegra-Hing E, Brink C. Patient-specific scatter correction in clinical cone beam computed tomography imaging made possible by the combination of Monte Carlo simulations and a ray tracing algorithm. *Acta Oncol.* 2013;52(7):1477-1483.

24. Watson PG, Mainegra-Hing E, Tomic N, Seuntjens J. Implementation of an efficient Monte Carlo calculation for CBCT scatter correction: phantom study. *Journal of Applied Clinical Medical Physics.* 2015;16(4).

25. Zhang Z, Han X, Pearson E, Pelizzari C, Sidky EY, Pan X. Artifact reduction in short-scan CBCT by use of optimization-based reconstruction. *Physics in medicine and biology.* 2016;61(9):3387.

26. Zhen X, Gu X, Yan H, Zhou L, Jia X, Jiang SB. CT to cone-beam CT deformable registration with simultaneous intensity correction. *Physics in medicine and biology.* 2012;57(21):6807.

27. Park S, Robinson A, Quon H, et al. Accurate tracking of tumor volume change during radiotherapy by CT-CBCT registration with intensity correction. Paper presented at: SPIE Medical Imaging2016.

28. Ganesh Narayanasamy GN, Jonathan Feddock JF, John Gleason JG. CBCT-based dosimetric verification and alternate planning techniques to reduce the normal tissue dose in SBRT of lung patients. *International Journal of Cancer Therapy and Oncology.* 2015;3(2):3218.

29. Berlinger K, Knorr S, Schlossbauer C, Rehs J. CBCT and X-Ray Combined Setup with X-Ray Verification of Patient Positioning. In.: US Patent 20,150,335,915; 2015.

30. Mali SB. Adaptive Radiotherapy for Head Neck Cancer. *Journal of Maxillofacial and Oral Surgery.* 2016. doi: 10.1007/s12663-016-0881-y.

31. Garcia V, Commowick O, Malandain G. A robust and efficient block matching framework for non linear registration of thoracic CT images. Paper presented at: Grand Challenges in Medical Image Analysis (MICCAI workshop)2010.

32. Commowick O, Wiest-Daesslé N, Prima S. Automated diffeomorphic registration of anatomical structures with rigid parts: Application to dynamic cervical MRI. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention2012.

33. Huger S, Graff P, Harter V, et al. Evaluation of the Block Matching deformable registration algorithm in the field of head-and-neck adaptive radiotherapy. *Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics.* 2014;30(3):301-308.

34. Hoffmann C, Krause S, Stoiber EM, et al. Accuracy quantification of a deformable image registration tool applied in a clinical setting. *Journal of applied clinical medical physics.* 2014;15(1):237-245.

35. Zambrano V, Furtado H, Fabri D, et al. Performance validation of deformable image registration in the pelvic region. *Journal of radiation research.* 2013;54 Suppl 1:i120-128.

36. Ibanez L, Schroeder W, Ng L, Cates J. The ITK software guide. 2005.

37. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE transactions on medical imaging.* 1999;18(8):712-721.

38. Klein S, Pluim JPW, Staring M, Viergever MA. Adaptive Stochastic Gradient Descent Optimisation for Image Registration. *International Journal of Computer Vision.* 2008;81(3):227-239.

39. Oliveira FP, Tavares JMR. Medical image registration: a review. *Computer methods in biomechanics and biomedical engineering.* 2014;17(2):73-93.

40. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging.* 2010;29(1):196-205.

41. Powell MJ. The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge.* 2009.

42. Toussaint N, Souplet J-C, Fillard P. MedINRIA: medical image navigation and research tool by INRIA. Paper presented at: Proc. of MICCAI2007.

43. Johnson HJ, McCormick MM, Ibanez L. The ITK Software Guide Book 1: Introduction and Development Guidelines Fourth Edition Updated for ITK version 4.7. *Kitware, Inc(January 2015).* 2015.

44. Gobbi DG, Peters TM. Generalized 3D nonlinear transformations for medical imaging: an object-oriented implementation in VTK. *Computerized Medical Imaging and Graphics.* 2003;27(4):255-265.

45. Tustison N, Gee J. Introducing Dice, Jaccard, and other label overlap measures to ITK. *Insight J.* 2009.1-4.

46. Pinter C, Lasso A, Wang A, Jaffray D, Fichtinger G. SlicerRT: Radiation therapy research toolkit for 3D Slicer. *Medical physics.* 2012;39(10):6332-6338.

47. http://plastimatch.org. Hausdorff_distance Class Reference (May, 2016). Retrived from http://plastimatch.org/doxygen/classHausdorff___distance.html#details. 2016.

48. Hardcastle N, Tome WA, Cannon DM, et al. A multi-institution evaluation of deformable image registration algorithms for automatic organ delineation in adaptive head and neck radiotherapy. *Radiation oncology.* 2012;7:90.

49. Ahmad S, Khan MF. Topology preserving non-rigid image registration using time-varying elasticity model for MRI brain volumes. *Computers in biology and medicine.* 2015;67:21-28.

50. Papiez B, Matuszewski B. Symmetric image registration with directly calculated inverse deformation field. *Annals of the BMVA.* 2012;2012(6):1-14.

51. Vercauteren T, De Gersem W, Olteanu LA, et al. Deformation field validation and inversion applied to adaptive radiation therapy. *Physics in medicine and biology.* 2013;58(15):5269.

52. Modat M, Cardoso MJ, Daga P, Cash D, Fox NC, Ourselin S. Inverse-consistent symmetric free form deformation. Paper presented at: International Workshop on Biomedical Image Registration2012.

53. Woerner AJ, Choi M, Harkenrider MM, Roeske JC, Surucu M. Evaluation of Deformable Image Registration-Based Contour Propagation From Planning CT to Cone-Beam CT. *Technology in Cancer Research & Treatment.* 2017.1533034617697242.

54. Kirby N, Chuang C, Ueda U, Pouliot J. The need for application-based adaptation of deformable image registration. *Medical Physics.* 2013;40(1):011702.

55. Li X, Zhang Y, Shi Y, et al. Comprehensive evaluation of ten deformable image registration algorithms for contour propagation between CT and cone-beam CT images in adaptive head & neck radiotherapy. *PloS one.* 2017;12(4):e0175906.

56. Thor M, Petersen JB, Bentzen L, Hoyer M, Muren LP. Deformable image registration for contour propagation from CT to cone-beam CT scans in radiotherapy of prostate cancer. *Acta Oncol.* 2011;50(6):918-925.

57. Hardcastle N, Van Elmpt W, De Ruysscher D, Bzdusek K, Tomé WA. Accuracy of deformable image registration for contour propagation in adaptive lung radiotherapy. *Radiation oncology.* 2013;8(1):1.

58. Kim J, Kumar S, Liu C, et al. A novel approach for establishing benchmark CBCT/CT deformable image registrations in prostate cancer radiotherapy. *Physics in medicine and biology.* 2013;58(22):8077.

59. Bender ET, Hardcastle N, Tome WA. On the dosimetric effect and reduction of inverse consistency and transitivity errors in deformable image registration for dose accumulation. *Medical physics.* 2012;39(1):272-280.

60. Tao G, He R, Datta S, Narayana PA. Symmetric inverse consistent nonlinear registration driven by mutual information. *Computer methods and programs in biomedicine.* 2009;95(2):105-115.

61. Katsuta Y, Kadoya N, Fujita Y, et al. Assessment of a commercially available automatic deformable image registration. Paper presented at: World Congress on Medical Physics and Biomedical Engineering May 26-31, 2012, Beijing, China2013.

62. Kadoya N, Fujita Y, Katsuta Y, et al. Evaluation of various deformable image registration algorithms for thoracic images. *Journal of radiation research.* 2013;55(1):175-182.

63. Wu Z, Rietzel E, Boldea V, Sarrut D, Sharp GC. Evaluation of deformable registration of patient lung 4DCT with subanatomical region segmentations. *Medical physics.* 2008;35(2):775-781.

64. Han EY, Chao M, Zhang X, Penagaricano J. Feasibility study on deformable image registration for lung SBRT patients for dose-driven adaptive therapy. *International Journal of Medical Physics, Clinical Engineering and Radiation Oncology.* 2015;4(03):224.

65. Abe T, Tamaki T, Makino S, et al. Assessing cumulative dose distributions in combined radiotherapy for cervical cancer using deformable image registration with pre-imaging preparations. *Radiation oncology.* 2014;9:293.

# Chapter 4

# Adaptive radiation therapy strategies in the treatment of prostate cancer patients using hypofractionated VMAT

This chapter focuses on the research work in which eight adaptive radiation therapy strategies were comprehensively evaluated using several dose and dose-volume metrics. The study included 20 prostate cancer patients who underwent hypofractionated VMAT treatment. The results demonstrated that daily on-line adaptation approaches (i.e. every fraction) were the most impactful, although strategies adapting every other fraction also were impactful while reducing relative workload as well. Offline treatment adaptations were shown to be less beneficial due to increased dose delivered to bladder and rectum compared to other ART strategies. This chapter has been accepted for publication in the peer-reviewed Journal of Applied Clinical Medical Physics.

## 4.1 Introduction

Radiation therapy is a major treatment option for patients with prostate cancer. However, variations in the patients' anatomy during radiotherapy treatment can present challenges.

Numerous studies show that the variable location of the prostate, bladder filling, and pockets of gas often present in the rectum might significantly compromise dose coverage of the target structure and increase the dose delivered to critical organs[1-5].

Hypo-fractionated radiotherapy, delivering fewer, higher fraction doses increases the dosimetric impact of anatomic variability compared to conventional fractionation schemes. For prostate cancer patients, moderate hypofractionated (70 Gy in 28 fractions, 2.5 Gy/fraction) IMRT and VMAT were proven to keep early normal tissue toxicity at acceptable levels[6,7]. More aggressive hypofractionation delivering 33.5-37.5 Gy in five fractions has also been shown to achieve acceptable toxicity and quality of life[9,10], even for high risk and very high risk (including node-positive) prostate cancer patients.

The safe and effective delivery of high radiation doses in hypofractionated schemes requires a high level of precision, but inter- and intra-fractional patient anatomical variation is present and known to compromise dosimetric aspects of the treatment[8,9]. Adaptive radiation therapy (ART) strategies, in particular on-line ART, have the ability to account for systematic anatomic changes of prostate swelling as well as random anatomic changes such as inter- and intra-fraction bladder and rectal filling, in addition to independent movement and deformation of multiple targets[8,10]. The necessity and the benefits of ART application in SBRT prostate treatments have been shown in other recent studies[9-12]. It should also be noted that existing IGRT (Image Guided Radiation Therapy) techniques, although allowing for prostate motion management, have some limitations. For example, they might require an invasive procedure carrying the risk of bleeding, infection and discomfort for the patient (radiopaque intraprostatic fiducial markers method). Another prostate IGRT example is a technique that utilizes inserted

electromagnetic transponders. In this case, patient eligibility criteria are very strict as only patients without hip prosthesis, metal implants, peacemaker, or other electromagnetic devices are eligible, as well as relatively thinner patients due to the maximum range of the beacon detection by a required external array (reading) device[13].

Adaptive Radiation Therapy is currently an active area of research and there are still many novel ART approaches that have not been explored yet but could make a significant contribution to the field. The current study focuses on a comprehensive evaluation of several ART methods that have not been explored for the prostate VMAT hypofractionation schemes examined here. The purpose of this research was to retrospectively investigate eight adaptive radiation therapy strategies (including both online and offline scenarios) for hypofractionated VMAT treatments based on imaging and treatment plan data of 20 prostate cancer patients with the application of deformable image registration. The online and offline adaptations considered were compared to the *non-ART* (not adapted) delivery scenario.

## 4.2 Material and Methods

### 4.2.1 Patients' Data

The imaging and treatment planning data for twenty prostate cancer patients, with an average age of 77 ($\pm$7) years were retrospectively used for this study. The study was approved by the local research ethics board (University of Manitoba).

All twenty patients had previously received a 40Gy/5 fraction treatment regimen and were treated at CancerCare Manitoba. One pre-treatment, planning Computed

Tomography (pCT) imaging scan, and five sets of on-treatment Cone Beam CT (CBCT) imaging scans were obtained for each patient. CBCT images were acquired during each treatment fraction right before radiation delivery to ensure proper patient positioning. Anatomic structures considered in the dosimetric analysis included the clinical target volume (CTV), the planning target volume (PTV=CTV+0.5cm margin) as well as organs-at-risk (OARs) - bladder, rectum, and femoral heads. An experienced radiation oncologist segmented these structures on both CT and CBCT imaging datasets. These structures were also used for plan adaptation and optimization purposes.

The pCT images, at 512x512 pixels, were obtained with a spatial resolution of 1.17 mm x 1.17 mm per pixel and 3.0 mm slice thickness (total of ~210 slices) on a Philips Brilliance Big Bore CT scanner. The CBCT images, at 384x384 pixels, were obtained with a spatial resolution of 1.17 mm x 1.17 mm per pixel and 2.5 mm slice thickness (total of 64 slices) using an OBI Cone-Beam CT unit (Varian Medical Systems, Palo Alto, CA).

### 4.2.2 Dose Delivery & Treatment Planning

The treatment was delivered using volumetric modulated arc therapy (VMAT) with two full arcs. Every patient was treated with a full bladder and empty rectum as per local clinical protocol. The intent of the radiation therapy was curative for all patients. Treatment plans were created in the External Beam Planning module of the Eclipse treatment planning system, version 13.6 (Varian Medical System, Palo Alto, CA). Dose and dose-volume objectives for the radiation treatment are summarized in Table 4.1. The beam energy and the maximum dose rate for both arcs were 6MV and 1000MU/min (SRS mode). All the *non-ART* plans were normalized to the dose received by 95% of the PTV

**Table 4.1.** Treatment Planning Objectives for Target and OAR structures.

| Structure | Prescription | | | Fraction Dose [cGy] | Total Dose [cGy] | Planning Objective Symbols* |
|---|---|---|---|---|---|---|
| **CTV** | At least | 99.0 | % receives more than | 800.0 | 4000.0 | $V_{40Gy} > 99\%$ |
| **PTV** | At least | 99.0 | % receives more than | 760.0 | 3800.0 | $V_{38Gy} > 99\%$ |
| **PTV** | Maximum Dose | | is | 856.0 | 4280.0 | $D_{max} < 4280$ cGy |
| **Rectum** | At most | 15.0 | % receives more than | 640.0 | 3200.0 | $V_{32Gy} < 15\%$ |
| **Rectum** | At most | 20.0 | % receives more than | 560.0 | 2800.0 | $V_{28Gy} < 20\%$ |
| **Bladder** | At most | 15.0 | % receives more than | 640.0 | 3200.0 | $V_{32Gy} < 15\%$ |
| **Bladder** | At most | 20.0 | % receives more than | 560.0 | 2800.0 | $V_{28Gy} < 20\%$ |
| **Femur-RT** | At most | 5.0 | % receives more than | 560.0 | 2800.0 | N/A |
| **Femur-LT** | At most | 5.0 | % receives more than | 560.0 | 2800.0 | N/A |

***the percentage deviations from these objectives for all ART and *non-ART* plans have been illustrated in the Figure 4.6b for CTV and PTV as well as in the Figure 4.7b for Bladder and Rectum. Femoral Heads were not included in the plan evaluation as the radiation doses for all plans were significantly below the planning objective thresholds.

volume. Specifically, the dose was determined based on the 95% of the PTV volume on the dose volume histogram of the original treatment plan. The *non-ART* plans were not normalized.

### 4.2.3 Adaptive Radiation Therapy

#### Deformable Image Registration

Daily CBCT images were acquired before the delivery of every fraction. Planning CT images were then registered to CBCT data sets using a Bspline[14] based automated deformable image registration (DIR) algorithm available in Velocity AI, version 3.2 (Varian Medical System, Palo Alto, CA). Deformed pCT images ('dCT') were then used

for daily dose estimation with respect to the treatment plan (adapted and non-adapted) that was delivered to the patient during a particular fraction. Since our study was retrospective, the plan delivery was simulated (i.e., it was not actually delivered to the patient). Contours that allowed for the estimation of the dose delivered to the considered anatomical structures were delineated by an experienced physician on CBCT and then propagated to daily dCT image scans. To adapt to the current patient anatomy, and as needed for some of the strategies studied here including adaption while accounting for the dose delivered in the previous fraction(s), the subsequent fraction treatment plans were re-optimized simulating offline or/and online scenarios.

**Plan Optimization**

For the optimization of VMAT plans in this study, the Progressive Resolution Optimizer[15], version 10.0.28 was used, while for dose calculation, the AAA algorithm (v.10.0.28) was utilized with a 0.25 cm calculation grid resolution. The total dose delivered to the patient after performing a given plan adaptation was estimated by mapping daily doses back to the reference (planning CT) image using an inverted deformation vector field obtained through DIR, and then by performing dose accumulation using the Velocity AI software. The objectives were consistent and unmodified throughout the optimization process relative to the *Planned* plans.

**Adaptive Radiation Therapy Strategies**

For this study, online, offline and dose feedback approaches were examined by simulation using the available daily anatomical CBCT data set. Online plan adaptations were simulated to occur immediately before a dose delivery while offline modifications were simulated to occur between fractions n and n+1. Dose feedback adaptation was simulated

as an offline strategy and utilized the dose delivered in the previous fraction to guide a plan adaptation (re-optimization) for the next fraction. Overall, eight adaptive radiation therapy strategies were simulated for all patients as described below.

i. *DF 2-4* – A combination of the *non-ART* treatment plan and dose feedback (DF) adaptation. In the 1st, 3rd, and 5th fraction, the *non-ART* plan was delivered. In the 2nd and 4th fraction, the *non-ART* plan was re-optimized based on the dose delivered during the previous fraction. The reasoning behind performing re-optimization only during the 2nd and 4th fraction is as follows: The second fraction is the first fraction that can use the feedback from the dose delivered in the previous (1st) fraction. During the third fraction, the *non-ART* treatment plan was delivered because the dose delivered during the 2nd fraction accounted for dose discrepancies resulting from dose delivery during the first fraction, and thus up to the 3rd fraction the plan was assumed to be delivered optimally. Therefore, the time-consuming dose feedback adaptation was not used during the 3rd fraction. To examine the performance of dose-feedback applied to every fraction (for example with the availability of *in vivo* patient dosimetry), we have also tested a continuous dose-feedback adaptation (*Cont.+DF* approach).

   The dose delivered in the previous fraction was incorporated in the optimization process for the current fraction using the "Dose Based" plan optimization module of Eclipse. Before the start of the optimization process, the dose delivered in the previous fraction was mapped to the patient's anatomy of the current fraction using deformable image registration. Once the optimization was initiated, the optimizer compensated for regions of lower than or higher than intended dose by delivering a higher/lower dose to those regions, so that the total accumulated dose delivered during the previous and

the current fraction would meet the treatment plan objectives. The application of the dose feedback in the subsequent plan adaptation scenarios was performed in the same manner.

ii. *Offline* – based on the offline adaptation of the *non-ART* treatment plan. In the 1st fraction, the *non-ART* plan was delivered. To deliver the dose during the 2nd, 3rd, 4th, and 5th fraction the treatment plan was adapted by plan re-optimization using the dCT from the previous fraction (obtained based on the previous fraction's CBCT). As it implies, in this adaptation scenario only changes in the patients' anatomy detected on the daily CBCT image dataset relative to the planning CT images were accounted for. The dose delivered during the previous fraction was *not* considered as described in approach (i).

iii. *Offline + DF* – based on the combination of offline adaptation of the *non-ART* plan and dose feedback adaptation. In the 1st fraction, the *non-ART* plan was delivered. In the 2nd fraction, the treatment plan was re-optimized using the daily image data sets (daily dCT) and the dose delivered during the previous fraction (dose feedback adaptation). In the 3rd fraction, the dose was delivered using an offline adapted plan (based on the previous fraction image data sets). In the 4th and 5th fractions, dose feedback and offline adaptation were applied, respectively.

iv. *Online* – based on the daily online adaptation of the *non-ART* plan. In all fractions, before the daily dose was delivered, the treatment plan was re-optimized according to the patient's anatomy just before treatment delivery.

v. *Online + DF* – based on the combination of daily online adaptation and dose feedback adaptation. In the 1st, 3rd, and 5th fraction online adaptation was performed while in the 2nd and 4th fraction dose feedback adaptation was performed.

vi. *Cont. + DF* – based on the combination of continuous dose feedback adaptation. In the 1st fraction, the *non-ART* plan was delivered. During the remaining, 2nd – 5th fractions, the treatment plan was adapted using the patient's daily anatomy. The plan re-optimization was performed based on the total dose delivered over all the previous fractions.

vii. *Online 1-3-5* – based on the daily online adaptations and *non-ART* treatment plan. In the 1st, 3rd, and 5th fraction the treatment plan was adapted using an online approach while in the 2nd, and 4th fraction the *non-ART* treatment plan was delivered.

viii. *Offline+Online* – based on the combination of online and offline adaptations. In the 1st, 3rd, and 5th fraction the treatment plan was adapted using an online approach while in the 2nd, and 4th fraction using offline plan adaptation.

The purpose behind creating various adaptive radiation therapy strategies was to find an optimal solution for treatment plan adaptation which minimizes the negative impact of changes in the patient's anatomy on the accuracy of the delivered dose while also minimizing the time it would take to perform such adaptations in a clinical environment. For example, the rationale behind examining the *Online 1-3-5* strategy, where online adjustments of the treatment plan were performed every second fraction instead of every

fraction, was to decrease the total time of adaptation relative to a full *Online* strategy, where online adjustments of the treatment plan were performed every fraction. Another example is a *DF 2-4* strategy (where *DF*=dose feedback). Incorporating a dose feedback step allows accounting for potentially inaccurate dose delivery in the previous fraction and the change in the patient's anatomy because optimization involving the DF step is performed on the daily imaging data. Although the implementation of those steps will increase total treatment time it is expected that the improved accuracy in the dose delivery will justify the additional workload.

In this study, we define the reference *non-ART* plan as the one that was created based on the pCT data and was delivered at every treatment fraction without modifications (dose delivered was calculated based on the CBCT and mapped back to the reference pCT image). This is an estimate of what dose is actually delivered by the conventional *non-ART* approach.

*Planned* delivery reflects the intended (ie. prescribed) dosimetry of the treatment plan, as approved by the radiation oncologist.

### 4.2.4 Evaluation of Adaptive Radiation Therapy Strategies

The dosimetric effectiveness of Adaptive Radiation Therapy strategies was evaluated using a variety of dose and dose-volume metrics for target and OARs as specified in Table 4.2. Metrics were selected based on the relevant literature to ensure general applicability[16-23]. The values of each metric were associated with anatomic structures, ART strategies, and the reference plan. Where applicable, the evaluation metrics for ART plans were presented relative to the reference plan as well to quantify the dosimetric benefit of

applying plan adaptations compared to the situation where the adaptations were not incorporated in the treatment process.

**Table 4.2.** Quantitative metrics used for evaluation of adaptive radiation therapy strategies. Apart from all listed metrics, maximum dose ($D_{max}$), mean dose ($D_{mean}$), and minimum dose ($D_{min}$) for each structure were determined as well.

| **CTV & PTV** | **Bladder & Rectum** | **Femur-LT & Femur-RT** |
|---|---|---|
| $D_{1\%}$ [cGy] | $D_{1\%}$ [cGy] | $D_{1\%}$ [cGy] |
| $D_{95\%}$ [cGy] | $D_{1cc}$ [cGy] | $D_{1cc}$ [cGy] |
| $D_{98\%}$ [cGy] | $D_{2\%}$ [cGy] | $D_{2\%}$ [cGy] |
| $D_{99\%}$ [cGy] | $V_{20\ Gy}$ [% of volume] | $D_{5\%}$ [cGy] |
| $V_{100\%}$ [% of volume] | $V_{32\ Gy}$ [% of volume] | |
| $V_{105\%}$ [% of volume] | $V_{38\ Gy}$ [% of volume] | |
| HI* [%] | | |
| CI** [%] (PTV only) | | |
| $V_{95\%}$ [% of volume] (PTV only) | | |

$D_{v\%}$ - minimum dose delivered to the "hottest" v% of the volume, D1cc – minimum absolute dose for the "hottest" 1cm$^3$ of the volume $V_{d\ Gy}$ - a volume that received absolute dose of d, HI – homogeneity index calculated as a ratio ($D_{2\%}$ - $D_{98\%}$)/$D_{50\%}$ [23], CI – conformity index calculated as a ratio $V_{95\%}$/Volume of PTV [24].

The calculation of the percentage of plans that passed the treatment planning criteria for CTV, PTV, and OAR structures was reported as well. Due to their clinical relevance, the percentage deviations from the treatment planning objectives were also reported. Importantly, our conclusions with respect to superiority and inferiority of particular ART strategies relative to other dose delivery approaches were mostly driven by analysis of treatment planning objectives. Specifically, the larger the passing rate, the more clinically feasible we considered a given ART approach. For CTV and PTV, the smaller the absolute percentage deviations from the planning objectives were considered better. For

OARs, the smaller percentage deviations were considered better. The large number of other dose-volume metrics (Table 4.2) that we have included in our study provide a more comprehensive view on the dosimetry of all considered plans but were not considered in the exact treatment planning objectives used to derive the plans (these were included in Table 4.1).

The statistical significance of the results comparing the *non-ART* plan to all the other plans was determined using paired T-tests, using the $p$-value associated with a 95% confidence level. The time efficiency of the best performing ART method was also reported. The time required for each ART strategy was estimated with the consideration of plan optimization, dose calculation and deformable image registration procedures.

### 4.2.5 Qualitative Assessment of Image Registration

In order to ensure that the image registration did not introduce any major errors in terms of the patient's anatomy deformations, a qualitative (i.e. visual) evaluation of deformed images and deformable vector fields was performed. In particular, the deformed images were compared to the target images by using image overlays, checkboard filters, dynamic magnifying window focusing on soft tissue and bone tissue alignment as well as external body contour, for every registration. The analysis of the vector fields included the inspection of the deformed grid that reflected the magnitudes and directions in the field. The inspection was conducted using tools available in the commercial image registration software Velocity AI (as specified in the section 4.2.3).

## 4.3 Results

### 4.3.1 Maximum, Minimum and Mean Doses

Figure 4.1 shows the percentage differences relative to the *non-ART* (i.e. reference, not adapted) plan in $D_{max}$, $D_{mean,}$ and $D_{min}$ for CTV, PTV (Figure 4.1a) as well as in $D_{max}$ and $D_{mean}$ for OARs (Figure 4.1b). $D_{min}$ for OARs was calculated as well but due to their limited applicability were not included in the results. Appendix 8.1 (Table 8.1) contains the detailed tabular data for Figure 4.1 including $D_{min}$ for OAR and standard deviations for all metrics.

All the metrics in Figure 4.1a indicate that values of $D_{max}$, $D_{mean}$ and $D_{min}$ in the case of CTV and PTV for adapted plans - were closer to the original (*Planned*) plans compared to the unadapted reference plan. Overall, in terms of dose metrics reported in Figure 4.1a, continuous dose feedback adaptation outperformed other ART strategies and had a performance close to online adaptations The *Online* and *Online 1-3-5* plans scored as well as *Planned* plans in terms of the maximum dose delivered to PTV. The most apparent difference between adapted plans and the dose delivered by the non-adapted plan was reflected in the value of the minimum dose to PTV. In particular, the unmodified plan resulted in the delivery of approximately 17% lower $D_{min}$ to PTV compared to the planned minimum dose. Most adapted plans significantly improved the delivered dose in terms of this evaluation metric. Overall, it can be noticed that *non-ART* plans delivered a lower radiation dose to both CTV and PTV in terms of $D_{max}$, $D_{mean}$ and $D_{min.}$

Figure 4.1b for the OARs shows that $D_{max}$ for both bladder and rectum in the *non-ART* plan differed from the planned dose by around 5% (decrease). However, adapted plans


**Figure 4.1.** Comparison of $D_{max}$ and $D_{mean}$, and $D_{min}$ metrics between the *Planned* plan as well as non-ART and adapted plans for a) target structures and **b)** for OARs. Bars represent percentage differences (averaged over 20 patients) in particular metrics relative to the *Planned* treatment. The measure of 0% on the y-axis is the reference point reflecting the value of a given metric for the *Planned* plan (ie. reference plan).

were able to closely match the *Planned* plans decreasing the difference in $D_{max}$ to around 1-2%. In contrast, the mean dose for bladder and rectum showed larger variations for the adapted plans relative to the *non-ART* plan. In the case of the bladder, the majority of adaptations increased the $D_{mean}$ by less than 5%. Only *Cont.+DF* plans escalated the mean dose by around 6%. *Online* and *Online 1-3-5* plans were able to slightly reduce the $D_{mean}$ for the bladder relative to the planned dose. The rectum received approximately 5% lower mean dose upon delivery of the *non-ART* plan compared to the *Planned* dose. As can be seen, *Cont.+DF*, *Offline+DF,* and online strategies were able to decrease that difference to roughly 0.5-2%. The dose to the femoral heads was spared the most through the application of *Online* and *Online 1-3-5*. As for the right femoral head, those two online adaptation techniques delivered maximum doses significantly smaller even compared to the planned dose. However, the *Cont.+DF* adaptation resulted in higher than intended dose to both left and right femoral heads – a 6% and over 10% increase in the mean dose, respectively.

### 4.3.2 Dose-Volume Metrics

#### Target Structures

Figure 4.2 shows the relative values of dose-volume metrics that were calculated within the evaluation of adaptive radiation therapy strategies for the CTV and PTV. Overall, the dose delivered with the various adaptive strategies was consistently closer to the planned dose compared to the non-ART approach. In that regard, *Offline+Online* and *Offline* plans demonstrated the lowest while *Cont.+DF* and online adaptations demonstrated the highest dosimetric performance for both target structures. The PTV benefited from ART more than CTV as shown by $D_{95\%}$, $D_{98\%}$ and $D_{99\%}$, metric. Notably the values of these three metrics for *non-ART* plans were approximately 4% lower for CTV and as much as 11%

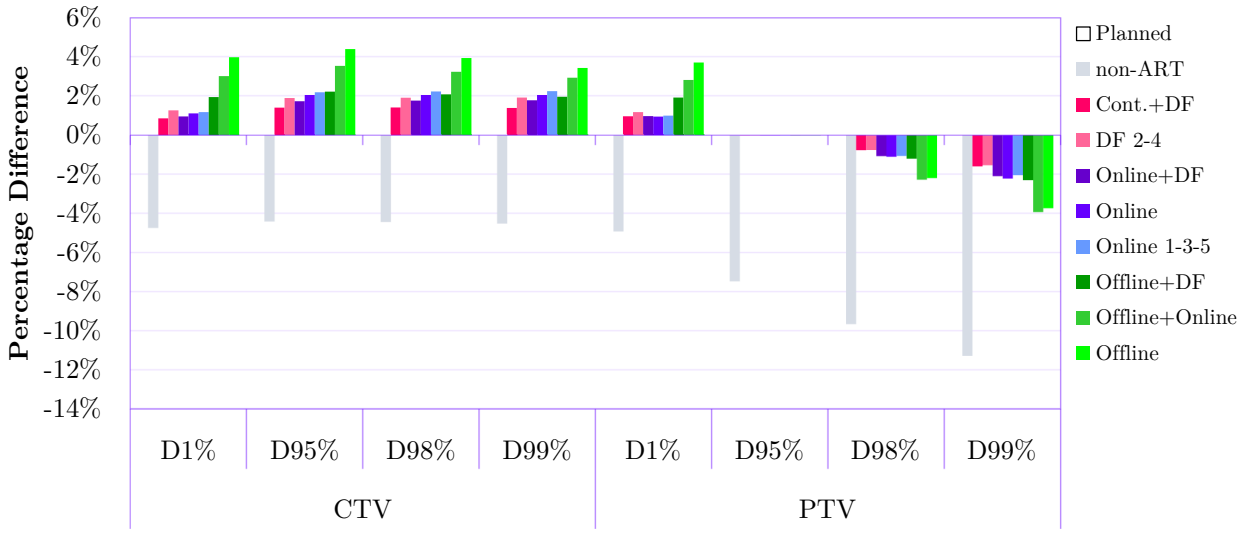



**Figure 4.2.** Comparison of dose-volume metrics between the *Planned* plan as well as non-ART and adapted plans for the Clinical Target Volume (CTV) and Planning Target Volume (PTV). Bars represent percentage differences (averaged over 20 patients) in particular metrics relative to the *Planned* treatment. The measure of 0% on the y-axis is the reference point reflecting the value of a given metric for the *Planned* plan.

lower for PTV compared to *Planned* plans. Tabular data for Figure 4.2 with standard deviations are included in the Appendix 8.2 (Table 8.2). When it comes to HI for CTV (Figure 4.3), the majority of adapted plans improved the homogeneity (i.e., a lower HI indicates a higher homogeneity level) of the dose distribution relative not only to the *non-ART* plan but also to the *Planned* standard treatment. *Online* and *Online 1-3-5* delivered the highest benefit. For PTV, HI was also improved among all adapted plans. Specifically, online and dose feedback plans performed similarly and outperformed offline strategies by around 20% to 40%. Compared to the *non-ART* dose delivery, CI for the planning target volume was approximately 10% higher when adapted plans were utilized. *Online* ART, in particular, very closely matched the *Planned* treatment. According to most metrics presented in Figure 4.3 *Offline* and *Offline+Online* adaptations were not meaningfully beneficial to the radiation treatment dosimetry.

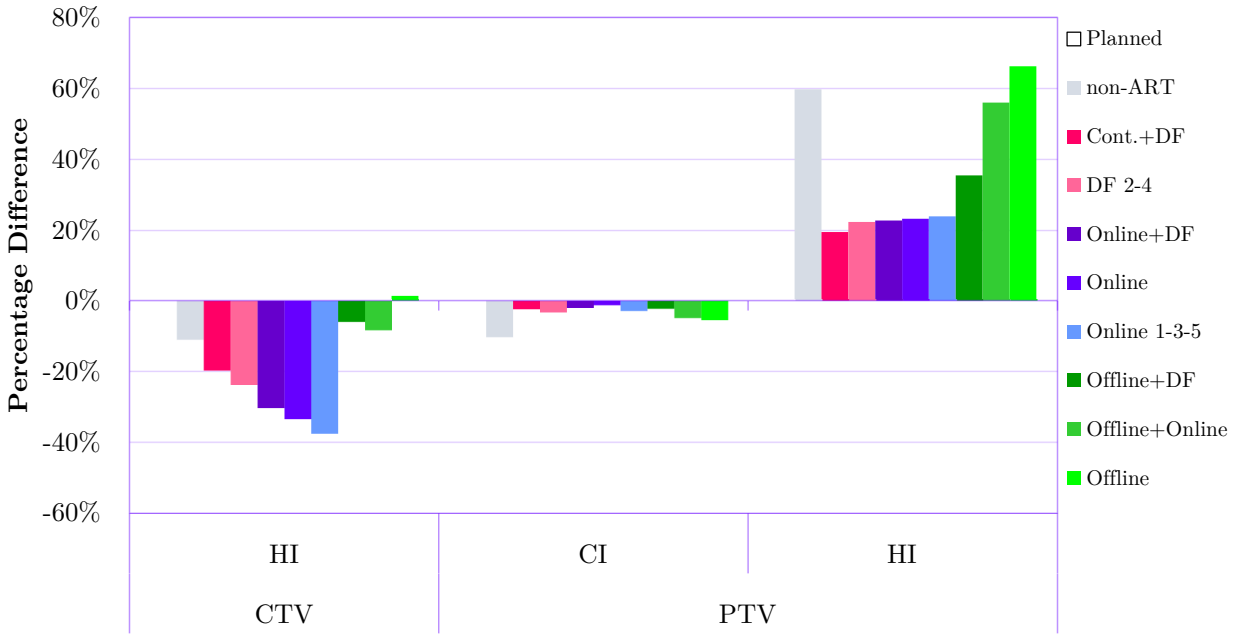



**Figure 4.3.** Comparison of homogeneity index (HI), and conformity index (CI) between the *Planned* plan as well as non-ART and all adapted plans for Clinical Target Volume (CTV) and Planning Target Volume (PTV). Bars represent percentage differences (averaged over 20 patients) in the particular metrics relative to the *Planned* treatment. The measure of 0% on the y-axis is the reference point reflecting the value of a given metric for *the Planned* plan.

### Organs at Risk

Figure 4.4 illustrates the dose-volume metrics for the bladder and rectum. In the case of the bladder, $D_{1\%}$, $D_{1cc}$, and $D_{2\%}$ did not differ significantly from the planned dose for adapted plans but were lower by approximately 7% for non-ART plans. Only *Cont.+DF,* and a few *Offline* adaptations, showed slightly higher values compared to the *Planned* delivery. Larger variations in magnitude were observed in $V_{15\%}$, $V_{20\%}$, and $V_{50\%}$ metrics. The most desirable results were obtained through *Online* and *Online 1-3-5* strategies. *Cont.+DF* adaptive plans showed poor performance with large volumes receiving 15%,

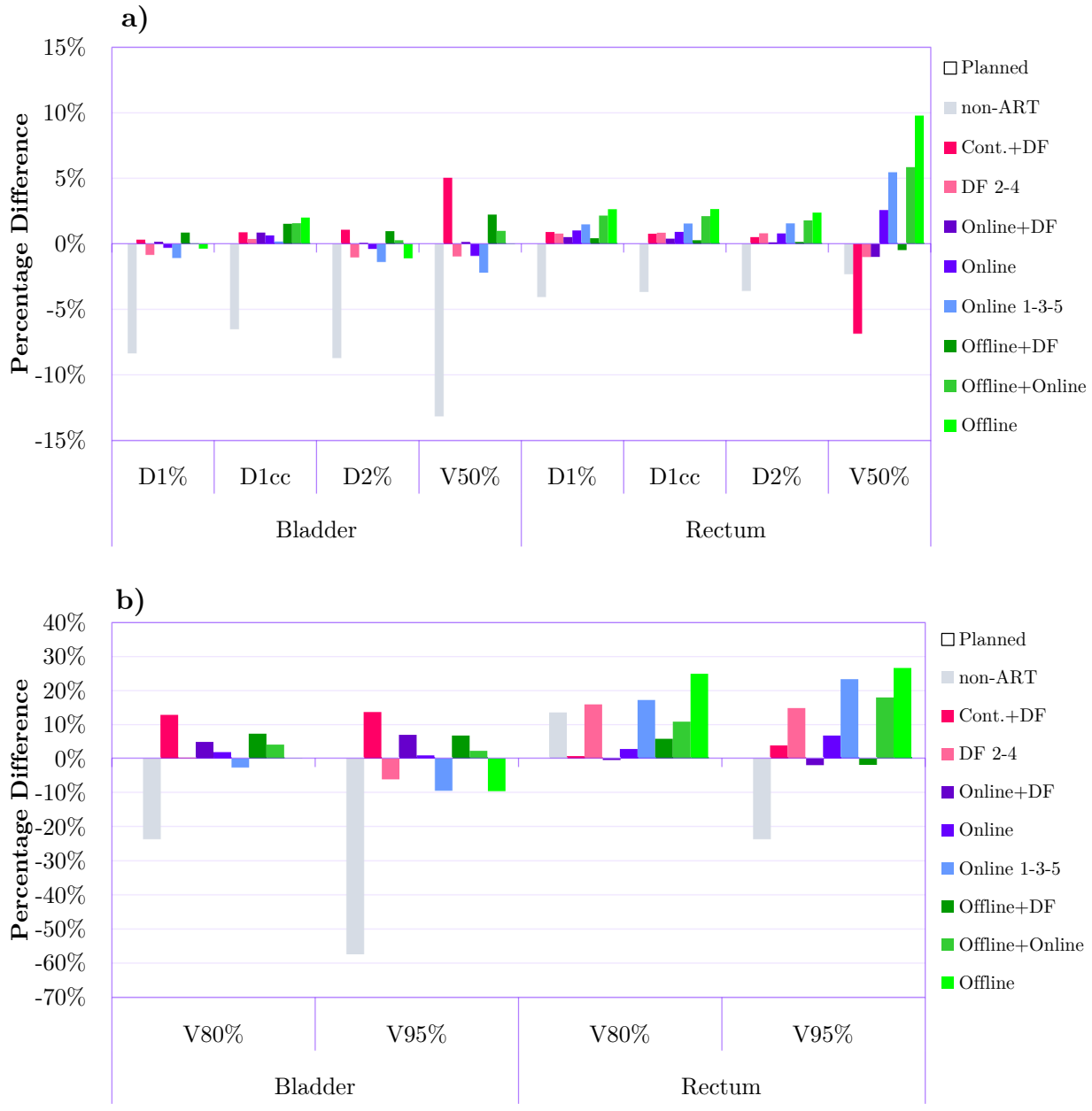



**Figure 4.4.** Comparison of dose-volume metrics between the *Planned* plan as well as non-ART and all adapted plans for bladder and rectum. Bars represent percentage differences (averaged over 20 patients) in particular metrics relative to the *Planned* treatment. The measure of 0% on the y-axis is the reference point reflecting the value of a given metric for the *Planned* plan. The charts **a)** and **b)** were separated for better visualization of the results due to the large differences in y-axis values between $V_{80\%}$, $V_{95\%}$, and the rest of the metrics.

20%, and 50% of the prescription dose as can be seen in Figure 4.4a. The same observation can be made in Figure 4.4b showing $V_{80\%}$ and $V_{95\%}$ metrics for the bladder. Considering results for the rectum, most ART strategies were able to closely match *Planned* values of $D_{1\%}$, $D_{1cc}$, and $D_{2\%}$ (Figure 4.4a) as well as values of $V_{80\%}$ and $V_{95\%}$ (*Cont.+DF, Online+DF, Online and Offline+DF* in Figure 4.4b). For $V_{15\%}$ and $V_{20\%}$, the majority of adapted plans delivered nearly the same results except for *Cont.+DF* for which approximately 2% volume increase was noted for 15% and 20% dose prescription levels. Compared to adapted, the *non-ART* plans were closer to the *Planned* plans for those two volume metrics. $V_{50\%}$ for *Planned* treatment was approximately the same as for *DF 2-4* and *Online+DF* adaptations. The *Cont.+DF* approach resulted in $V_{50\%}$ being around 7% lower compared to the *Planned* delivery.

Figure 4.5 demonstrates that for the left femoral head only the *Online 1-3-5* plans were able to keep the $D_{1\%}$, $D_{1cc}$, $D_{2\%,}$ and $D_{5\%}$ at the level close to the planned treatment. All the other plan modifications, except for non-ART resulted in doses higher than the Planned plans by around 4% to 11%. In contrast, the majority of adapted plans (except for *Cont.+DF* and *DF 2-4*), especially *Online* and *Online 1-3-5*, deliver a lower $D_{1\%}$, $D_{1cc}$, $D_{2\%,}$ and $D_{5\%}$ by up to 5% to the right femoral head compared to *Planned* dose delivery.

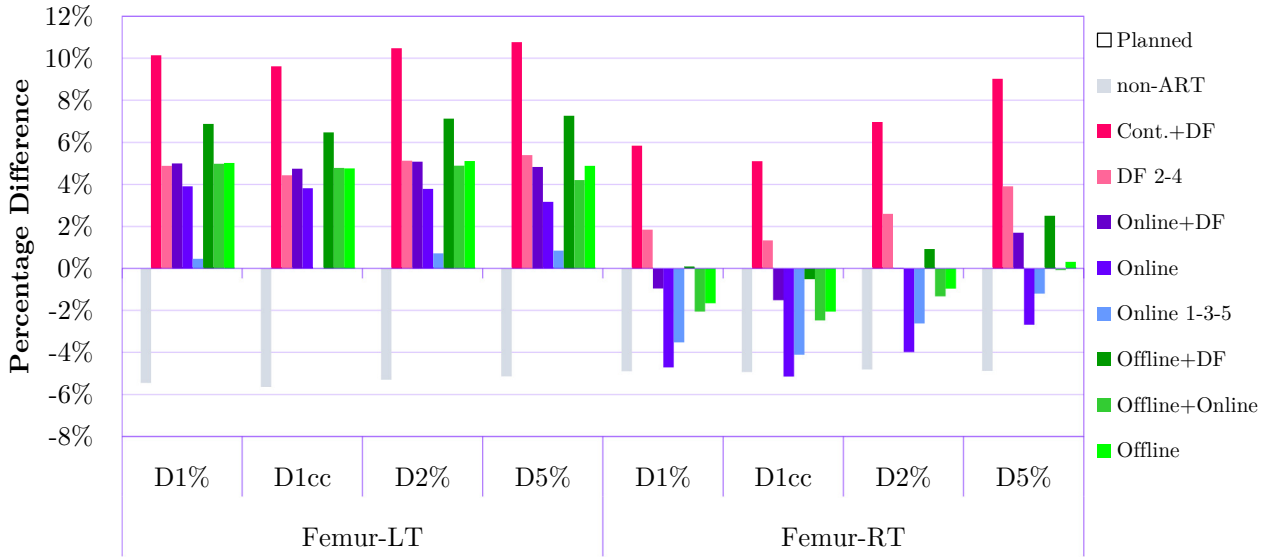



**Figure 4.5.** Comparison of dose-volume metrics between the *Planned* plan as well as non-ART and all adapted plans for left and right femoral heads. Bars represent percentage differences (averaged over 20 patients) in particular metrics relative to the *Planned* treatment. The measure of 0% on the y-axis is the reference point reflecting the value of a given metric for the *Planned* plan. The lack of expected symmetry in the dose delivered by adapted plans to both femoral heads is explained in the Discussion section.

### 4.3.3 Comparison to Treatment Planning Criteria

#### Target Structures

Figure 4.6a demonstrates the percentage of patients for whom a given treatment plan met the treatment planning objectives specified in Table 4.1 (section 4.2.2) for CTV and PTV structures. For CTV, all plans had at least a 90% passing rate except for *non-ART* plans for which no patients passed CTV or PTV planning objectives. 100% of patients passed the CTV criteria in the case of *DF 4-2* and three *Online* adaptations. For the PTV these same four strategies were able to achieve a 60% to 80% passing rate for the $D_{max}$ objective, while other ART approaches received 50% and lower rates. The $V_{38\%}$ criterion for PTV was very challenging to reach even for well-performing online adaptations. The maximum passing rate was achieved by *Online* ART and was equal to slightly above 20%.

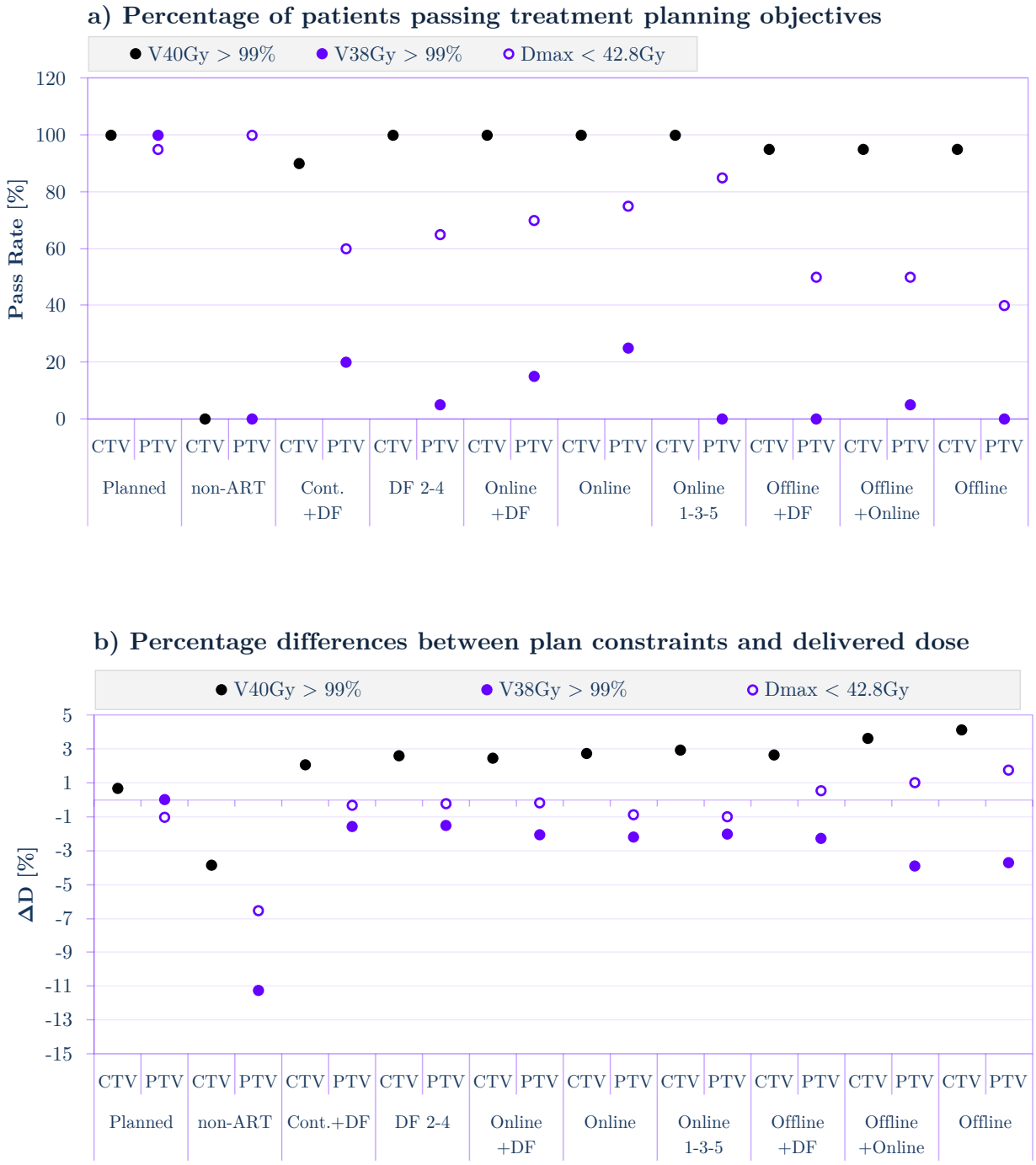



**Figure 4.6. a)** Passing rates for Clinical Target Volume (CTV) and Planning Target Volume (PTV) for three treatment planning objectives; **b)** Dose differences between the planning objectives and the dose delivered by the specific treatment plan.

Figure 4.6b details the dose difference, ΔD, between the considered criteria and the value achieved by the plan. For the CTV, nearly all the plans were able to meet the treatment planning objective and ΔD is positive ranging from 1% to 4%. The dose difference in $D_{max}$ for the PTV was approximately 1% for investigated ART approaches. It is noted that even though the passing rate for offline adaptations was lower in comparison to the rest of the ART strategies, the ΔD is, on average, positive for *Offline+DF*, *Offline+Online,* and *Offline* plans. This clearly shows that several patients delivered higher doses to the CTV so that it was able to cause the increase in the dose averaged over all 20 patients. The ΔD for $V_{38Gy}$ objective for PTV ranged from around -4% in the case of *Offline* plans to approximately -2% for *Online* adaptations. Compared to all the ART approaches the non-ART plans differs significantly from the *Planned* plans by -4% to -11% depending on the planning criteria.

In summary, the implementation of the majority of the ART strategies improved the overall passing rate and ΔD for most of the plans, especially for daily online adaptations compared to the delivery of an unchanged *non-ART* treatment plans.

**Organs at Risk**

The passing rate presented in Figure 4.7a is equal to 100% across all the plans for the bladder in the case of both plan objectives. Consistently, Figure 4.7b shows that ΔD for the bladder is significantly (around 60%) below tolerance doses ($D_{15\%}$ and $D_{20\%}$). In the case of the rectum, the passing rate for three *Offline* adaptations and the *non-ART* plan was in the range of 60% to 85% for the $V_{28Gy}$ metric and from 80% to 100% for the $V_{32Gy}$ metric. For *Online* plans, the analogous range was from 80% to 95% for $V_{28Gy}$ and from 95% to 100% for $V_{32Gy}$. ΔD was negative for all the plans and had nearly the same


**Figure 4.7. a)** Passing rates for Organs at Risk (OARs) in relation to the treatment planning objectives. **b)** Dose differences between the planning objective and the dose value delivered by the specific treatment plan.

magnitude (of 10%) for most of the adaptations for the $V_{32Gy}$ and $V_{28Gy}$ criteria. The absence of femoral heads in this analysis is addressed in the discussion section.

### 4.3.4 Statistical Significance

As mentioned in the Material and Methods section, the statistical significance of the results was calculated using paired, two-tailed T-tests. The analysis was based on the comparison of dose and dose-volume metrics for target and OARs structures between the *non-ART* plan and the other treatment strategies for all 20 patients. The relevant *p*-values with a 95% confidence level are presented in Table 4.3 for $D_{max}$, $D_{min}$, and $D_{mean}$ as well as in Appendix 8.1 (Table 8.3) for the remaining metrics. Both tables also summarize the total number of metrics for which the test determined the statistical significance at the level of $p<0.05$.

Table 4.3 shows that the majority of the results were statistically significant for the *non-ART* treatment. Among ART plans the lowest *p*-value was observed for *Cont.+DF* and offline adaptations, while the highest were observed for the remaining plan modification strategies. The results for PTV and CTV indicate a similar level of statistical significance. When considering organs at risk, the results for the rectum demonstrated a statistical significance similar to that of the bladder with the exception of minimum dose for which results corresponding to bladder where more significant compared to the rectum. It is also worth to note that the null-hypothesis for a 2-sample t-test (here, two tailed) is that the two groups are equal. Low p-values (<0.05) indicate that there was strong evidence against the null hypothesis (the null-hypothesis can be rejected with 95% confidence). High p-values imply strong evidence in favour of null hypothesis (null-hypothesis cannot be rejected).

**Table 4.3.** The results of paired, two-tailed T-test for $D_{max}$, $D_{min}$, and $D_{mean.}$ The fields for which $.01<p\leq.05$ are highlighted in green (ie. significant), while those with $p\leq.01$ were highlighted in red (ie. strongly significant).

| | | | Adaptive Radiation Therapy Strategies | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Structure** | **Metric** | **non-ART** | **Cont. +DF** | **DF 2-4** | **Online +DF** | **Online** | **Online 1-3-5** | **Offline +DF** | **Offline +Online** | **Offline** |
| **CTV** | Max. Dose [%] | 0.0000 | 0.4591 | 0.1416 | 0.2500 | 0.3321 | 0.2263 | 0.0089 | 0.0245 | 0.0007 |
| | Mean Dose [%] | 0.0000 | 0.0136 | 0.0017 | 0.0058 | 0.0050 | 0.0002 | 0.0002 | 0.0022 | 0.0000 |
| | Min. Dose [%] | 0.0000 | 0.0059 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0046 | 0.0000 | 0.0004 |
| **PTV** | Max. Dose [%] | 0.0000 | 0.1893 | 0.0796 | 0.1263 | 0.7760 | 0.9321 | 0.0037 | 0.0477 | 0.0026 |
| | Mean Dose [%] | 0.0000 | 0.0132 | 0.0019 | 0.0060 | 0.0090 | 0.0004 | 0.0002 | 0.0036 | 0.0000 |
| | Min. Dose [%] | 0.0000 | 0.0223 | 0.0013 | 0.0234 | 0.0293 | 0.0013 | 0.0027 | 0.0018 | 0.0001 |
| **Bladder** | Max. Dose [%] | 0.0000 | 0.1813 | 0.3109 | 0.3115 | 0.9035 | 0.6836 | 0.0410 | 0.1550 | 0.0115 |
| | Mean Dose [%] | 0.0007 | 0.1261 | 0.9888 | 0.7669 | 0.9129 | 0.4376 | 0.3785 | 0.7188 | 0.8258 |
| | Min. Dose [%] | 0.0000 | 0.0021 | 0.0002 | 0.0004 | 0.0002 | 0.0000 | 0.0003 | 0.0001 | 0.0001 |
| **Rectum** | Max. Dose [%] | 0.0000 | 0.0792 | 0.2371 | 0.7090 | 0.3807 | 0.1985 | 0.0515 | 0.0676 | 0.0046 |
| | Mean Dose [%] | 0.5983 | 0.3821 | 0.0064 | 0.1046 | 0.0079 | 0.0003 | 0.0144 | 0.0063 | 0.0000 |
| | Min. Dose [%] | 0.1078 | 0.5133 | 0.4591 | 0.4791 | 0.4282 | 0.4094 | 0.5968 | 0.5261 | 0.6682 |
| **Femur-LT** | Max. Dose [%] | 0.0000 | 0.0144 | 0.1429 | 0.1583 | 0.3177 | 0.0421 | 0.0631 | 0.2094 | 0.1201 |
| | Mean Dose [%] | 0.0000 | 0.0022 | 0.0101 | 0.5798 | 0.7615 | 0.9542 | 0.1118 | 0.9400 | 0.3883 |
| | Min. Dose [%] | 0.1249 | 0.2865 | 0.5871 | 0.8242 | 0.8768 | 0.9340 | 0.4703 | 0.6609 | 0.6645 |
| **Femur-RT** | Max. Dose [%] | 0.0000 | 0.4516 | 0.6915 | 0.0677 | 0.0043 | 0.0010 | 0.1863 | 0.0716 | 0.0856 |
| | Mean Dose [%] | 0.0000 | 0.0000 | 0.0001 | 0.0035 | 0.1629 | 0.0711 | 0.0036 | 0.0403 | 0.0236 |
| | Min. Dose [%] | 0.1061 | 0.5409 | 1.0000 | 0.7075 | 0.4057 | 0.5919 | 0.8840 | 0.4950 | 0.8847 |
| **The percentage of statistically significant metrics at $.01< p\leq.05$** | | **78%** | **44%** | **44%** | **33%** | **39%** | **44%** | **56%** | **50%** | **61%** |
| **The percentage of statistically significant metrics at $p\leq.01$** | | **78%** | **22%** | **44%** | **28%** | **33%** | **39%** | **44%** | **33%** | **50%** |

### 4.3.5. Time Efficiency

The time required to manually complete key steps in the ART loop in the clinic was: i) plan optimization ~4 min 30 s.; ii) dose calculation ~1 min 30 s.; iii) deformable image registration ~1 min. 50 s, for a total time of around 7 min 50 s per online plan adaptation for a single fraction. The limitation of our time estimation is that it did not include CBCT image acquisition, data processing, and possible verification step that may be required for a newly adapted treatment plan.

### 4.3.6 Qualitative Assessment of Image Registration

Thorough visual inspection of image registration results did not reveal any major, non-physical image deformations that could negatively impact the dosimetric results of ART strategies explored in this study. The alignment of the soft and bone tissues as well as external body contours that were inspected with overlays, checkerboards filter and dynamic magnifying window confirmed a high quality of image registrations. The deformation vector field was smooth without folding distortions, indicating that only realistic deformations of the patient anatomy occurred during the registration process.

## 4.4 Discussion

*Offline*, *Offline* and *Offline+Online* plan adaptations resulted in the highest delivered dose to CTV and PTV compared to other ART strategies which was demonstrated by the majority of dose-volume metrics (Figure 4.2). As explained in the Material and Methods section, *Offline* adaptation relies on the patient's anatomy from the previous fraction to modify the treatment plan that will be delivered in the next fraction. The possible issue in that approach is that if the magnitude of interfractional changes in the patient's anatomy is significant, the offline adapted treatment plan may not be able to correct the *non-ART* plan as intended. However, literature findings have shows that at least in some situations, online and offline ART approaches can deliver similar dosimetric performance (which can be patient specific). For example, Sun et al. investigated 22 prostate cancer patients who underwent IMRT treatment (total dose of 64Gy in 20 fractions) and found that both the online and offline adaptations performed similarly[26].

Figure 4.2 also demonstrates that relative to CTV the dose coverage of the PTV is more sensitive to anatomical changes. The $D_{98\%}$ and $D_{99\%}$ for that structure were on average 5%

and 6% lower for the unmodified plan. The importance of proper PTV coverage highlights the significance of ART application, as it has proven its ability to improve and maintain PTV coverage. Overall, both the dose and dose-volume metrics shown in Figure 4.1 and Figure 4.2 show that non-ART plans delivered consistently lower doses to CTV and PTV compared to the *Planned* and all the ART plans.

Regarding the HI index (Figure 4.3), although we found that for CTV the daily plan re-optimizations were generating satisfactory dose distributions, their combination in the dose accumulation step resulted in an even higher level of homogeneity for the majority of patients. Despite the promising results in the HI index for CTV, the accumulated dose distribution needs to be carefully examined in clinical practice. This is because the dose accumulation can result in the appearance of hot and cold spots depending on the spatial relation between daily dose distributions. This effect is most likely responsible for the presence of relatively high doses delivered to the bladder by the *Cont.+DF* adaptation as seen in Figure 4.4a ($D_{1\%}$, $D_{1cc}$, and $D_{2\%}$) and Figure 4.4b ($V_{80\%}$ and $V_{95\%}$). It is also worth mentioning that for *Online* adaptations, a clinically acceptable value of HI was not associated with an increased dose in organs-at-risk. This is contrary to the findings of Banaei et. al who conducted a study based on 15 prostate cancer patients that were delivered IMRT treatment[27]. Researchers reported inverse exponential relationships between the OAR sparing and HI, which might be the case due to the differences in dose delivery techniques. Banaei et. al used nine static IMRT beams, while this study utilized VMAT techniques with two arcs. Chow et. al showed that in the case of prostate cancer, VMAT compared to IMRT provides more desirable PTV coverage in terms of both HI (0.09 vs 0.12) and CI index (0.94 vs 0.89)[28]. It should also be noted that based on the

data presented in the Figure 4.4, the *non-ART* plans delivered lower radiation doses to both bladder and rectum compared to all other delivery approaches. Although it can be seen as a desirable outcome, the *non-ART* plans are not optimal because as mentioned in the previous paragraph, both CTV and PTV coverage significantly differ from the *Planned* dose.

It is also important to notice that, as shown in Figure 4.5, the dose delivered by adapted plans to the right femoral head was noticeably lower than for the left femoral head. It is not very clear why the dose sparing is not approximately symmetrical for these two structures. The authors suspect that because the dose delivered to femoral heads was significantly lower than planning objective threshold the observation of asymmetry is a form of systematic noise and an indication of limited priority given to those structures by the optimizer during plan re-optimization. The direction of the first arc may also contribute to this, as the second arc (in the opposite direction) typically has a lesser dose impact than the first arc (ie. the second arc 'fine tunes' the dose). In the early adoption of VMAT in clinical practice, Hardcastle et al.[29] reported such asymmetry but admitted that the reason was also not clear, further adding that gantry rotation direction did not affect the asymmetry of the dose distribution. Also, Tran et al.[23] indicated that the left femoral head received a higher dose than the right femoral head however due to the fact that dose-volume objectives were met, this observation was not discussed.

Figure 4.6a clearly shows the advantage of ART application in that the passing rate for $D_{99\%}$ in CTV was around 100% for the *Online*, *Online 1-3-5 and Online+DF* adaptations compared to 0% for the unaltered *non-ART* plan. The 100% passing rate for the bladder for all plans, as shown in Figure 4.7a was anticipated because VMAT plans have been proven to be able

to decrease the dose delivered to that organ very effectively (large $\Delta$D~60% on Figure 4.7b) in prostate cancer radiotherapy[30]. Limiting the dose to the rectum by the application of various ART strategies was more challenging (small $\Delta$D~10% in Figure 4.7b, passing rate as low as 65% for $D_{20\%}$ in Figure 4.7a) due to the position, size, and the increased daily movement of the rectum. Also, the *Online* approach resulted in the highest passing rate for rectum planning objectives. The analysis of treatment planning objective passing rates and dose deviations $\Delta$D shows that *Online* and *Online 1-3-5* strategies are very promising for ART adoption in the clinical environment. It can be noticed that the dose feedback approaches resulted in the lowest absolute values of $\Delta$D (Figure 4.6b) for target structures, however, compared to online strategies they are significantly more resource intensive thus may not be an optimal choice in practical ART applications.

It is also interesting to observe that the passing rate for PTV objective ($D_{99\%}$>3800cGy, Figure 4.6a) was very low. In particular *non-ART, Online 1-3-5, DF 2-4, Offline+DF, Offline+Online, Offline* all had zero or close to zero passing rates. Figure 4.6b shows why this might be the case. We can see that even the original (planned) plans barely meet those criteria ($\Delta$D is almost equal to zero; 0.02%). This means that any, even the smallest random error related to any aspect of treatment delivery, patient positioning etc. could easily invalidate this particular objective. In our case, the source of this error could be very small random contouring variability. Only the most resource intensive strategies (*Cont.+DF*, *Online+DF* and *Online*) were able to slightly mitigate this discrepancy.

Although we found that Online 1-3-5 approach was the most efficient and effective, one may ask why not to adapt the plan simply upon detection of discrepancy, without relying on any predetermined strategy. The main reason for this is that ART aims to keep high

accuracy of dose delivery at all times if possible, preventing geometric or dosimetric objectives from violating acceptable ranges of accuracy.

One of the study limitations is that the comprehensive quantitative evaluation of the deformable image registration algorithm as well as the spatial relationship between the registration error and its impact on the accuracy of dose estimation were beyond the scope of this work. However, it is expected to be a relatively small effect compared to the impact of the various adaptation strategies[31]. Another limitation is that the plan re-optimizations were performed automatically using original plan objectives. The future work could potentially explore a possibility of adapting those plans manually by experienced radiation therapy planner.

Table 4.3 shows that the statistical significance of the results is, in general, higher for *Online 1-3-5* plans compared to *Online* ART. Although overall the *Online* adaptation delivered a dosimetric performance slightly closer to the initially intended plan, the *Online 1-3-5* strategy is 40% faster due to fewer adaptations required. The trade-off between the time efficiency and the dosimetric results presented can be useful for both busy clinics and centers with larger time allocation per treatment plan.

As mentioned in the introduction, ART is an active area of research, however, the number of papers reporting the comprehensive evaluation of various adaptation scenarios is limited. Often authors study just a single or few approaches[32-36]. Therefore, we believe that our comprehensive approach brings value in evaluating ART focused on the two full arcs hypofractionated VMAT treatments for prostate cancer patients considered in this study.

The evaluation of a variety of ART strategies will help to easier identify an ART approach that is best suited for individual clinics.

## 4.5 Conclusion

The aim of this research was to investigate and quantify the dosimetric benefits resulting from the application of several different adaptive radiation therapy strategies for hypofractionated VMAT treatments for prostate cancer patients. The findings of our work quantify these improvements and indicate that performing daily online adaptations every fraction or every second fraction improves the dosimetric outcomes of delivered radiotherapy treatment compared to the plan that was created solely based on the pre-treatment planning CT scan and was then delivered without accounting for interfractional changes in the patient's anatomy. The strategy of adapting every second fraction achieves nearly the same dosimetric benefit to the patient but with significantly reduced resources used and may represent the most clinically attractive strategy examined here for significantly hypofractionated prostate cancer patients.

## 4.6 References

1. Huang T-C, Chou K-T, Yang S-N, Chang C-K, Liang J-A, Zhang G. Fractionated changes in prostate cancer radiotherapy using cone-beam computed tomography. *Medical Dosimetry.* 2015;40(3):222-225.

2. Chen Z, Yang Z, Wang J, Hu W. Dosimetric impact of different bladder and rectum filling during prostate cancer radiotherapy. *Radiation Oncology.* 2016;11(1):103.

3. Poli APDF, Dias RS, Giordani AJ, Segreto HRC, Segreto RA. Strategies to evaluate the impact of rectal volume on prostate motion during three-dimensional conformal radiotherapy for prostate cancer. *Radiologia brasileira.* 2016;49(1):17-20.

4. Wahl M, Descovich M, Shugard E, et al. Interfraction anatomical variability can lead to significantly increased rectal dose for patients undergoing stereotactic body radiotherapy for prostate cancer. *Technology in Cancer Research & Treatment.* 2017;16(2):178-187.

5. Nigay E, Bonsall H, Meyer B, Hunzeker A, Lenards N. Offline adaptive radiation therapy in the treatment of prostate cancer: a case study. *Medical Dosimetry.* 2019;44(1):1-6.

6. Jorgo K, Polgar C, Major T, et al. Acute and late toxicity after moderate hypofractionation with simultaneous integrated boost (SIB) radiation therapy for

prostate cancer. A single institution, prospective study. *Pathology & Oncology Research.* 2020;26(2):905-912.

7. Kim H, Lee J, Kim W. Early toxicity of moderate hypofractionated volumetric modulated Arc radiotherapy for localized prostate cancer. *International Journal of Radiation Research.* 2019;17(2):293-300.

8. Pathmanathan AU, van As NJ, Kerkmeijer LG, et al. Magnetic resonance imaging-guided adaptive radiation therapy: a "game changer" for prostate treatment? *International Journal of Radiation Oncology* Biology* Physics.* 2018;100(2):361-373.

9. Mancosu P, Clemente S, Landoni V, et al. SBRT for prostate cancer: challenges and features from a physicist prospective. *Physica Medica.* 2016;32(3):479-484.

10. Böckelmann F, Putz F, Kallis K, Lettmaier S, Fietkau R, Bert C. Adaptive radiotherapy and the dosimetric impact of inter-and intrafractional motion on the planning target volume for prostate cancer patients. *Strahlentherapie und Onkologie.* 2020.1-10.

11. Gorovets D, Burleson S, Jacobs L, et al. Prostate SBRT with intra-fraction motion management using a novel LINAC-based MV-kV imaging method. *Practical Radiation Oncology.* 2020.

12. Vanhanen A, Poulsen P, Kapanen M. Dosimetric effect of intrafraction motion and different localization strategies in prostate SBRT. *Physica Medica.* 2020;75:58-68.

13. Dang, Audrey, et al. "Image-guided radiotherapy for prostate cancer. *Translational andrology and urology* 7.3 (2018): 308.

14. Lawson JD, Schreibmann E, Jani AB, Fox T. Quantitative evaluation of a cone-beam computed tomography–planning computed tomography deformable image registration method for adaptive radiation therapy. *Journal of applied clinical medical physics.* 2007;8(4):96-113.

15. Vanetti E, Nicolini G, Nord J, et al. On the role of the optimization algorithm of RapidArc® volumetric modulated arc therapy on plan quality and efficiency. *Medical physics.* 2011;38(11):5844-5856.

16. Alongi F, Cozzi L, Arcangeli S, et al. Linac based SBRT for prostate cancer in 5 fractions with VMAT and flattening filter free beams: preliminary report of a phase II study. *Radiation Oncology.* 2013;8(1):1-8.

17. Murray LJ, Cosgrove V, Lilley J, et al. Developing a class solution for prostate stereotactic ablative body radiotherapy (SABR) using volumetric modulated arc therapy (VMAT). *Radiotherapy and Oncology.* 2014;110(2):298-302.

18. Ruggieri R, Naccarato S, Stavrev P, et al. Volumetric-modulated arc stereotactic body radiotherapy for prostate cancer: dosimetric impact of an increased near-maximum target dose and of a rectal spacer. *The British journal of radiology.* 2015;88(1054):20140736.

19. Hussein M, South CP, Barry MA, et al. Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy. *Radiotherapy and Oncology.* 2016;120(3):473-479.

20. Choi HS, Jo GS, Chae JP, et al. Defining the optimal time of adaptive replanning in prostate cancer patients with weight change during volumetric arc radiotherapy: a dosimetric and mathematical analysis using the gamma index. *Computational and mathematical methods in medicine.* 2017;2017.

21. Kang SW, Chung JB, Kim JS, et al. Optimal planning strategy among various arc arrangements for prostate stereotactic body radiotherapy with volumetric modulated arc therapy technique. *Radiology and Oncology.* 2017;51(1):112-120.

22. Seppälä J, Suilamo S, Tenhunen M, et al. Dosimetric comparison and evaluation of 4 stereotactic body radiotherapy techniques for the treatment of prostate Cancer. *Technology in Cancer Research & Treatment.* 2017;16(2):238-245.

23. Tran A, Zhang J, Woods K, et al. Treatment planning comparison of IMPT, VMAT and 4π radiotherapy for prostate cases. *Radiation Oncology.* 2017;12(1):1-9.

24. Doses A. 3. Special considerations regarding absorbed-dose and dose—volume prescribing and reporting in IMRT,". *Journal of the ICRU.* 2010;10(1):27-40.

25. Salimi M, Abi KST, Nedaie HA, et al. Assessment and comparison of homogeneity and conformity indexes in step-and-shoot and compensator-based intensity modulated radiation therapy (IMRT) and Three-dimensional conformal

radiation therapy (3D CRT) in prostate cancer. *Journal of medical signals and sensors.* 2017;7(2):102.

26. Qin A, Sun Y, Liang J, Yan D. Evaluation of online/offline image guidance/adaptation approaches for prostate cancer radiation therapy. *International Journal of Radiation Oncology* Biology* Physics.* 2015;91(5):1026-1033.

27. Banaei A, Hashemi B, Bakhshandeh M, Mofid B. Trade-off between the conflicting planning goals in correlation with patient's anatomical parameters for intensity-modulated radiotherapy of prostate cancer patients. *Journal of Radiotherapy in Practice.* 2019;18(3):232-238.

28. Chow JC, Jiang R, Kiciak A, Markel D. Dosimetric comparison between the prostate intensity-modulated radiotherapy (IMRT) and volumetric-modulated arc therapy (VMAT) plans using the planning target volume (PTV) dose–volume factor. *Journal of Radiotherapy in Practice.* 2016;15(3):263-268.

29. Hardcastle N, Tomé WA, Foo K, Miller A, Carolan M, Metcalfe P. Comparison of prostate IMRT and VMAT biologically optimised treatment plans. *Medical Dosimetry.* 2011;36(3):292-298.

30. Scobioala S, Kittel C, Elsayad K, et al. A treatment planning study comparing IMRT techniques and cyber knife for stereotactic body radiotherapy of low-risk prostate carcinoma. *Radiation Oncology.* 2019;14(1):143.

31. Paganelli C, Meschini G, Molinelli S, Riboldi M, Baroni G. Patient-specific validation of deformable image registration in radiation therapy: Overview and caveats. *Medical physics.* 2018;45(10):e908-e922.

32. Fung WWK, Wu VWC, Teo PML. Developing an adaptive radiation therapy strategy for nasopharyngeal carcinoma. *Journal of radiation research.* 2014;55(2):293-304.

33. Böck M, Eriksson K, Forsgren A, Hårdemark B. Toward robust adaptive radiation therapy strategies. *Medical Physics.* 2017;44(6):2054-2065.

34. Anaya VM, Fairfoul J. Assessing the feasibility of adaptive planning for prostate radiotherapy using Smartadapt deformable image registration. *Medical engineering & physics.* 2019;64:65-73.

35. Antico M, Prinsen P, Fracassi A, Isola A, Cobben D, Fontanarosa D. Comparison between conventional IMRT Planning and a novel real-time adaptive planning strategy in hypofractionated regimes for prostate cancer: A proof-of-concept planning study. Paper presented at: Healthcare2019.

36. Heukelom J, Fuller CD. Head and neck cancer adaptive radiation therapy (ART): conceptual considerations for the informed clinician. Paper presented at: Seminars in radiation oncology2019.

# Chapter 5

# U-net architecture with embedded Inception-ResNet-v2 image encoding modules for automatic segmentation of organs-at-risk based on computed tomography scans of head and neck cancer patients

The segmentation of critical organs is an essential step in any radiation therapy treatment planning process. However, the manual delineation of patient anatomy is very resource-intensive. It requires the expertise of a radiation oncology physician and a significant amount of time to complete. This chapter describes the study in which 25 critical organs, imaged on computed tomography (CT) scans of head and neck cancer patients, were automatically segmented using a deep learning model. Expert level accuracy, short contouring time, and realistic prediction reasoning made the model proposed in this work a feasible solution for head and neck CT imaging segmentation in a clinical environment. This chapter is under review in the Journal of Physics in Medicine & Biology.

## 5.1 Introduction

The segmentation of the head and neck region for the purpose of creating radiation therapy treatment plans takes a significant amount of time – on average 2.7 to 3.0 hours for tumor and OARs[1]. Additionally, the delineation of specific anatomical structures suffers from inter- and intra-observer variability[2,3] which has been shown to have a significant impact on radiation plan dosimetry[4]. Automatic segmentation focuses on solving these problems by providing a fast, accurate, and consistent contouring tool[5].

In recent years there have been several studies that explored the application of deep learning algorithms for the task of head and neck region segmentation. Although U-Net belongs to the most commonly used model architectures, various modifications of this network have been proposed. For example, Van Rooij et al. used 3D U-net and additionally applied a dropout regularization to network layers, randomly setting some units in the network to zero during training, which reduced the tendency to overfit[6]. Zhu et al. extended their 3D U-net architecture (called 'AnatomyNet') by introducing a scheme for segmentation of the entire CT volume (without a subdivision into local patches), embedding 3D 'squeeze-and-excitation' residual blocks in encoding layers for the improvement of feature representation and creating a new loss function by combining the Dice Score with a focal loss to facilitate model training[7]. Gao et al. proposed a two-stage

network, FocusNet-v2, to contour small organs with specifically designed small-organ localization and segmentation sub-networks while maintaining the accuracy of large organ segmentation[8].

The growing interest in new delineation methods as well as the rapid development of deep learning algorithms in the field of computer vision provides many research opportunities that have not yet been considered in the automatic segmentation of head and neck anatomy.

The purpose of this study is to use the recently modified U-Net architecture with embedded Inception-ResNet-v2 blocks, to contour organs-at-risk (OARs) based on the CT scans of head and neck cancer patients who underwent radiation therapy treatment. Compared to the U-Net architecture utilized in previous studies for Head and Neck anatomy segmentation[9,10], our U-net modification combines the flexibility of the inception modules in discovering complex image patterns through the application of parallel convolutional layers with the residual blocks containing a skip connection that allow for fast signal propagation throughout the network. Our goal was to create higher-performing and more time-efficient CNN architecture. To the best of the author's knowledge this modified architecture has not been utilized or validated for automatic segmentation of the head and neck region. Twenty-five organ's-at-risk are included in this study, greatly expanding

beyond the limited number of OARs (ie. major salivary glands, spinal cord and brainstem) typically examined in the automatic segmentation literature for the head and neck region and described as a research need in a recent review article by Vrtovec et al.[11]. Another novel aspect of this work is that a large number of patient imaging and contouring data sets were obtained through publicly available data collections contributed to by multiple institutions. While many automatic segmentation investigations focus on datasets from a single institution, the use of CT datasets from multiple institutions presents a larger challenge to the accuracy of the training but is more relevant for real-world applications and may result in more robust segmentation performance across different institutions. This use of multi-institutional datasets was also identified as a research need in the recent review paper by Vrtovec et al. (as well as multi-modal imaging analysis which is not specifically addressed in the current work). Furthermore, this is the first study that provides analysis to investigate comprehensive clinical reasoning behind the predictions of a deep learning based medical image segmentation model. This is achieved through the use of Guided Gradient-weighted Class Activation Maps (Guided Grad-CAMs) that previously have been used to solely interpret image classification outcomes[12] but have not yet been applied to radiotherapy deep learning segmentation applications. Therefore, the current work makes

significant contributions to the early-stage development and adoption of artificial intelligence in clinical practice.

## 5.2 Material and Methods

### 5.2.1 Patients' Data

CT scans of the head and neck region delineated by expert radiation oncologists were acquired from the publicly available database TCIA that aggregated medical images from multiple institutions[13-121]. The total number of patients involved in this study was 964. Figure 5.1 details the absolute number of patients who had a particular organ-at-risk contoured. In total, the number of delineated anatomical structures was 25. Figure 5.1

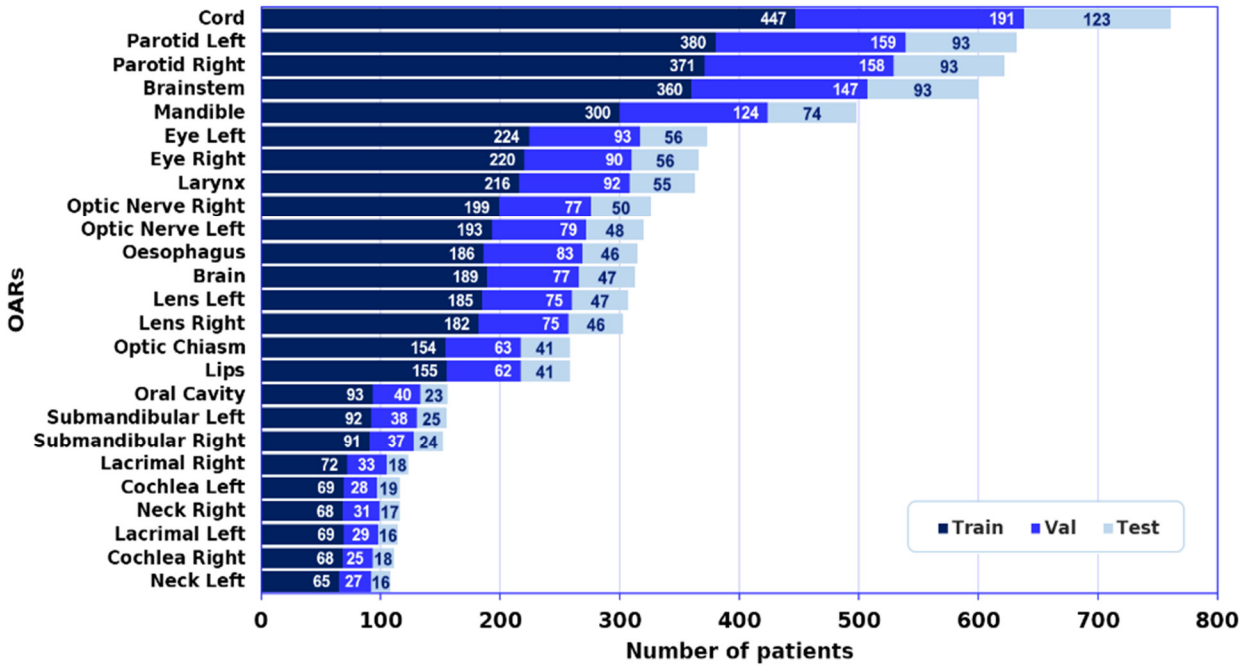



**Figure 5. 1.** The number of patients from the TCIA database whose CT scans and contouring data were used in the current work (for a test, validation, and training). Although the total number of patients considered in this study was 964, the number and type of structures contoured varied significantly from patient to patient within the database.

also shows the proportion of training, validation, and test data was approximately 65%, 25%, and 10% respectively. In total the data came form 9 institutions as specified in Table 5.1.

**Table 5.1**. Institutions contributing the data that were used in this study.

| **Institution** | **Number of patients** |
|---|---|
| Princes Margaret Cancer Centre | 300 |
| MD Anderson Cancer Center | 173 |
| Maastricht University Medical Center | 137 |
| Centre Hospitalier Universitaire de Sherbrook | 97 |
| Washington University St. Louis, School of Medicine | 96 |
| Hospital General Juif de Montreal | 89 |
| University of Miami, Miller School of Medicine | 31 |
| Hospital Maisonneuve-Rosemont de Montreal | 30 |
| Centre Hospitalier de l'Universite de Montreal | 11 |
| **TOTAL** | 964 |

### 5.2.2 Imaging and Segmentation Data Processing

CT images were pre-processed by the application of mask-invariant, contrast-limited, adaptive histogram equalization (CLAHE) to improve the soft tissue contrast (Figure 5.2). Additionally, each scan was augmented using random shift, scale, and rotate transformations with 15%, 20%, and 10% alteration limits, respectively, using the Albumentations software library[22]. The amount of additional augmented data generated was determined for every structure individually as follows. First, the number of CT slices per structure was multiplied by the number of patients who had the considered structure

contoured in order to calculate the total number of labeled CT slices ('ε') for a given structure. For ε<600, or 600<ε<3000, or ε>3000, the amount of existing data was multiplied by 12, 6, and 3, respectively, through augmentation. This approach was found to be sufficient in terms of limiting overfitting and maintaining the time efficiency of model training.

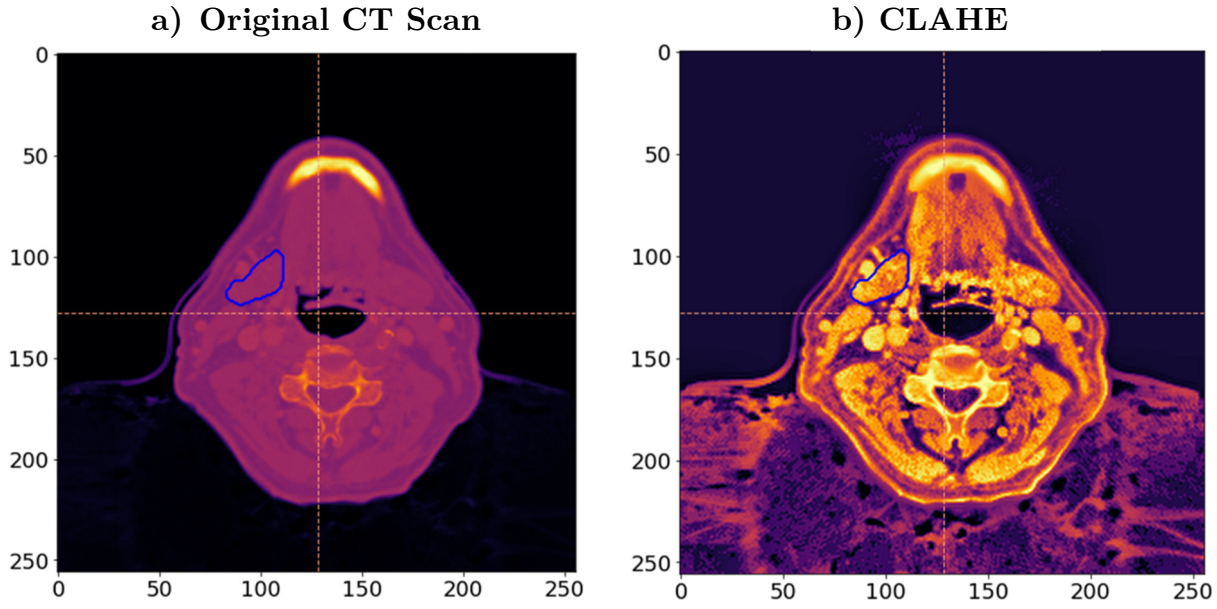



**Figure 5. 2. a)** Original CT scan with the contour of an example structure (right submandibular gland); **b)** The same CT scan after pre-processing with a contrast-limited, adaptive histogram equalization algorithm (CLAHE).

The images were cropped from 512x512 pixels into 256x256 symmetrically relative to the image's center of mass. This approach ensured that the relevant parts of patients' anatomy and associated contours were within the field-of-view.

Binary masks for anatomical contours were subjected to simple morphological operations to properly represent the original DICOM contours created by the radiation oncologist. The operations included filling the empty regions inside the contour, morphological closing

as well as the combination of Gaussian smoothing and Otsu's thresholding. The impact of each operation on the mask is shown in the Appendix 8.2 (Figure 8.1 and 8.2). Also, masks were augmented in the same way as CT scans i.e., using random shift, scale, and rotate transformations, also through the Albumentations software.

### 5.2.3 Model Architecture

In this work, a 2D U-net network with embedded Inception-ResNet-v2 blocks was used for automated segmentation of head and neck CT scans. In general, U-net is a symmetrical architecture that includes the encoder (contracting path) followed by the decoder (expansive path). The purpose of the encoder network is to transform the input image into feature representations using convolution blocks at multiple resolution levels. Specifically, after all the convolution operations are completed at a given resolution level the feature maps are downsampled by a max-pooling operation. As a result, the depth of the feature map is doubled while its width and height are halved. In this study, the U-net architecture consisted of six resolution levels that consecutively down-sampled the 256x256 input image into an 8x8 feature map during the contracting path. The purpose of the decoder is to up-sample feature maps to the resolution level of the input image while recovering their spatial representation. This is accomplished in a few consecutive steps. First, feature maps are up-sampled with transposed convolution, then concatenated with feature maps derived from the encoding network at the same resolution level which is followed by two convolution layers each connected to a rectified linear unit (ReLU). Those operations are repeated until the input image resolution is reached. The next step involves the application of $1 \times 1$ convolution to reduce the depth of the last feature map

to match the depth of the segmentation mask. Finally, to obtain the binary segmentation mask the sigmoid activation function is applied to that feature map[23].

In the U-net architecture proposed in this study the encoder's learning units (convolution layers at each resolution level) were replaced with Inception-ResNet-v2 units[24] i.e., the combination of residual[25] and inception[26] blocks. The rationale behind introducing the new blocks was that the residual units are able to avoid the vanishing gradient problem and decrease model training time, while the inception units increase the depth of the network which improves its ability to learn more complex representations of image-to-mask mapping. As mentioned in the recent review article by Fu et al. the application of both Inception and ResNet modules has been studied in terms of CNN architecture modifications in limited number of publications, however not in the context of combining them with U-Net architecture for automatic segmentation[10].

### 5.2.4 Training Parameters

The model was trained over 100 epochs with a batch size of 32 and an adaptive learning rate optimizer (Adam)[27]. The loss function was created by the combination of Jaccard Index[28] and Binary Cross Entropy[29] (equally-weighted addition). The maximization of the first metric ensures the overlap between model generated and target segmentations (global alignment) while the maximization of the second metric facilitates a fine, pixel-wise alignment between two binary masks (i.e. local alignment). The model was implemented in Python using a Tensorflow framework[30,31] with Keras library[32] and trained on a single NVIDIA Tesla V-100-SXM2 GPU or Tesla P-100-PCIe GPU. The model was trained on each anatomical structure separately so that in total 25 model instances were created.

### 5.2.5 Model Evaluation

The performance of the model was evaluated on testing data sets with several well-known metrics[11,33] that measured the similarity between the OAR segmentations generated by the model and the corresponding contours delineated by an expert radiation oncologist. These quantitative metrics included the Dice Score[34], Jaccard Index[28], and the Hausdorff Distance (regular, mean, 95-percentile, 5-percentile)[35]. The time required for the model training and inference was also reported. Additionally, we have provided visual examples of CT scans contoured by both radiation oncologists and the developed model, for several OARs. Figures were generated using Python's Matplotlib package.

### 5.2.6 Model Interpretability

The heatmap of the pixels in the image that contributed the most to prediction masks generated by the model was used for model interpretation. In particular, to visualize the contributions we used the combination of two methods proposed by Selvaraju et al.[12]. The first method is a pixel-space gradient visualization based on the guided backpropagation which provides high-resolution maps highlighting fine-grain details in the image but is not class-discriminative. The second method is based on Gradient-weighted Class Activation Mapping (Grad-CAM) which in contrast is highly class-discriminative but does not enable visualization of fine-grained details. The combination of those two approaches allows obtaining both high-resolution and class-discriminative Guided Grad-CAM visualizations of pixel-wise contributions.

Selvaraju et al.[12] indicated that their class-discriminative localization technique can generate visual explanations for any CNN-based network, and have been commonly applied for segmentation, classification, detection and feature extraction tasks of other

specific regions, imaging modalities (both medical and non-medical) and various deep learning algorithms[36-40]. However, this is the first study that provides this interpretability for segmentation of multiple critical structures on computed tomography scans of head and neck area achieved with U-net networks.

## 5.3 Results

### 5.3.1 Quantitative Evaluation

Figure 5.3 shows the values of Dice Score calculated on the testing data for all the organs-at-risk considered in this work. The Dice Score averaged over all structures and patients was 0.82±0.10. The average Dice Score for structures that had a score higher and lower than the overall average was 0.88±0.04 and 0.73±0.09 respectively. The higher of these two average Dice Scores was associated with organs that were supported by 428 patients (i.e. the average number of patients having a Dice Score *above* the average Dice Score was 428) while the lower score was for structures supported by 202 patients (i.e. the average number of patients having a Dice Score *below* the average Dice Score was 202).

Figure 5.4 illustrates the values of mean Hausdorff Distance ($HD_{mean}$) that were determined on the testing data for all the organs-at-risk. The $HD_{mean}$ averaged over all structures and patients was 1.51±1.17 mm. The average $HD_{mean}$ for structures that had a score higher and lower than the overall average was 1.82±1.22 mm and 0.65±0.15 mm respectively. The higher of those two measures was for structures that were supported by 255 patients while the lower score was for organs-at-risk supported by 362 patients.

The numerical values of mean, standard deviation, and median for Dice Score and Hausdorff metrics as well as the remaining metrics are listed in Table 5.2.

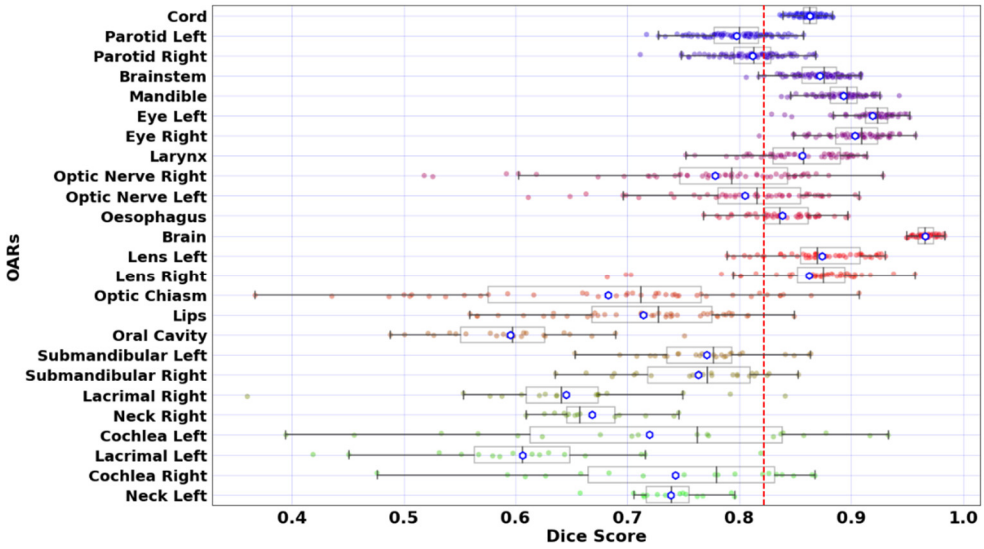



**Figure 5. 3.** A box plot for individual Dice Scores for all patients in the testing data set and for all critical structures. A single white dot denotes a mean value over all the patients for a given organ, while the dashed, red line indicates the mean value over all the patients and organs. The neck lymph nodes (on the right and left side of the neck) are denoted as Neck Left and Neck Right.

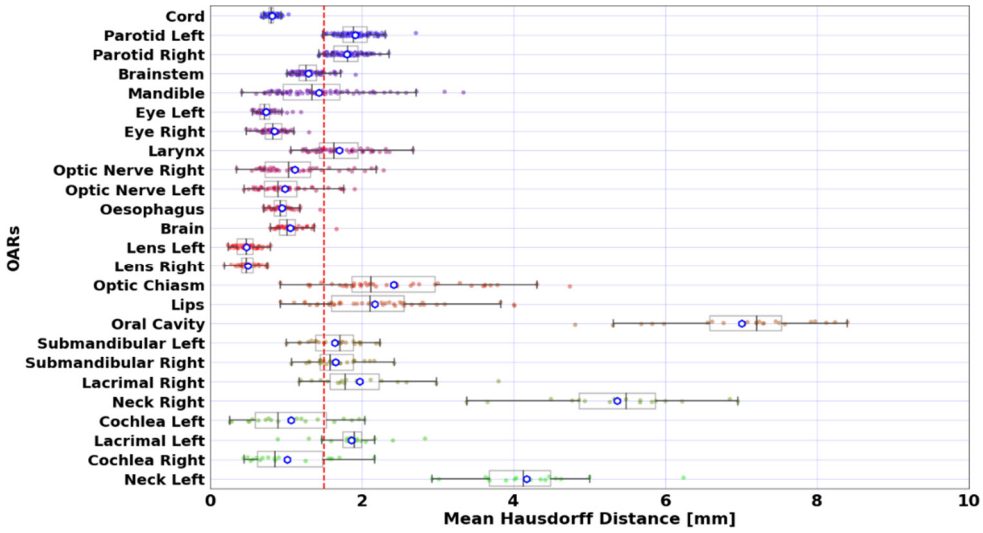



**Figure 5. 4.** A box plot for individual mean Hausdorff Distances for all patients in the testing data set and for all critical structures. A single white dot denotes a mean value over all the patients for a given organ, while the dashed, red line indicates the mean value over all the patients and organs. The neck lymph nodes (on the right and left side of the neck) are denoted as Neck Left and Neck Right.

**Table 5.2.** Mean, standard deviation and median for all model evaluation metrics.

| Structure | Statistical Metric | Dice Score | Dice score comparison to (Rhee et al)[43] | HD [mm] | Jaccard Index | $HD_{mean}$ [mm] | $HD_{mean}$ [mm] comparison to (Rhee et al)[43] | $HD_{5\%}$ [mm] | $HD_{95\%}$ [mm] |
|---|---|---|---|---|---|---|---|---|---|
| **Brainstem** | Mean (±SD) | 0.87 (0.10) | 0.864 (0.079) | 3.3 (2.0) | 0.78 (0.13) | 1.3 (0.8) | 8.8 (19.6) | 1.0 (0.7) | 2.9 (1.9) |
| | Median | 0.90 | | 2.9 | 0.81 | 1.1 | | 1.0 | 2.2 |
| **Brain** | Mean (±SD) | 0.97 (0.07) | 0.984 (0.003) | 5.7 (7.2) | 0.94 (0.09) | 1.1 (1.2) | 12.3 (15.2) | 0.5 (0.9) | 3.5 (5.2) |
| | Median | 0.98 | | 2.9 | 0.97 | 0.7 | | 0 | 1.9 |
| **Cochlea Left** | Mean (±SD) | 0.72 (0.20) | 0.652 (0.119) | 2.6 (1.9) | 0.59 (0.21) | 1.1 (0.7) | 4.7 (3.6) | 0.8 (0.7) | 2.4 (1.7) |
| | Median | 0.81 | | 1.9 | 0.67 | 0.8 | | 1.0 | 1.5 |
| **Cochlea Right** | Mean (±SD) | 0.74 (0.16) | 0.676 (0.131) | 2.7 (2.4) | 0.61 (0.17) | 1.0 (0.7) | 4.5 (3.5) | 0.9 (0.5) | 2.4 (2.2) |
| | Median | 0.78 | | 1.9 | 0.64 | 0.8 | | 1.0 | 1.7 |
| **Cord** | Mean (±SD) | 0.86 (0.08) | 0.834 (0.064) | 1.9 (2.0) | 0.77 (0.11) | 0.8 (0.5) | 7.1 (13.1) | 0.6 (0.6) | 1.7 (1.9) |
| | Median | 0.88 | | 1.9 | 0.78 | 0.7 | | 1.0 | 1.4 |
| **Eye Left** | Mean (±SD) | 0.92 (0.07) | 0.888 (0.037) | 1.9 (0.9) | 0.86 (0.10) | 0.7 (0.4) | 4.9 (14.2) | 0.5 (0.5) | 1.6 (0.8) |
| | Median | 0.94 | | 1.9 | 0.88 | 0.6 | | 1.0 | 1.4 |
| **Eye Right** | Mean (±SD) | 0.90 (0.10) | 0.893 (0.036) | 2.2 (1.9) | 0.84 (0.13) | 0.8 (0.5) | 3.5 (1.0) | 0.6 (0.6) | 1.8 (1.5) |
| | Median | 0.94 | | 1.9 | 0.88 | 0.7 | | 1.0 | 1.4 |
| **Lacrimal Left** | Mean (±SD) | 0.61 (0.22) | N/A | 5.4 (3.2) | 0.47 (0.21) | 1.9 (1.0) | N/A | 1.4 (1.0) | 4.9 (3.1) |
| | Median | 0.66 | | 4.2 | 0.49 | 1.6 | | 1.0 | 4.0 |
| **Lacrimal Right** | Mean (±SD) | 0.66 (0.21) | N/A | 8.7 (19.2) | 0.53 (0.22) | 1.9 (1.4) | N/A | 1.2 (0.7) | 6.2 (9.9) |
| | Median | 0.71 | | 4.4 | 0.55 | 1.5 | | 1.0 | 3.7 |
| **Larynx** | Mean (±SD) | 0.86 (0.16) | N/A | 4.6 (4.2) | 0.77 (0.18) | 1.7 (1.5) | N/A | 1.3 (1.3) | 3.9 (3.9) |
| | Median | 0.91 | | 3.1 | 0.84 | 1.1 | | 1.0 | 2.7 |
| **Lens Left** | Mean (±SD) | 0.87 (0.07) | 0.729 (0.139) | 1.2 (0.5) | 0.78 (0.11) | 0.5 (0.3) | 4.2 (11.0) | 0.2 (0.4) | 1.1 (0.4) |
| | Median | 0.90 | | 1.0 | 0.81 | 0.4 | | 0 | 1.0 |
| **Lens Right** | Mean (±SD) | 0.87 (0.10) | 0.704 (0.145) | 1.3 (0.7) | 0.78 (0.13) | 0.5 (0.3) | 3.5 (6.7) | 0.3 (0.5) | 1.2 (0.6) |
| | Median | 0.89 | | 1.0 | 0.80 | 0.4 | | 0 | 1.0 |
| **Lips** | Mean (±SD) | 0.71 (0.20) | N/A | 10.4 (16.1) | 0.59 (0.21) | 2.2 (2.3) | N/A | 1.2 (1.4) | 7.4 (9.4) |

| Structure | Statistical Metric | Dice Score | Dice score comparison to (Rhee et al)[43] | HD [mm] | Jaccard Index | HDmean [mm] | HDmean [mm] comparison to (Rhee et al)[43] | HD5% [mm] | HD95% [mm] |
|---|---|---|---|---|---|---|---|---|---|
| | Median | 0.78 | | 5.7 | 0.64 | 1.4 | | 1.0 | 3.5 |
| **Mandible** | Mean (±SD) | 0.89 (0.12) | 0.868 (0.033) | 8.4 (17.0) | 0.82 (0.15) | 1.4 (3.4) | 12.8 (9.5) | 0.4 (1.1) | 6.3 (15.4) |
| | Median | 0.93 | | 2.2 | 0.87 | 0.6 | | 0 | 1.4 |
| **Neck Left** | Mean (±SD) | 0.74 (0.21) | N/A | 17.8 (19.1) | 0.62 (0.20) | 4.2 (5.7) | N/A | 2.4 (4.1) | 14.0 (16.3) |
| | Median | 0.80 | | 10.5 | 0.67 | 2.2 | | 1.0 | 7.1 |
| **Neck Right** | Mean (±SD) | 0.67 (0.27) | N/A | 19.0 (17.4) | 0.55 (0.25) | 5.4 (6.6) | N/A | 3.4 (5.0) | 15.7 (16.4) |
| | Median | 0.76 | | 11.7 | 0.62 | 2.9 | | 1.4 | 8.7 |
| **Oesophagus** | Mean (±SD) | 0.84 (0.17) | 0.807 (0.070) | 2.8 (3.6) | 0.75 (0.18) | 0.9 (0.7) | 10.8 (10.9) | 0.7 (0.8) | 2.4 (2.4) |
| | Median | 0.89 | | 1.9 | 0.80 | 0.7 | | 1.0 | 1.9 |
| **Optic Chiasm** | Mean (±SD) | 0.68 (0.21) | 0.407 (0.139) | 13.6 (24.0) | 0.55 (0.21) | 2.4 (1.5) | 9.6 (4.0) | 1.6 (1.3) | 6.3 (3.9) |
| | Median | 0.75 | | 7.6 | 0.60 | 2.1 | | 1.0 | 4.9 |
| **Optic Nerve Left** | Mean (±SD) | 0.81 (0.12) | 0.679 (0.092) | 4.5 (4.2) | 0.69 (0.15) | 1.0 (0.7) | 6.9 (7.7) | 0.5 (0.5) | 3.6 (3.7) |
| | Median | 0.83 | | 2.7 | 0.71 | 0.8 | | 0 | 2.0 |
| **Optic Nerve Right** | Mean (±SD) | 0.77 (0.17) | 0.693 (0.085) | 4.9 (4.5) | 0.66 (0.19) | 1.1 (0.9) | 7.4 (7.2) | 0.6 (0.6) | 4.1 (4.1) |
| | Median | 0.83 | | 3.1 | 0.71 | 0.8 | | 1.0 | 2.2 |
| **Oral Cavity** | Mean (±SD) | 0.60 (0.25) | N/A | 19.0 (10.4) | 0.47 (0.25) | 7.0 (4.3) | N/A | 5.2 (4.3) | 16.7 (10.2) |
| | Median | 0.65 | | 17.2 | 0.48 | 5.8 | | 3.5 | 14.6 |
| **Parotid Left** | Mean (±SD) | 0.80 (0.15) | 0.826 (0.064) | 6.6 (4.8) | 0.68 (0.17) | 1.9 (1.0) | 13.4 (5.6) | 1.2 (0.7) | 5.4 (3.7) |
| | Median | 0.84 | | 5.7 | 0.73 | 1.7 | | 1.0 | 4.3 |
| **Parotid Right** | Mean (±SD) | 0.81 (0.14) | 0.827 (0.048) | 6.9 (5.5) | 0.70 (0.17) | 1.8 (1.1) | 13.9 (5.8) | 1.2 (0.7) | 5.5 (4.7) |
| | Median | 0.86 | | 5.5 | 0.75 | 1.5 | | 1.0 | 4.03 |
| **Submandibular Left** | Mean (±SD) | 0.77 (0.17) | N/A | 5.2 (7.1) | 0.65 (0.19) | 1.7 (1.2) | N/A | 1.2 (0.8) | 4.2 (3.6) |
| | Median | 0.83 | | 3.5 | 0.70 | 1.3 | | 1.0 | 2.9 |
| **Submandibular Right** | Mean (±SD) | 0.76 (0.18) | N/A | 5.2 (6.8) | 0.64 (0.20) | 1.7 (1.1) | N/A | 1.2 (0.8) | 4.3 (3.8) |
| | Median | 0.83 | | 3.5 | 0.70 | 1.3 | | 1.0 | 2.9 |

The OAR results for Dice Score and Hausdorff distances are comparable to those of other deep learning studies for head and neck anatomy segmentation summarized in the recent review paper by Fu et al.[10]. However, since those studies were applied solely on the 2015 MICCAI Head and Neck Auto-Segmentation Challenge datasets (and only included nine OARs), our results are not exactly comparable. We provide a more detailed comparison of our results to those of the work of Rhee et al.[43], which utilizes the largest number of CT training datasets (1169 head and neck datasets) in deep learning methods for head and neck segmentation published to date, and also includes a reasonably high number of OARs examined (16 vs the 25 examined here), included in Table 5.2. Even though the work of Rhee et al. [43] utilized two CNN models (V-Net for small structures and FCN-8s for large structures), and the current work utilized a single CNN, we exceed those results for every OAR except brain, left parotid, and right parotid. As expected, the standard deviations for the current work are somewhat higher than many previous publications, most likely due to the multi-institutional nature (and therefore higher contouring variability) of our training CT datasets.

Figure 5.5 demonstrates how the number of samples available and used for training the proposed U-net model affects its performance. A logarithmic trendline was used since it was found to be the best fit to the data amongst considered log, linear, exponential, and power-based options, but is not based on a demonstrated theoretical relationship. Overall, both evaluation metrics show that the model performance degrades as the amount of available data decreases, as expected.

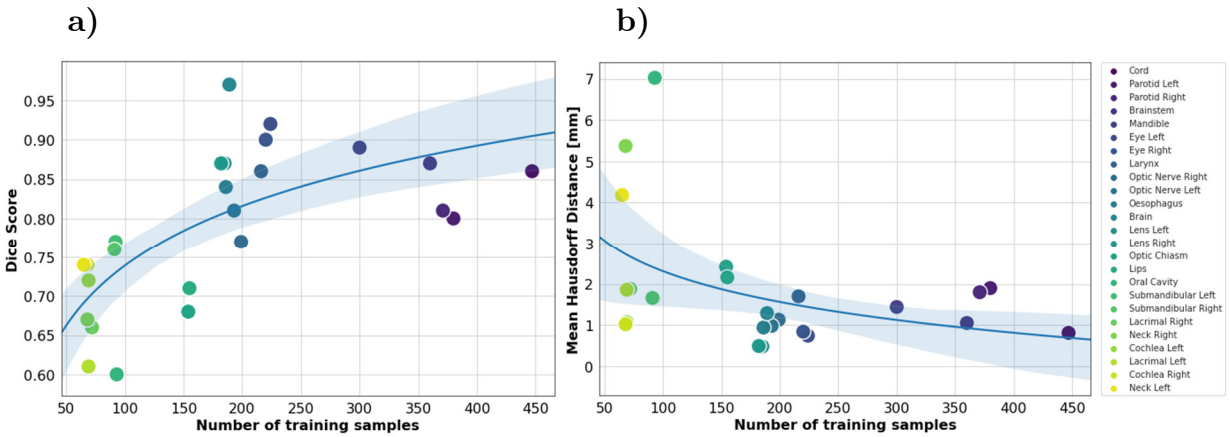
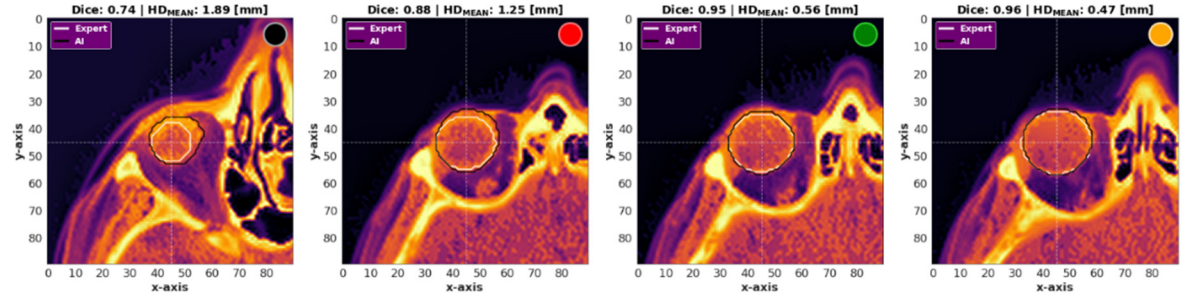



**Figure 5. 5.** The impact of the dataset size of individual OAR structures available for training on the model performance as measured by **a)** Dice Score and **b)** Mean Hausdorff Distance. The number of training samples is equal to the number of patients for which a particular structure was contoured. The logarithmic trendline is accompanied by 95% confidence intervals.

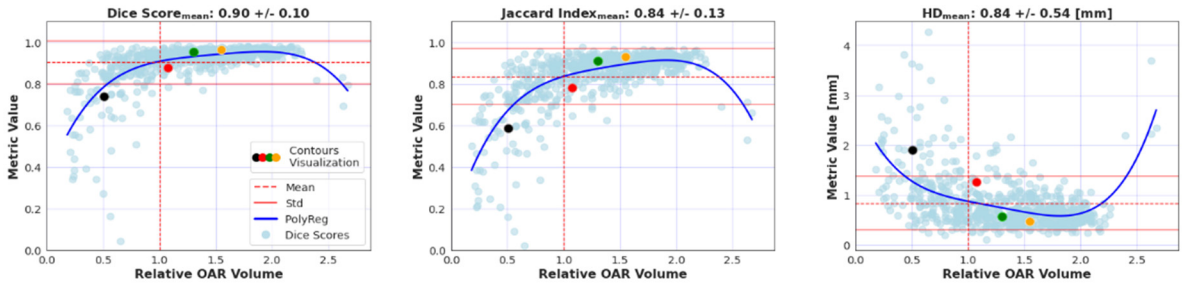
### 5.3.2 Visual Evaluation

In addition to quantitative metrics, we have also performed a visual assessment of contours generated by the implemented neural network, for easier clinical interpretation of the results. The extensive collection of selected CT scans with overlaid expert and AI contours was included in Appendix 8.3 for all the OARs. As an example, Figure 5.6 shows the visual comparison of three organs (right eye, right parotid, and optic chiasm) accompanied by graphs that illustrate the relationship between the cross-sectional area of the contour on a single CT slice and the level of similarity between model-generated segmentations and the radiation oncologist delineated segmentations.

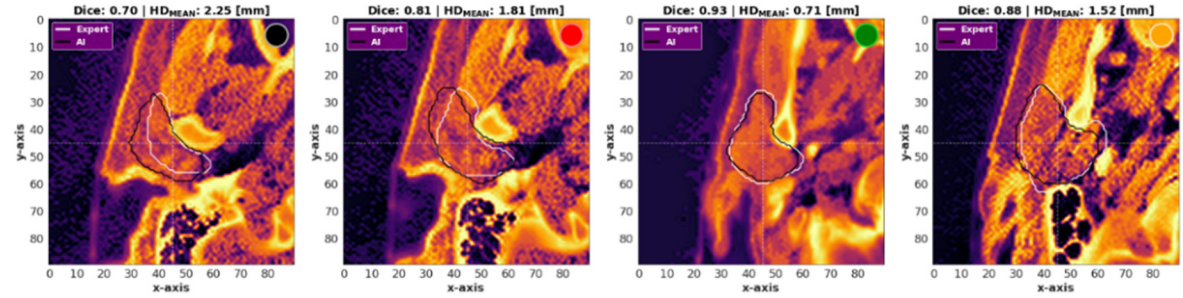
a)




b)




c)




d)

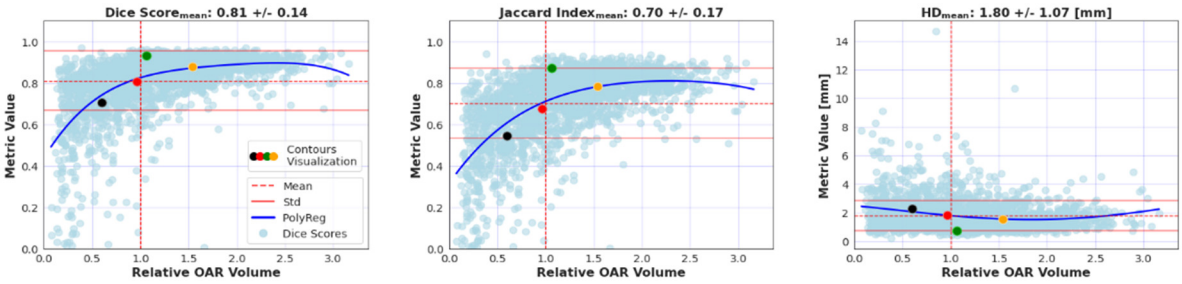



**Figure Continues on the next page →**

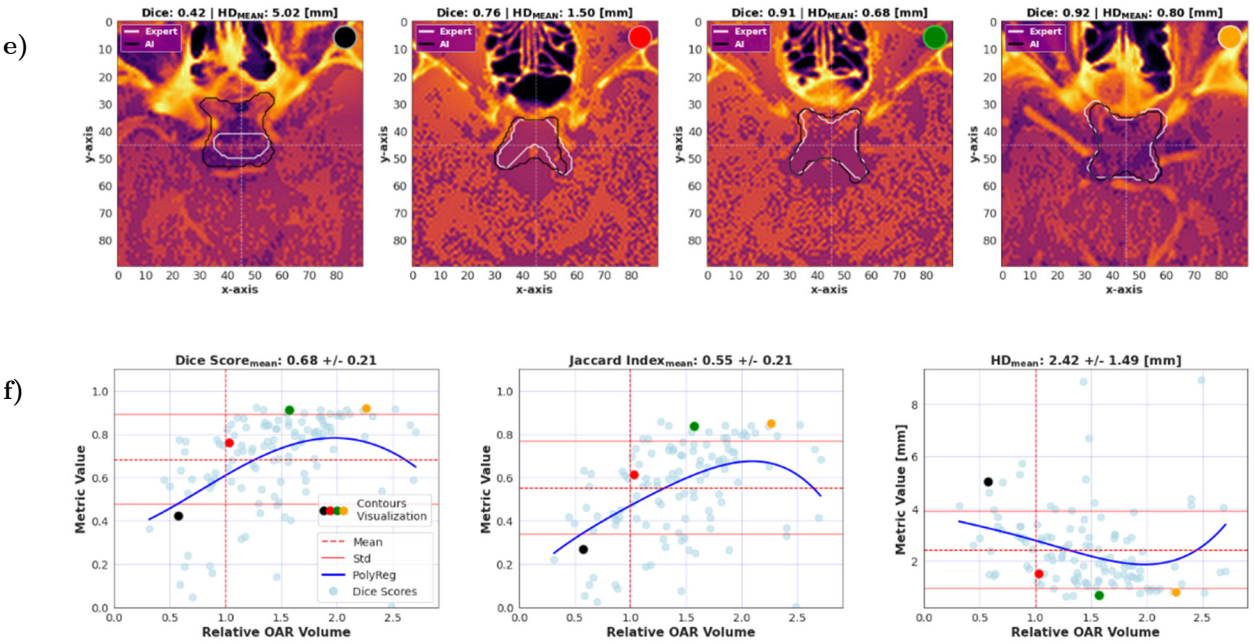



**Figure 5. 6.** Comparison of expert and AI-generated contours for four example patients demonstrating below-average, average, and above-average results for the right eye **a)** the right parotid gland **c)** and optic chiasm **e)**. Subfigures **b)**, **d)**, **f)** show three evaluation metrics as a function of mean-normalized voxel volume of a structure on a single CT slice in the axial-view. In general, the larger the cross-sectional area of OAR, the better the agreement between expert and AI segmentations. The subfigures b) and d) titles indicate the mean and the standard deviation of a given metric for all patients. The CT images and scatter plots are related via color labels located in the upper-right corner of subfigures a), c) and e). Visualizations for the remaining OARs are included in Appendix 8.3.

### 5.3.3 Training and Inference Time

On average the training and inference times were 23ms/CT slice and 31ms/CT slice, respectively. After accounting for the number of CT slices per given structure and the number of structures per patient (up to 25), the full segmentation of the head and neck region took 6.8 seconds. Model training time took 8 hours per anatomical structure on average. Detailed specifications on model training time and inference for each structure, as well as GPU usage are included in Appendix 8.2 (Table 8.4).

### 5.3.4 Model Explainability

The model explainability can be reported in the form of Guided Grad-CAMs highlighting those areas of the CT scan that the model was mostly focused on, in terms of pixel contribution, when performing a prediction about the segmentation mask for a given anatomical structure. Figure 5.7 shows several examples of these activation maps for the mandible, brain, left parotid gland, and oral cavity (Grad-CAMs for remaining structures are included in Appendix 8.3). It can be noticed that the model's rationale behind contouring an OAR is associated mainly with the location of the organ but also with information in more peripheral parts of the CT image (most apparent for oral cavity). Additionally, the pixel contribution gradients are significantly larger at the boundary of those delineated structures that exhibited high Dice Scores and low Hausdorff Distances compared to structures with low Dice Scores and high Hausdorff Distances. This indicates that more accurate contours are associated with the model having higher confidence in differentiating the contour mask of the targeted organ from the surrounding anatomy.

Another observation is that the more challenging it is to contour a particular organ the more peripheral regions of the anatomy are involved in making a prediction, which means that the model uses some information from parts of the image that are not in close proximity to the contoured structure. This is apparent, for example, when examining the activation map for the oral cavity and left parotid compared to maps for mandible and brain. The visualizations of model reasoning in Figure 5.7 also indicate that high tissue contrast improves the prediction accuracy, which also reflects clinical experience during manual OAR contouring (ie. better contrast results in more accurate segmentation).

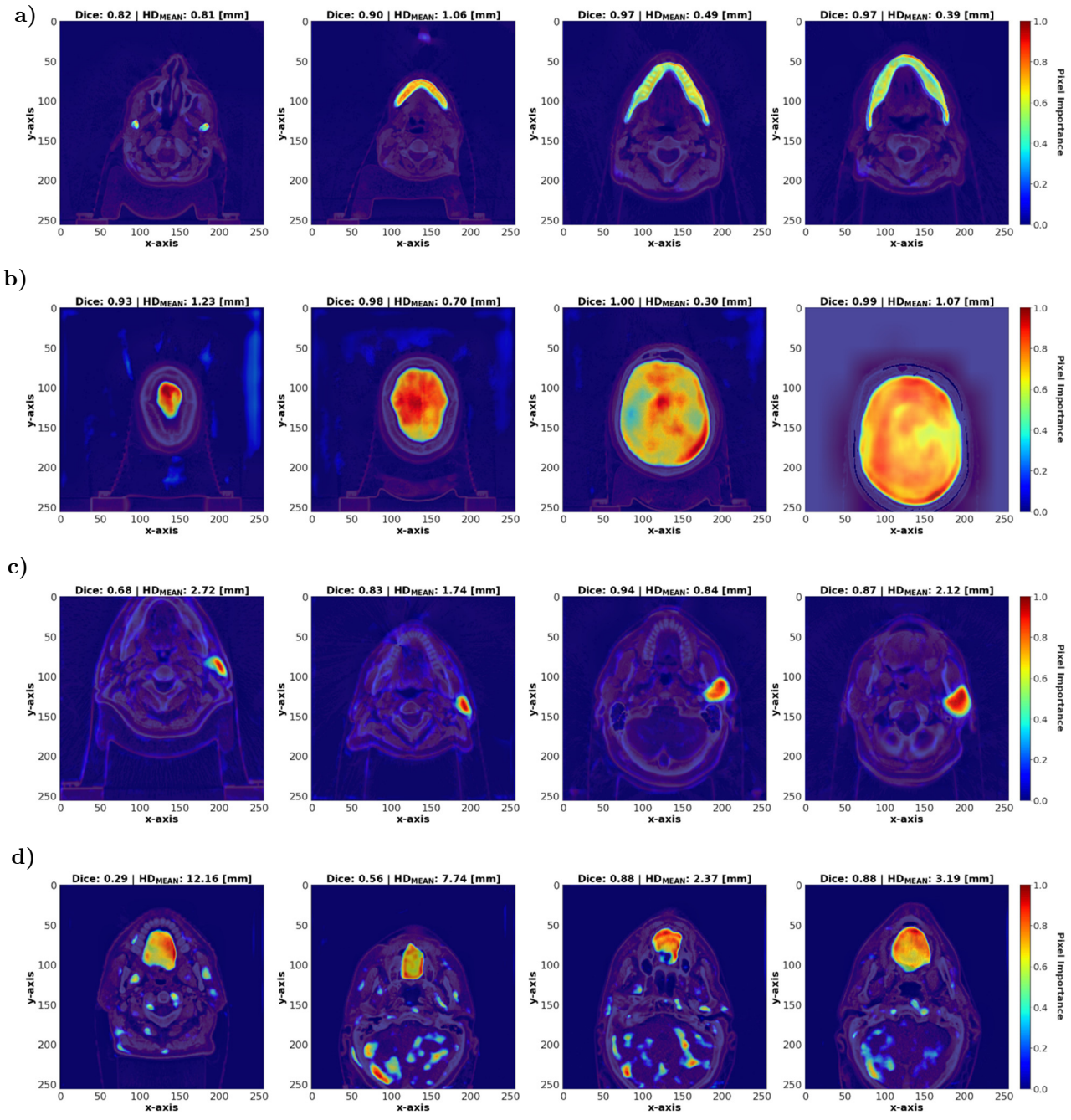



**Figure 5. 7.** Guided Gradient-Weighted Class Activation Maps for mandible **a)**, brain **b)**, left parotid gland **c)** and oral cavity **d)**. Dice Scores and mean Hausdorff Distances were calculated based on the visible 2D contours only. Grad-CAMs for remaining anatomical structures are included in Appendix 8.3.

## 5.4 Discussion

Model evaluation metrics shown in Figure 5.3 demonstrate that poorer results correspond to limited availability of the associated training data for those particular structures. This is also directly observed in Figure 5.4, but there are also few other factors that should be considered in understanding this relationship. The total volume of the organ, the cross-sectional area on a single CT scan, soft tissue contrast as well as large inter-observer and inter-institutional contouring differences also play a role. For example, in our study, the average volume of the organs with Dice Scores higher than the overall average was nearly two times (~1.8) larger than for the organs with Dice Scores lower than the overall average. This is a well known effect associated with the definition of Dice Score as a measure of the overlap between two volumes (or areas)[44]. For instance, a few millimetres shift between a large structure (such as the brain) will have a significantly lower impact on it's Dice Score compared to the same shift applied to a much smaller structure (such as the submandibular glands). Some exceptions to this trend are noted. For instance, left and right eye contours had a much higher Dice Score, compared to the lymph nodes even though the lymph node chains are significantly larger in volume. This is because eyes are much better defined in the CT scans (i.e. have a regular shape and high contrast relative to surrounding tissues). Lymph nodes, however, do not have sharp boundaries causing difficulties for automated segmentation[41].

Figures 5.3 and 5.4 also show large differences in Dice Scores and mean Hausdorff Distances between patients, for several structures (i.e. those with unsatisfactory contour similarity metrics). Relatively large standard deviation values reported in Table 5.2 confirm this as well. This is due to insufficient consistency in the segmentation of the

same OARs between patients. This issue is commonly encountered when the considered organ is surrounded by soft tissue. Their limited image contrast often introduces ambiguity that decreases the contouring consistency even further. The well-known inter-observer variability due to the difference in physician experience, expertise, and (sometimes) followed contouring protocols, additionally reduces segmentation stability.

The lowest Dice Score and highest Hausdorff Distance were observed for the oral cavity. Qualitative visual assessment revealed that this was due to a high variability in the oral cavity segmentation in the training datasets.

Despite these shortcomings, the overall model performance was satisfactory, since 16/22 of OAR results exceeded (in terms of dice coefficient it shows higher and more consistent performance) the average inter-physician variability described in the literature[11] while 19/22 exceeded or were within this variability range as summarized in the Figure 5.8. Those structures were also the most commonly contoured among head and neck cancer patients undergoing radiation therapy treatment. This amplifies a positive clinical impact of these results relative to poorly or inconsistently segmented structures that were delineated less frequently.

It is also challenging to compare segmentation results between individual studies because the results depend more on the quality of original raw clinical data rather than on the deep learning model, approach to data processing, or even the number of samples. Nevertheless, Figure 5.8 compares the Dice Scores obtained in this study with the ranges of Dice Scores resulting from inter-observer variability of radiation oncology clinicians.

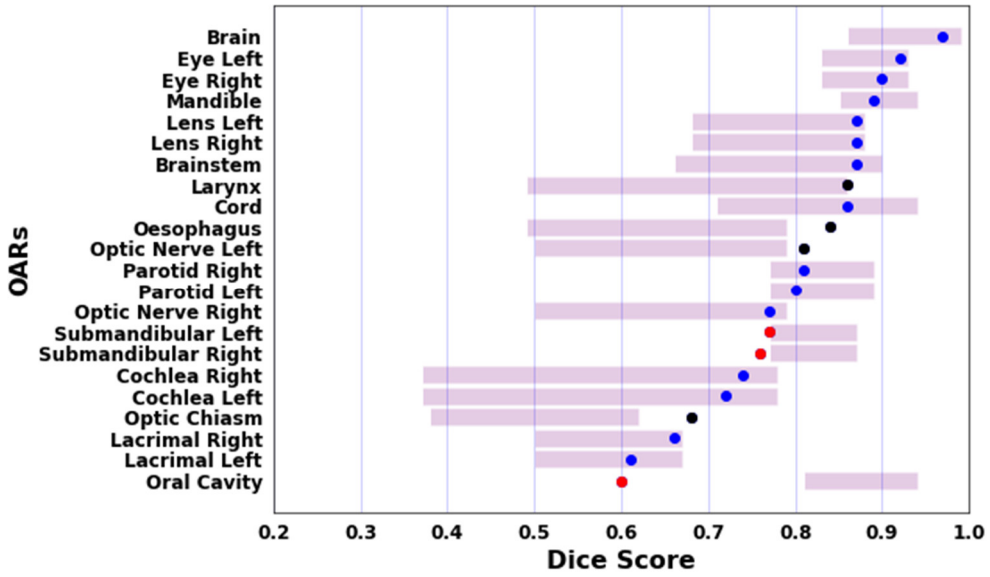



**Figure 5. 8.** Comparison of mean Dice Scores obtained in this study (color-coded dots) with the ranges of literature-based[11] interobserver variability (horizontal bars) observed among radiation oncology experts who manually contoured head and neck cancer patients. Those studies do not include the right and left sides of the neck, or lips; therefore, these structures are not included in the chart. Blue, red, and black dots define the dice score value within, worse, and better than the literature range respectively.

The ranges were derived from a large number of peer-reviewed publications on the subject matter that were summarized in the comprehensive review by Vrtovec et al.[11]. Overall, the comparison in Figure 5.8 shows that our results are within contouring deviations observed between expert physicians. Moreover, the majority of Dice Scores reported in this work are in the low end of uncertainty brackets (i.e. high Dice Score bracket). In the case of the brainstem, oesophagus, and left optic nerve the level of overlap between model generated and expert contours is even higher than it would be expected from the expert-to-expert comparison. Therefore, the deep learning auto-segmentation model evaluated in this work would be a feasible solution for a contouring task in the clinical environment. The only exception is the oral cavity for which, as

mentioned above, we obtained unsatisfactory results most likely due to the poor quality of the available training data.

While promising, this study has several limitations, which could be addressed in future work. First, the CT scans of each patient were delineated by one physician only, which potentially makes our results slightly biased towards a particular individual expert. The effect may be slightly reduced by the fact that the data used in this study came from multiple institutions, and thus were contoured by several experts. This common issue in medical image segmentation research is related to the fact that manual contouring of CT scans for radiation therapy purposes is very time-consuming. However, we have attempted to overcome this limitation by utilizing data from multiple institutions. We also compare our results to the results reported in the recent literature.

Furthermore, for many of the OARs considered in this study, the number of CT scan datasets was fairly limited even when accessing a large, publicly available database. We mitigated this challenge by performing image and contour augmentations which, although can not be considered a full substitute for additional clinical data, helped to reduce overfitting, and improved performance of the testing data[42],[45-48].

## 5.5 Conclusion

The purpose of this study was to utilize a deep learning model based on the U-Net architecture with embedded Inception-ResNet-v2 blocks to automatically contour critical organs on the CT scans of head and neck cancer patients who underwent radiation therapy treatment. CT datasets utilized for algorithm training were obtained from multiple institutions as opposed to a single institution. The number of OARs investigated is the

largest examined to date for head and neck anatomy and included several understudied structures. Reported results compare well with those found in the literature and are well-aligned within uncertainty related to interobserver variability between radiation oncologists responsible for anatomy segmentation in the clinical environment. The novel analysis of model interpretability for head and neck segmentation shows that the reasoning behind the algorithm's predictions is consistent with those important features that radiation oncology experts rely upon during OAR delineation. Additionally, the estimated total contouring time of 6.8 seconds for 25 structures makes the model a feasible solution for implementation in a busy clinical environment.

## 5.6 References

1. Kosmin M, Ledsam J, Romera-Paredes B, et al. Rapid advances in auto-segmentation of organs at risk and target volumes in head and neck cancer. 2019;135:130-140.

2. Conibear JR. *Assessment of Target Volume and Organ at Risk Contouring Variability within the Context of UK Head and Neck and Lung Cancer Radiotherapy Clinical Trials*, UCL (University College London); 2018.

3. Sandström H, Jokura H, Chung C, Toma-Dasu IJAO. Multi-institutional study of the variability in target delineation for six targets commonly treated with radiosurgery. 2018;57(11):1515-1520.

4. Aliotta E, Nourzadeh H, Siebers JJPiM, Biology. Quantifying the dosimetric impact of organ-at-risk delineation variability in head and neck radiation therapy in the context of patient setup uncertainty. 2019;64(13):135020.

5. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in auto-segmentation. Paper presented at: Seminars in radiation oncology2019.

6. van Rooij W, Dahele M, Brandao HR, Delaney AR, Slotman BJ, Verbakel WF. Deep Learning-Based Delineation of Head and Neck Organs at Risk: Geometric and Dosimetric Evaluation. *International Journal of Radiation Oncology* Biology* Physics.* 2019;104(3):677-684.

7. Zhu W, Huang Y, Zeng L, et al. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical physics.* 2019;46(2):576-589.

8. Gao Y, Huang R, Yang Y, et al. FocusNetv2: Imbalanced large and small organ segmentation with adversarial shape constraint for head and neck CT images. *Medical Image Analysis.* 2020;67:101831.

9. Tang, Hao, et al. "Clinically applicable deep learning framework for organs at risk delineation in CT images." Nature Machine Intelligence 1.10 (2019): 480-491.

10. Fu, Yabo, et al. "A review of deep learning based methods for medical image multi-organ segmentation." Physica Medica 85 (2021): 107-122.

11. Vrtovec, Tomaž, et al. "Auto-segmentation of organs at risk for head and neck radiotherapy planning: From atlas-based to deep learning methods." Medical physics 47.9 (2020): e929-e950.

12. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. Paper presented at: Proceedings of the IEEE international conference on computer vision2017.

13. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. 2013;26(6):1045-1057.

14. Bejarano T, De Ornelas Couto M, Mihaylov I. Head-and-neck squamous cell carcinoma patients with CT taken during pre-treatment, mid-treatment, and post-treatment dataset. the cancer imaging archive; 2018. In.

15. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of digital imaging.* 2013;26(6):1045-1057.

16. Bosch WR, Straube WL, Matthews JW, Purdy JA. Data from head-neck_cetuximab. *The Cancer Imaging Archive.* 2015.

17. Elhalawani H, Mohamed AS, White AL, et al. Matched computed tomography segmentation and demographic data for oropharyngeal cancer radiomics challenges. *Scientific data.* 2017;4:170077.

18. Vallieres M, Kay-Rivest E, Perrin LJ, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Scientific reports.* 2017;7(1):1-14.

19. Grossberg AJ, Mohamed AS, Elhalawani H, et al. Imaging and clinical data archive for head and neck squamous cell carcinoma patients treated with radiotherapy. *Scientific data.* 2018;5:180173.

20. Kwan JYY, Su J, Huang SH, et al. Radiomic biomarkers to refine risk models for distant metastasis in HPV-related oropharyngeal carcinoma. *International Journal of Radiation Oncology* Biology* Physics.* 2018;102(4):1107-1116.

21. Wee L, Dekker A. Data from Head-Neck-Radiomics-HN1. *The Cancer Imaging Archive,[Online] Available: https://doi org/10.* 2019;7937.

22. Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. Albumentations: fast and flexible image augmentations. *Information.* 2020;11(2):125.

23. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical image computing and computer-assisted intervention2015.

24. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint arXiv:160207261.* 2016.

25. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition2016.

26. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition2016.

27. Kingma DP, Ba JJapa. Adam: A method for stochastic optimization. 2014.

28. Jaccard PJNp. The distribution of the flora in the alpine zone. 1. 1912;11(2):37-50.

29. Khened M, Kollerathu VA, Krishnamurthi G. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Medical image analysis.* 2019;51:21-45.

30. Abadi M, Agarwal A, Barham P, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. 2016.

31. Abadi M, Barham P, Chen J, et al. Tensorflow: A system for large-scale machine learning. Paper presented at: 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)2016.

32. Chollet F. Keras. *https://kerasio.* 2015.

33. Taha AA, Hanbury AJBmi. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. 2015;15(1):29.

34. Dice LRJE. Measures of the amount of ecologic association between species. 1945;26(3):297-302.

35. Hausdorff F. *Set Theory.* Chelsea Publ.; 1991.

36. Baek S, He Y, Allen BG, et al. What does ai see? deep segmentation networks discover biomarkers for lung cancer survival. *arXiv preprint arXiv: 190311593.* 2019.

37. Diamant A, Chatterjee A, Vallières M, Shenouda G, Seuntjens J. Deep learning in head & neck cancer outcome prediction. *Scientific reports.* 2019;9(1):1-10.

38. Halicek M, Shahedi M, Little JV, et al. Head and neck cancer Detection in Digitized Whole-Slide Histology Using convolutional neural networks. *Scientific reports.* 2019;9(1):1-11.

39. Lee H, Yune S, Mansouri M, et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nature Biomedical Engineering.* 2019;3(3):173.

40. Liu F, Wang K, Liu D, Yang X, Tian J. Deep pyramid local attention neural network for cardiac structure segmentation in two-dimensional echocardiography. *Medical Image Analysis.* 2020;67:101873.

41. Gorthi S, Duay V, Houhou N, et al. Segmentation of head and neck lymph node regions for radiotherapy planning using active contour-based atlas registration. *IEEE Journal of selected topics in signal processing.* 2009;3(1):135-147.

42. Eaton-Rosen Z, Bragman F, Ourselin S, Cardoso MJ. Improving data augmentation for medical image segmentation. 2018.

43 . Rhee, Dong Joo, et al. "Automatic detection of contouring errors using convolutional neural networks." Medical physics 46.11 (2019): 5086-5097.

44. Taha, Abdel Aziz, and Allan Hanbury. "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool." BMC medical imaging 15.1 (2015): 1-28. 4a

45. Ho D, Liang E, Chen X, Stoica I, Abbeel P. Population based augmentation: Efficient learning of augmentation policy schedules. Paper presented at: International Conference on Machine Learning2019. 41

46. Lim S, Kim I, Kim T, Kim C, Kim S. Fast autoaugment. Paper presented at: Advances in Neural Information Processing Systems2019. 42

47. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *Journal of Big Data.* 2019;6(1):60. 43

48. Xie Q, Dai Z, Hovy E, Luong M-T, Le QV. Unsupervised data augmentation for consistency training. *arXiv preprint arXiv:190412848.* 2019 44

# Chapter 6

# Machine learning for dose-volume histogram based clinical decision-making support system in radiation therapy plans for brain tumors

The application of the knowledge derived from radiation therapy treatment plans that were delivered in the past to a certain group of patients was proven to be valuable in the process of creating new treatment plans for a similar group of patients. Adaptive radiation therapy focuses on creating new anatomy-adapted radiotherapy plans, therefore for ART, the application of knowledge about previously delivered plans is of high significance as well. In particular, the historical data about the dosimetric trade-offs between the radiation dose delivered to the critical organs and the dose coverage of the target volume could be utilized to identify patients who may require ART. The goal of the research presented in this chapter was to develop a machine learning model that, based on the historical DVH data of brain tumors treatment plans, would categorize new plans into those for which the PTV treatment planning objective was met and those for which the PTV objective was not met due to the priority trade-off made in order to meet one or more organs at risk constraints

(plans for patients who would likely require or benefit from ART). The study demonstrated that the trained model achieved satisfactory accuracy and not only could be used clinically for more efficient allocation of limited ART resources but can be used by medical physicists in a data-driven quality assurance program as well as by radiation oncologists to support their decision-making process in terms of treatment plan approval and potential plan modifications. This chapter has been published in the Journal of Clinical & Translational Radiation Oncology.

## 6.1 Introduction

In modern radiation therapy, the main steps in the treatment planning process are well established. After the treatment plan is created it is then routinely reviewed by a clinician to ensure that the treatment objectives are met and dosimetric trade-offs, when required, are at acceptable levels[1-3]. For those plans that may require a dosimetric trade-off, this sometimes complex decision-making process could benefit from the knowledge of similar plans that were developed, approved, and successfully delivered in the past to patients. The analysis of historical plans can provide valuable insights about dose-volume metrics, help to gain evidence-based confidence when approving new treatment plans as well as improve the efficiency and quality of the treatment plan[4-8]. However, in practice, the extraction, analysis, and interpretation of meaningful information from relevant historical data are very time-consuming and not achievable by radiation oncology professionals in the busy clinical environment. Machine learning (ML) helps to overcome those difficulties and can be used to assist medical physicists and physicians to make better informed, data-driven decisions in the radiation therapy process.

During recent years there has been growing interest in the application of ML models to develop quality assurance (QA) tools and support the treatment planning process. For example, Hirashima et al. and Wall et al. used XGBoost and Extra-Trees methods respectively to predict the gamma passing rate for patient-specific QA results for volumetric modulated radiation therapy (VMAT) plans[9,10]. Osman et al. utilized an artificial neural network for the prediction of the MLC leaf position deviations during dynamic IMRT treatment delivery using log file data[11]. ML has also been explored in many other QA applications in medical physics[12-16].

Machine learning-based enhancement of the treatment planning process has also been of strong recent interest. Knowledge-based planning (KBP) is a commonly studied application that leverages relevant features of previous, successfully delivered treatment plans in order to predict specific treatment planning parameters or the possible attainable dose-volume histograms (DVHs). Predicted DVHs for targets and OARs can then help radiotherapy planners to manually determine dose-volume constraints or become a part of an automated treatment planning workflow[17]. For example, Tambe et al. built a KBP model to reduce the variance among VMAT plans via the prediction of the minimum possible dose-volume metrics for advanced-stage lung cancer patients[18]. KBP has been successfully used across various clinical sites such as head and neck[19,20], prostate[21,22], lung[23-25], rectum[26,27], breast[28,29], pelvis[30], and brain[31].

The purpose of this study was to apply ML techniques to create a novel decision support application that has not been investigated before. Specifically, a machine learning model was trained to classify previously delivered VMAT plans of brain tumor patients into two categories created in collaboration with expert radiation oncologists. The first category contained plans

that met PTV treatment planning objectives. The second category included plans for which PTV objectives were not met due to the priority given to one or more OARs (i.e. a trade-off was required); those plans, however, were still clinically acceptable and delivered. Once trained, the ML algorithm would be able to indicate which new plans required a compromise (or not). This is a novel ML application and could have a very practical impact on the decision-making process in a clinical environment (more details about clinical use and future utility of the system are included in the discussion section). Furthermore, our study applied double nested cross-validation for model selection and tuning as well as comprehensive global and local model explainability analysis. In the literature, numerous studies use k-fold cross-validation[32], but few apply nested cross-validation[33-35], and none that are similar to our application. Model explainability is rarely performed in applications of ML in radiation therapy but has been identified as a strong need in the research community in order to properly interpret model results. The explainability analysis included in this work may be particularly valuable to both medical physics and radiation oncology professionals working in a clinical environment.

## 6.2 Material and Methods

### 6.2.1 Treatment Plans Data

This study involved 79 brain tumor patients that were prescribed a total dose of 60Gy delivered in 30 fractions, 2Gy per fraction using two-arc VMAT plans. The data necessary to train machine learning models were derived from dose-volume histograms and anatomical contours delineated by two experienced radiation oncologists. DVHs provided dosimetric information while segmentations provided geometric information for feature extraction. Structures considered for analysis included PTV, brainstem, left and right cochlea as well as optic chiasm.

### 6.2.2 Model Inputs

#### Dosimetric features

Figure 6.1a-b summarizes the dosimetry data and deviations from the treatment objectives for each structure. Forty-one plans met all dose objectives, while 12, 13, 9, and 4 plans did not meet one, two, three, and four dose objectives, respectively. Appendix 8.4 (Figure 8.3) shows DVHs for all plans and structures as well as indicates plans for which a specific number of treatment planning objectives were not met.

#### Geometric features

The first type of geometric features included the minimum distance $\Delta$d measured in millimetres between OARs and PTV calculated by the $5^{th}$ percentile of minimum Hausdorff distances (i.e. the set of minimum distances between two segmented structures)[36] as shown in Figure 6.1c. The second type of geometric feature was PTV volume measured in $cm^3$ as presented in Figure 6.1d.

It is also worth noting that some OAR objectives were not met due to other reasons than a trade-off, therefore instead of 41 plans mentioned above we have 54 plans for class-0 and, similarly, instead of 38 plans, we have 25 plans for class-1.

### 6.2.3 Model Outputs

The models in this study were trained for the binary classification task thus the model output was represented by the binary categorical variable. The value of 'zero' corresponds to plans for which the PTV treatment planning objective was met while the value of 'one'

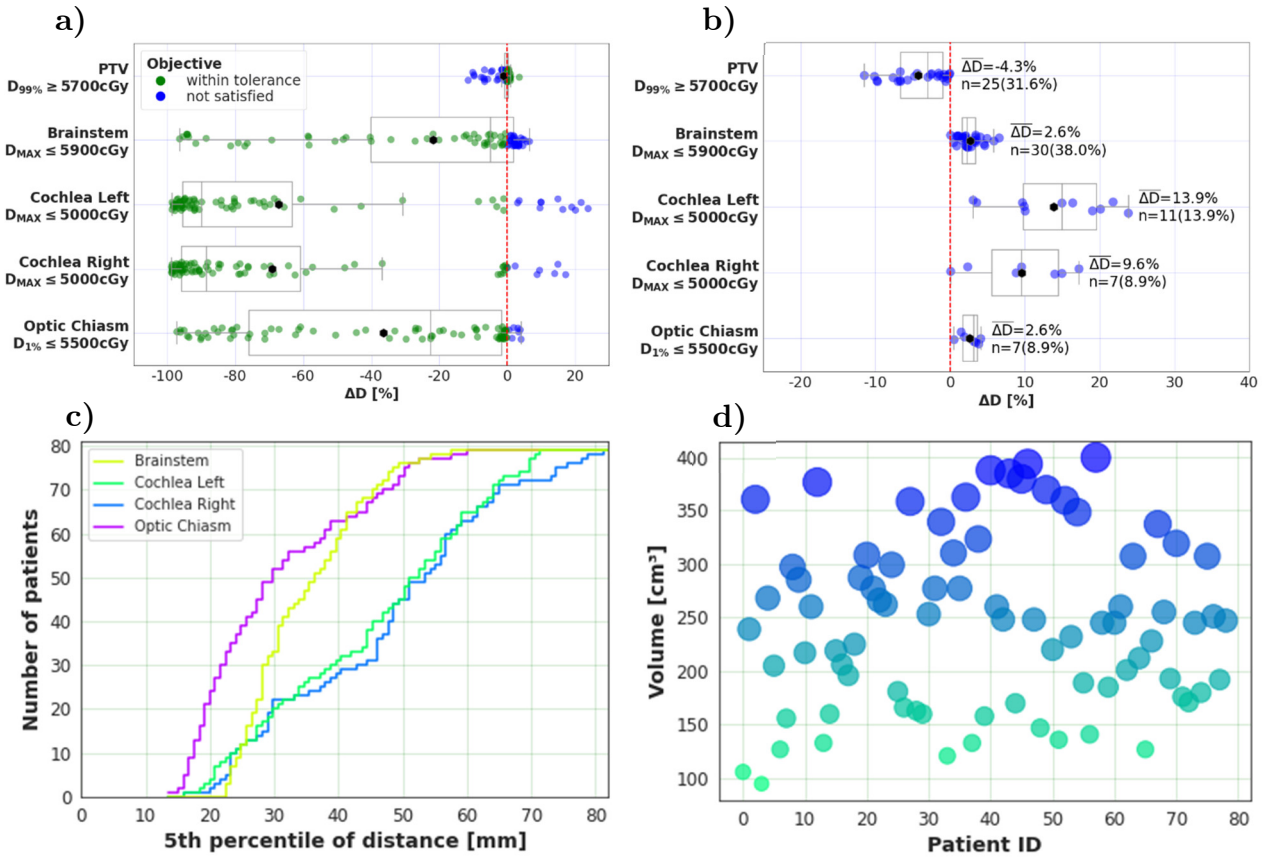



**Figure 6.1. a)** Box plots of the percentage deviations from treatment planning objectives ΔD for all 79 patients and associated structures, where the red dashed line indicates the boundary between positive and negative ΔD; **b)** box plots of percentage deviations from treatment planning objectives for plans who did not meet a specified objective. The percentage values of ΔD were used as dosimetric features for training the machine learning model. $\overline{\Delta D}$ correspond to the mean deviation for n plans. **c)** Cumulative distribution of the $5^{th}$ percentile of Hausdorff distances between the PTV and organs-at-risk. **d)** The absolute PTV volume for each plan. The size and the colors of the markers are proportionate and correspond to the PTV volume measure.

is associated with plans for which the PTV objective was not met due to the priority trade-off to meet one or more OAR constraints. This classification was established through independent review by two radiation oncologists. The proportion of plans belonging to

classes zero and one is 68.4% (54 plans) and 31.6% (25 plans) respectively. This class imbalance is not significant and was not observed to impact results.

### 6.2.4 Model and Hyperparameters Selection

Four models commonly used in ML were considered: Support Vector Machine Classifier, Elastic Net, Logistic Regression, and Random Forest Classifier (as implemented in the scikit-learn Python module[37]). *A priori* justification for use of particular models and their parameters is desirable from a scientific standpoint but is challenging in practice due to a large number of possible combinations of hyperparameters. Therefore, in this work, the classification algorithms were selected due to simplicity, computational efficiency, and common usage in the ML community.

Only one model was selected for further comprehensive analysis of the results on unseen testing data. The model selection process was based on the nested cross-validation technique that was shown to be superior relative to single cross-validation in minimizing the bias for model and hyperparameters selection as well as reducing overfitting[38,39]. Figure 6.2 shows the principle of operation of nested cross-validation. The training data for each fold in the outer loop is divided into training and validation sets within subsequent folds in the inner loop where hyperparameters are tuned and model performance reported as the best single score for one fold in the outer loop. Once the procedure is repeated over all folds in the outer loop, the mean score of all inner loop best scores is reported. The model with the best mean score is then selected for further analysis.

It can be noticed that even though nested cross-validation allows selecting the particular model it is often a case that the hyperparameters optimal for fold $n$ might not be the

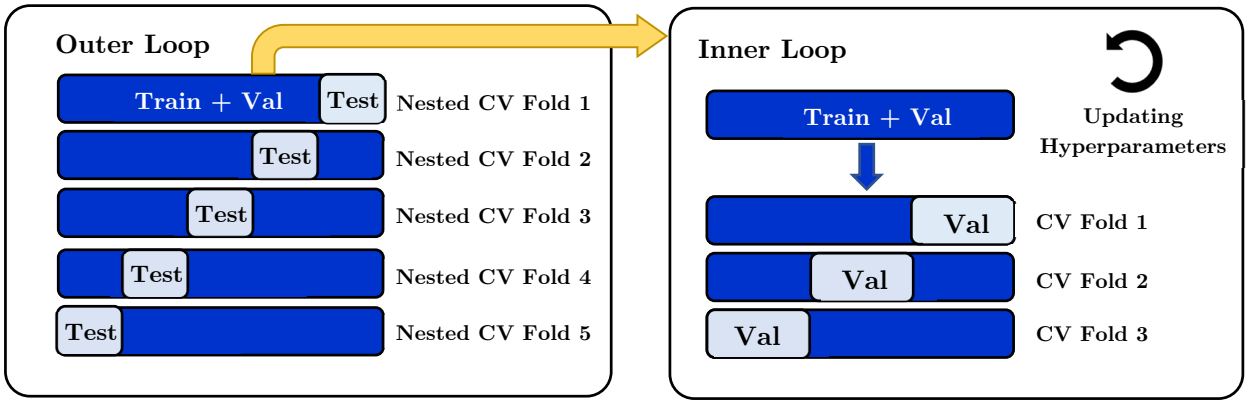



**Figure 6. 2.** Schematic illustrating the principle of operation of nested cross-validation. The cross-validation split in the outer and inner loop was conducted with stratified k-folds where k=5 for the outer loop and k=3 for the inner loop.

same as hyperparameters optimal for any other fold. For this reason, to select the best set of hyperparameters that would provide desirable model performance (low bias and low variance) the nested cross-validation for each set of best inner loop hyperparameters was performed. The test data (here in the outer loop) were not used for model and hyperparameter selection as they will be used for the final evaluation of the selected model using stratified k-fold cross-validation. Stratified split ensures that each target class in each fold contains approximately the same percentage of samples. The proportion of training to test data in the outer loop was 80/20 (5 folds) while for the inner loop was 67/33 (3 folds, due to a smaller number of available samples). The metric used for scoring the models was the area under the curve (AUC) of receiver operating characteristic (ROC)[37].

Hyperparameters and their ranges for selected models are specified in Appendix 8.4 (Table 8.5). The search for optimal hyperparameters was performed using a grid search method

that explores all the possible combinations of hyperparameters. For large data sets, it is often not a viable option, however for the sample size and number of features selected for this study, it was reasonably computationally efficient. The processor used for computations was a 2-core Intel Xeon CPU @2.30GHz.

### 6.2.5 Model Evaluation

The model with hyperparameters tuned after nested cross-validation was evaluated on the test data that accounted for 20% (15 or 16 plans depending on the fold as 79 is not equally divisible by 5) of all the data (79 plans) using single 5-fold cross-validation. For each fold, the areas under the ROC curves were reported. Additionally, the confusion matrix with true and predicted classes together with the precision, recall, and accuracy metrics (standard ROC definitions as in Pepe et al.[40]) for one of the cross-validation folds were also included in the results for more intuitive performance interpretation.

### 6.2.6 Model Explainability

To better understand the predictions generated by the model we analyzed the Shapely additive explanation (SHAP) interaction values[41] for both global and local explainability. The Shapley value reflects a mean value of marginal feature contributions to the prediction and can be interpreted as a contribution to explaining the difference between the average prediction of the model and the actual individual prediction.

## 6.3 Results

### 6.3.1 Model and Hyperparameters Selection

Recall the first step included the model selection based on the nested cross-validation. The mean and standard deviation of the ROC score was 0.9726 (±0.0059) for Support Vector Machine, 0.9986 (±0.0028) for Elastic Net, 0.9994 (±0.0012) for Logistic Regression, and 0.9979 (±0.0025) for Random Forest Classifier. The performance of each model is similar both in terms of mean and standard deviation, however, Logistic Regression received the highest score and showed the lowest SD for inner loop evaluation metrics. The computational time of nested cross-validation for model selection was 50 seconds.

The hyperparameters for the best performing model (ie. Logistic Regression) also were selected with a cross-validation technique. Only two sets of hyperparameters were evaluated in this step because during model selection those two sets were associated with the best score for more than one fold. The hyperparameters were: pipeline 1 - regularization L1, C=0.077, and liblinear solver achieving a mean AUC of 0.9666 (±0.0668), and pipeline 2 - regularization L1, C=1.668, and liblinear solver achieving mean AUC of 0.9898 (±0.0074). The computational time of nested cross-validation for hyperparameter selection was 3 seconds.

### 6.3.2 Model Evaluation

The mean accuracy of the Logistic Regression model selected was 93.8±4.1% while the mean area under the ROC curve was 0.98±0.02 on the testing data. Figure 6.3 shows the confusion matrix with precision, recall, and f1 scores for two classes and one (ie. fourth) cross-validation fold.

The fourth fold was selected as an example because the resulting accuracy is approximately the same as the average accuracy of the model. The performance measures reported in Figure 6.3 are similar for all the remaining folds and are included in Appendix 8.4 (Figure 8.4). Their definition and interpretation are included in Appendix 8.4 (Table 8.6) as well.

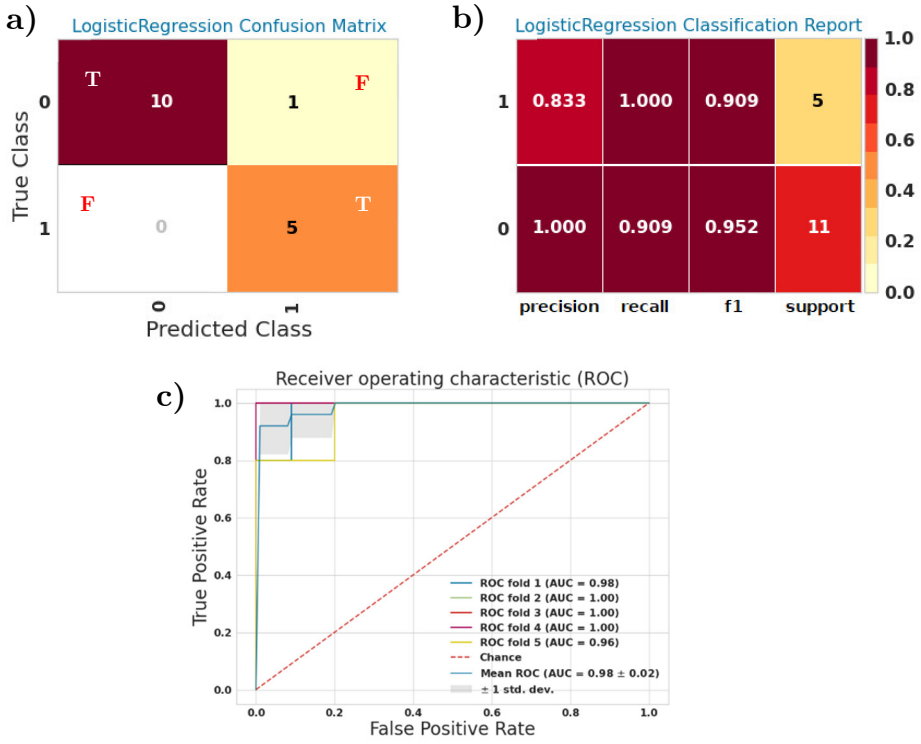



**Figure 6. 3.** Confusion matrices **a)** and Logistic Regression model evaluation metrics **b)** for fourth cross-validation fold. The meaning behind each metric was briefly summarized in Appendix 8.4 (Table 8.6); **c)** Receiver Operating Characteristic for five cross-validation folds created based on the testing data and the performance of the Logistic Regression model. The dashed diagonal line ('Chance' in the legend) represents the random assignment of classes.

Figure 6.3c shows the AUC values for ROC curves associated with each fold. All evaluation metrics were calculated based on the testing data.

### 6.3.3 Model Explainability

The logistic regression model explainability was addressed both globally and locally. Figure 6.4 shows global explainability as an impact of each feature on the model output. Figure 6.4a provides directional SHAP values and relative values of geometric and dosimetric model features. Figure 6.4b shows the average impact (ie. mean SHAP) of each feature on the model output. Overall, the SHAP analysis indicates that the deviation from the $D_{99\%}$ ($\Delta D_{99\%}$) metric for PTV had the greatest influence on the model predictions and is approximately 3.5 times larger (i.e. more important) than the second most important feature – $\Delta D_{1\%}$ for optic chiasm. The least important features were percentage deviations from the maximum dose delivered to the left and right cochlea. A more in-depth interpretation of this analysis is interesting from a clinical perspective and will be discussed further in the discussion.

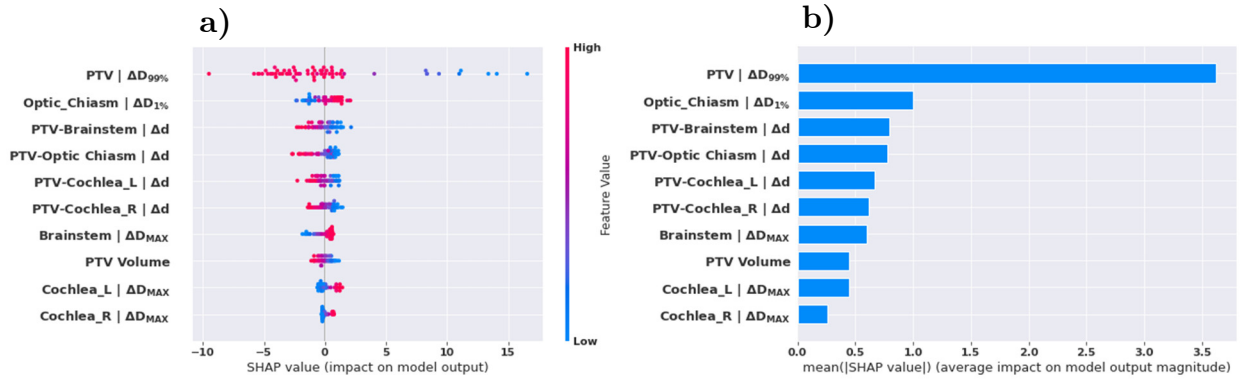
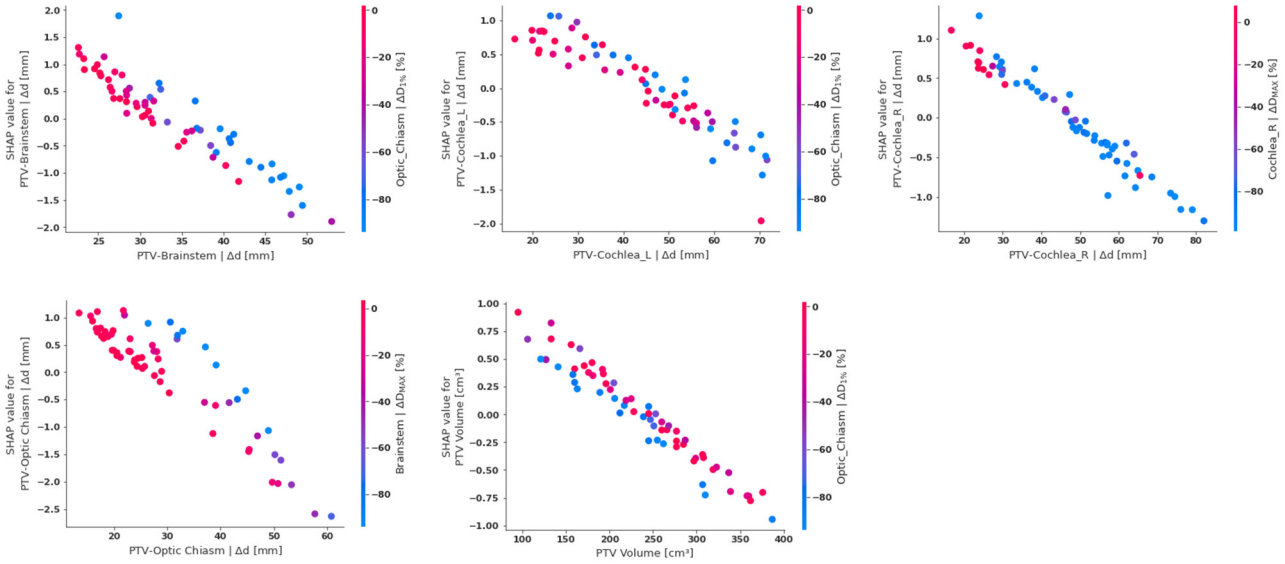



**Figure 6. 4. a)** Feature importance represented by the impact of directional SHAP values and particular feature values on the output of the model; **b)** The average feature contribution to the model output measured by mean absolute SHAP values.

Although individual SHAP values are important in identifying which features the model relies on when making a prediction, it is also very practical to determine the relationship

**a) Geometric Features**




**b) Dosimetric Features**

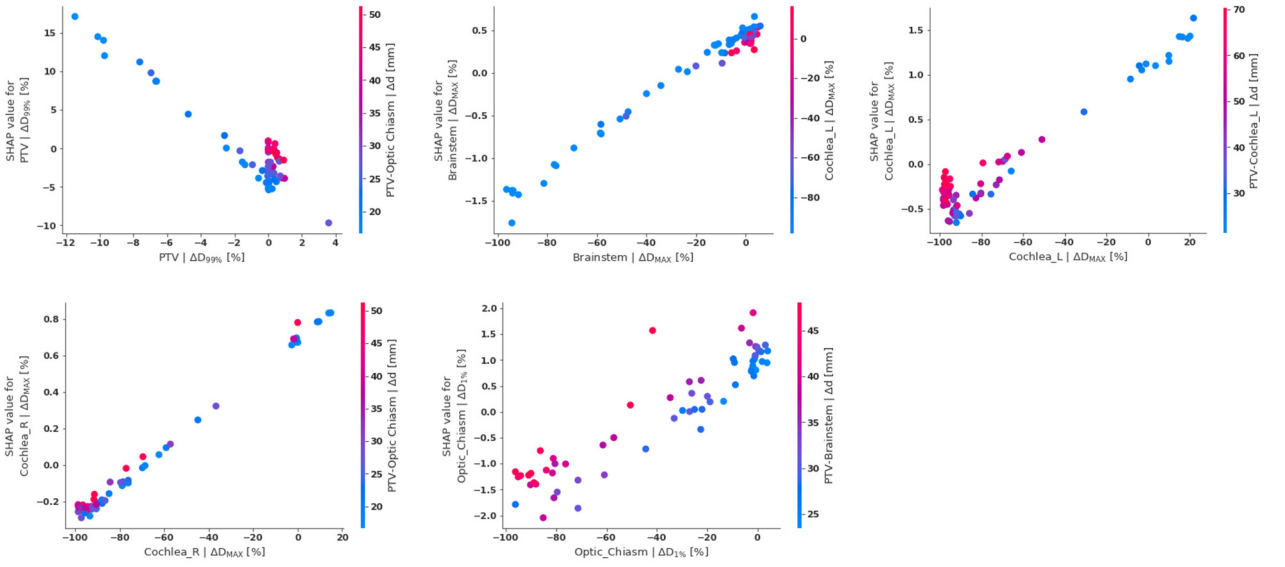



**Figure 6. 5.** Partial dependency charts for a) geometric features and b) dosimetric features. The color bars associated with each chart indicate the feature with which the evaluated feature (on the $x$-axis) has the strongest interaction. Specifically, the interaction indicates the influence those two features have on the model prediction. For example, if we consider a feature on the $x$-axis, then another feature on the color bar will be automatically selected in order to maximize the mutual impact of those two features on the model prediction. Partial dependency charts for all cross-validation folds are included in Appendix 8.5.

between original values (not scaled for the model training) of individual features and their global contribution to the model performance. Figure 6.5 shows these relationships for both geometric and dosimetric features of the Logistic Regression model.

Generally, for geometric features, it is seen that the larger the distance between the PTV and the OARs and the larger the PTV volume, the lower the SHAP value. Regarding the dosimetric features, for the majority of them, there is a linear and positively correlated relationship between the feature and its contribution to the model prediction. The only exception is $\Delta D_{99\%}$ for PTV; in this case, SHAP values are linearly but negatively correlated to this feature.

While the above analysis provides insights into the global model explainability, the model predictions can also be interpreted locally by examining individual model predictions as in Figure 6. In particular, the bar charts for each prediction show the feature importance in the form of directional SHAP values, the probability of the prediction belonging to class 0 and class 1 (ie. the class predicted by the model). The determination of whether the prediction was correct (or not) is also included. However, the practicality of this feature is limited to model testing. It can be noticed that out of 10 model features only 7 are present in the local explainability graphs. This is because the three remaining features were not significantly contributing to these individual predictions. Although Figure 6.6 shows the results of local model explainability for four example predictions, Appendix 8.5 illustrates local feature importance for all predictions.

## 6.4 Discussion

After executing double nested cross-validation of several models, the Logistic Regression algorithm was selected as the best performing model for further use. Double-nested cross-validation was employed instead of single-nested for model selection transparency. In terms of computational efficiency, this would also be a preferable method if the data set

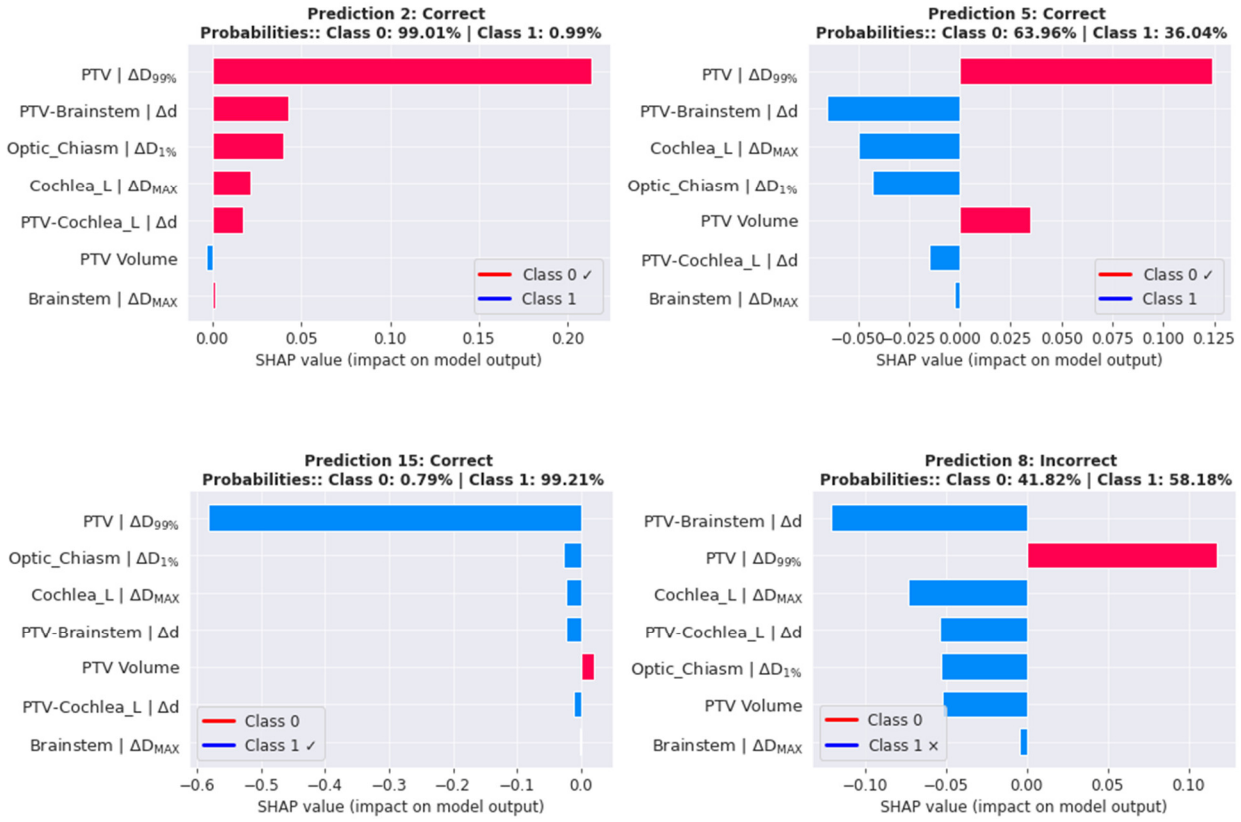



**Figure 6. 6.** Local explanations of the logistic regression model in the form of features importance and, generated by the model, binary class probabilities. For illustrative purposes, two randomly selected samples belonging to each class (including one incorrectly classified sample) were selected. The threshold for the class assignment was determined by the probability of 50%.

and the range of hyperparameters were larger. As for the Logistic Regression model, it is a relatively simple and easy-to-interpret model that can be trained and provide new predictions (model inference) quickly. The Logistic Regression model has also been of interest in recent radiation oncology research[42-45].

The model was selected based on its performance measured by the AUC, not accuracy. This is mainly because the ROC curve is insensitive to data sets with unbalanced classes and additionally reflects the classifier's performance for all values of the discrimination threshold. These characteristics make AUC a preferable metric in the evaluation of ML models[44-47]. However, because our study included imbalanced data we have also provided precision-recall curves in the Appendix 8.4. The model did not overfit the data because the model performance on testing data for each cross-validation fold is both satisfactory and consistent. These are promising results considering that the number of samples in the data set and the samples-to-features ratio were relatively limited.

The SHAP analysis results presented in Figure 6.4 requires further discussion. First, negative SHAP values do not mean that the feature importance is smaller than for positive SHAP values. Rather, SHAP values below zero drive the prediction towards class 0 while positive SHAP values drive the prediction towards class 1. This characteristic combined with the color-coded values of the particular feature delivers interesting model interpretations. For example, in the case of $\Delta D_{99\%}$ for PTV, it can be seen that what drives the predictions towards the class 0 are high values of $\Delta D_{99\%}$ that, if we look at Figure 6.1, correspond to a high probability of this treatment planning objective being met. In this scenario, the fact that class 0 is associated with plans for which the PTV objective was also met, shows that the model interpretation is consistent with the clinical

interpretation. Another example includes the geometric features i.e. Figure 6.4a illustrating that larger distances $\Delta$d between PTV and OARs tend to drive model predictions towards the acceptable plans (class 0) as well. This is also a very common observation in clinical practice because the larger $\Delta$d, the easier it is to create a treatment plan that would provide desirable PTV coverage and simultaneously spare OARs. Furthermore, the dosimetric features for OARs show that higher dose deviations drive the prediction to class 1 and correspond to plans for which the PTV objective was not met due to the priority trade-off in order to meet one or more OAR constraints. This also demonstrates that the analysis of model explainability provides valuable insights and confirms that the model relies on clinically valid logic when making predictions.

Figure 6.5 also confirms this and additionally shows the interaction between some of the individual features. For instance, Figure 6.5a shows that the smaller the distance between PTV and right cochlea, the higher the probability of a high dose being delivered to that organ. The same relationship can be observed between PTV-to-brainstem distance and $\Delta D_{1\%}$ for optic chiasm. Other inter-feature interactions are less pronounced, although the majority of geometric features have relatively strong interaction with $\Delta D_{1\%}$ for optic chiasm. By examining the partial dependency charts for dosimetric features in Figure 6.5b, it can be seen that the strongest mutual interaction of $\Delta D_{1\%}$ for optic chiasm exists with PTV-brainstem distance. The most important model feature, $\Delta D_{99\%}$ for PTV, had the strongest interaction with distance to the optic chiasm, demonstrating that if PTV is located in the proximity of optic chiasm there is a low probability of meeting the PTV dose objectives. This is consistent with the observation in the clinical practice in a

situation where the PTV coverage trade-off needs to be made in order to meet treatment objectives for critical structures.

It is also interesting to note that the best performing model is logistic regression, which is less complex than SVM. We believe that it is most likely due to the simplicity of our classification task (binary classification) and because of the presence of the feature(s) with strong predictive power as shown in the global explainability chart (Figure 6.4), that indicates the $\Delta$D99% metric for PTV as the feature that contributes to the model outputs the most significantly. This observation confirms an important characteristic of ML, namely that an increase in model complexity does not always lead to an increase in model performance.

The proposed ML classifier and model explainability work together to provide additional value to the clinical processes. After the plan is created and the algorithm classifies the treatment plan, the model explainability analysis (performed instantly) indicates the attributes behind the plan classification (i.e. which plan and patient-related attributes caused the plan to require trade-off). Therefore the clinician would not have to analyse the treatment plan and/or schedule a consultation with the treatment planner/dosimetrist or other radiation oncologists to find the cause. This opens up an opportunity of developing an automated notification system for clinicians (not explored in our study but possible for future work).

Furthermore for challenging plans (those with trade-offs), it is found to be useful both in our clinic as well as in the recent literature, to look at similar past plans, and make data-driven decisions regarding further steps in treatment planning (e.g. plan modifications). A new plan, once classified as a plan with trade-offs in the proposed ML system, is

automatically compared to similar plans delivered in the past. The clinician does not have to search through the clinical database to find similar plans.

Potential future applications of such an ML system could expand beyond a single institution. One can envision smaller radiation therapy centers sending challenging plans (ie. those with trade-offs) to a large, experienced center for ML analysis, and thus providing an invaluable planning QA tool. The large center could incorporate the new plans in the ML training set and refine their model (ie. continual learning). This data and technology-sharing environment could potentially equalize the standard of care regardless of the resource availability of a given medical facility. This example can intuitively be expanded to broad collaboration between many large and small institutions across the world.

Another potential clinical impact of the presented system is for management of adaptive radiation therapy. Specifically, patients with dosimetric compromises (trade-offs) might justify more accurate dose delivery and positioning. Therefore these patients could be identified as high priority for Adaptive Radiation Therapy (ART).

This study achieved promising model performance and provided clinically feasible interpretations, however, has two limitations. The first is a limited number of plan datasets, which may impact the model's robustness. This is mainly due to the limitations of data availability for local brain tumor plans qualified for this study (ie. total dose delivered, fractionation, delivery technique). Data availability is a common problem in radiation therapy studies involving the application of ML. There are many papers where ML models are trained using <150 and as little as 11 patients[9,31,48,49]. A second limitation of this study is the simplified, binary classification of plans. At the design stage of the

study, we found that plans, where the priority trade-offs were made for OARs, could have been additionally divided into plans with higher and lower priority trade-offs. However, our relatively small data set would cause those two potential classes to be significantly under-sampled, therefore ultimately the classification of trade-off priorities was not pursued in this study.

Future work might include building a model based on a larger amount of data. Additionally, the creation of relevant treatment plan classifications can depend on and vary among radiation oncology professionals and institutions, due to variability in experience, expertise, and clinical protocols. Therefore, to minimize intrinsic biases, future work could involve collaboration among many medical physicists and clinicians across several institutions, in designing and applying specific plan classification to maximize the potential of ML models and their impact on day-to-day clinical practice.

Additionally, a possible subsequent study could also involve testing the model using unapproved plans to further evaluate a model performance.

## 6.5 Conclusion

The purpose of this study was to train an ML model to classify previously delivered VMAT plans for brain tumor patients into two categories created in collaboration with expert radiation oncologists. The trained model achieved satisfactory accuracy on the test data and can be used by medical physicists in the data-driven quality assurance program as well as by radiation oncologists to support their decision-making process in terms of treatment plan approval and potential plan modifications. Model explainability analysis

facilitated a better understanding of the machine learning model reasoning for the generated predictions and showed consistency with clinical observations.

## 6.6 References


1. Chao KSC. *Practical Essentials of Intensity Modulated Radiation Therapy.* Wolters Kluwer Health; 2013.
2. Gaya A, Mahadevan A. *Stereotactic Body Radiotherapy: A Practical Guide.* Springer London; 2015.
3. Xia P, Godley A, Shah C, Gregory M. M. Videtic MDCMF, Suh J. *Strategies for Radiation Therapy Treatment Planning.* Springer Publishing Company; 2018.
4. Li N, Carmona R, Sirak I, et al. Highly efficient training, refinement, and validation of a knowledge-based planning quality-control system for radiation therapy clinical trials. *International Journal of Radiation Oncology* Biology* Physics.* 2017;97(1):164-172.
5. Wang J, Hu W, Yang Z, et al. Is it possible for knowledge-based planning to improve intensity modulated radiation therapy plan quality for planners with different planning experiences in left-sided breast cancer patients? *Radiation Oncology.* 2017;12(1):85.
6. Scaggion A, Fusella M, Roggio A, et al. Reducing inter-and intra-planner variability in radiotherapy plan output with a commercial knowledge-based planning solution. *Physica Medica.* 2018;53:86-93.
7. Kamima T, Ueda Y, Fukunaga J-i, et al. Multi-institutional evaluation of knowledge-based planning performance of volumetric modulated arc therapy (VMAT) for head and neck cancer. *Physica Medica.* 2019;64:174-181.

8. Yu S, Xu H, Sinclair A, Zhang X, Langner U, Mak K. Dosimetric and planning efficiency comparison for lung SBRT: CyberKnife vs VMAT vs knowledge-based VMAT. *Medical Dosimetry.* 2020.

9. Wall PD, Fontenot JD. Application and comparison of machine learning models for predicting quality assurance outcomes in radiation therapy treatment planning. *Informatics in Medicine Unlocked.* 2020;18:100292.

10. Hirashima H, Ono T, Nakamura M, et al. Improvement of prediction and classification performance for gamma passing rate by using plan complexity and dosiomics features. *Radiotherapy and Oncology.* 2020.

11. Osman AF, Maalej NM, Jayesh K. Prediction of the individual multileaf collimator positional deviations during dynamic IMRT delivery priori with artificial neural network. *Medical Physics.* 2020;47(4):1421-1430.

12. Li Q, Chan MF. Predictive time-series modeling using artificial neural networks for Linac beam symmetry: an empirical study. *Annals of the New York Academy of Sciences.* 2017;1387(1):84.

13. Guidi G, Maffei N, Meduri B, et al. A machine learning tool for re-planning and adaptive RT: a multicenter cohort investigation. *Physica Medica.* 2016;32(12):1659-1666.

14. Carlson JN, Park JM, Park S-Y, Park JI, Choi Y, Ye S-J. A machine learning approach to the accurate prediction of multi-leaf collimator positional errors. *Physics in Medicine & Biology.* 2016;61(6):2514.

15. Grewal HS, Chacko MS, Ahmad S, Jin H. Prediction of the output factor using machine and deep learning approach in uniform scanning proton therapy. *Journal of Applied Clinical Medical Physics.* 2020.

16. Valdes G, Scheuermann R, Hung C, Olszanski A, Bellerive M, Solberg T. A mathematical framework for virtual IMRT QA using machine learning. *Medical physics.* 2016;43(7):4323-4334.

17. Wang C, Zhu X, Hong JC, Zheng D. Artificial intelligence in radiotherapy treatment planning: present and future. *Technology in cancer research & treatment.* 2019;18:1533033819873922.

18. Tambe NS, Pires IM, Moore C, Cawthorne C, Beavis AW. Validation of in-house knowledge-based planning model for advance-stage lung cancer patients treated using VMAT radiotherapy. *The British Journal of Radiology.* 2020;93(1106):20190535.

19. Fogliata A, Cozzi L, Reggiori G, et al. RapidPlan knowledge based planning: iterative learning process and model ability to steer planning strategies. *Radiation Oncology.* 2019;14(1):187.

20. Zhang J, Ge Y, Sheng Y, et al. Knowledge-Based Tradeoff Hyperplanes for Head and Neck Treatment Planning. *International Journal of Radiation Oncology* Biology* Physics.* 2020.

21. Chatterjee A, Serban M, Faria S, Souhami L, Cury F, Seuntjens J. Novel knowledge-based treatment planning model for hypofractionated radiotherapy of prostate cancer patients. *Physica Medica.* 2020;69:36-43.

22. van Schie MA, Janssen TM, Eekhout D, et al. Knowledge-based assessment of focal dose escalation treatment plans in prostate cancer. *International Journal of Radiation Oncology* Biology* Physics.* 2020.

23. Delaney AR, Dahele M, Tol JP, Slotman BJ, Verbakel WF. Knowledge-based planning for stereotactic radiotherapy of peripheral early-stage lung cancer. *Acta Oncologica.* 2017;56(3):490-495.

24. Teichert K, Currie G, Küfer K-H, et al. Targeted multi-criteria optimisation in IMRT planning supplemented by knowledge based model creation. *Operations Research for Health Care.* 2019;23:100185.

25. van't Hof S, Delaney AR, Tekatli H, et al. Knowledge-based planning for identifying high-risk stereotactic ablative radiation therapy treatment plans for lung tumors larger than 5 cm. *International Journal of Radiation Oncology* Biology* Physics.* 2019;103(1):259-267.

26. Shepherd M, Bromley R, Stevens M, et al. Developing knowledge-based planning for gynaecological and rectal cancers: a clinical validation of RapidPlan™. *Journal of Medical Radiation Sciences.* 2020.

27. Wang M, Li S, Huang Y, et al. An interactive plan and model evolution method for knowledge-based pelvic VMAT planning. *Journal of Applied Clinical Medical Physics.* 2018;19(5):491-498.

28. Fan J, Wang J, Zhang Z, Hu W. Iterative dataset optimization in automated planning: Implementation for breast and rectal cancer radiotherapy. *Medical physics.* 2017;44(6):2515-2531.

29. Rice A, Zoller I, Kocos K, et al. The implementation of RapidPlan in predicting deep inspiration breath-hold candidates with left-sided breast cancer. *Medical Dosimetry.* 2019;44(3):210-218.

30. Kubo K, Monzen H, Ishii K, et al. Inter-planner variation in treatment-plan quality of plans created with a knowledge-based treatment planning system. *Physica Medica.* 2019;67:132-140.

31. Kishi N, Nakamura M, Hirashima H, et al. Validation of the clinical applicability of knowledge-based planning models in single-isocenter volumetric-modulated arc therapy for multiple brain metastases. *Journal of Applied Clinical Medical Physics.* 2020.

32. Chan MF, Witztum A, Valdes G. Integration of AI and Machine Learning in Radiotherapy QA. *Frontiers in Artificial Intelligence.* 2020;3:76.

33. Deist TM, Dankers FJ, Valdes G, et al. Machine learning algorithms for outcome prediction in (chemo) radiotherapy: An empirical comparison of classifiers. *Medical physics.* 2018;45(7):3449-3459.

34. Luo Y, McShan D, Ray D, et al. Development of a fully cross-validated bayesian network approach for local control prediction in lung cancer. *IEEE transactions on radiation and plasma medical sciences.* 2018;3(2):232-241.

35. Yu T, Lam SK, To LH, et al. Pretreatment prediction of adaptive radiation therapy eligibility using MRI-based radiomics for advanced nasopharyngeal carcinoma patients. *Frontiers in oncology.* 2019;9:1050.

36. Hausdorff F. *Set Theory.* Chelsea Publ.; 1991.

37. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research.* 2011;12:2825-2830.

38. Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of cheminformatics.* 2014;6(1):1-15.

39. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics.* 2006;7(1):91.

40. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction.* OUP Oxford; 2003.

41. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence.* 2020;2(1):2522-5839.

42. Boutilier JJ, Lee T, Craig T, Sharpe MB, Chan TC. Models for predicting objective function weights in prostate cancer IMRT. *Medical physics.* 2015;42(4):1586-1595.

43. Cooper BT, Li X, Shin SM, et al. Preplanning prediction of the left anterior descending artery maximum dose based on patient, dosimetric, and treatment planning parameters. *Advances in radiation oncology.* 2016;1(4):373-381.

44. Gabryś HS, Buettner F, Sterzing F, Hauswald H, Bangert M. Design and selection of machine learning methods using radiomics and dosiomics for normal tissue complication probability modeling of xerostomia. *Frontiers in oncology.* 2018;8:35.

45. Arimura H, Soufi M, Kamezawa H, Ninomiya K, Yamada M. Radiomics with artificial intelligence for precision medicine in radiation therapy. *Journal of radiation research.* 2019;60(1):150-157.

46. Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge and Data Engineering.* 2005;17(3):299-310.

47. Hackeling G. *Mastering Machine Learning with scikit-learn.* Packt Publishing Ltd; 2017.

48. Lee T, Hammad M, Chan TC, Craig T, Sharpe MB. Predicting objective function weights from patient anatomy in prostate IMRT treatment planning. *Medical physics.* 2013;40(12):121706.

49. Schubert C, Waletzko O, Weiss C, et al. Intercenter validation of a knowledge based model for automated planning of volumetric modulated arc therapy for prostate cancer. The experience of the German RapidPlan Consortium. *PLoS One.* 2017;12(5):e0178034.

# Chapter 7

# Thesis Summary & Future Work

## 7.1 Summary

One of the main challenges in radiation therapy is to ensure accurate delivery of the prescribed radiation dose when the patient's anatomy changes during the course of treatment. These changes may be compensated for through the process of Adaptive Radiation Therapy (ART), where the treatment is adapted during the course of therapy. ART continues to be the subject of significant research and development in the radiation oncology community over recent years.

Conventional radiation therapy planning is a complex and time-consuming process requiring synchronized cooperation of experts from multiple disciplines. Therefore, ART, in which a new radiotherapy plan (as a modification of the previous plan) needs to be created during ongoing patient treatment, requires application of new specific tools and procedures to make the ART process more time-efficient, accurate, and clinically effective. These tools and procedures include: i) CT to CBCT deformable image registration for daily dose accumulation, estimation of total delivered dose and contour propagation; ii) selection of the most effective ART approach to maximize the dosimetric benefit and

minimize the use of valuable time and human resources; iii) automatic segmentation of organs-at-risk for improved speed and consistency of the contouring process; iv) selection of those patients who would likely require ART for proper allocation of ART procedures and Quality Assurance of the treatment plan. The goal of this thesis was to propose and evaluate solutions to these aspects of adaptive radiation therapy to make it more reliable, accurate, and efficient.

Chapter 2 provided an introduction of fundamental topics to facilitate a better understanding of the research methodology used to achieve specific thesis goals. The introductory topics included the radiobiological basis of radiation therapy and modern irradiation methods (IMRT & VMAT), deformable image registration, online and offline adaptive radiation therapy as well as machine learning and deep learning methods.

The objective of the research work described in Chapter 3 was to evaluate several deformable image registration algorithms by registering computed tomography scans to daily CBCT images for five prostate cancer patients treated with intensity-modulated radiation therapy. Each patient received a total dose of 74 Gy over 37 fractions, 2Gy/fraction. The CBCT scans were obtained approximately in the middle of radiotherapy treatment ($18^{\text{th}}$ fraction) with the Varian On-Board Imager (OBI) system attached to the linac. The patient's anatomy was manually segmented on both CT and CBCT scans by an experienced radiation oncologist. Contoured structures that were used for the evaluation of DIR algorithms included PTV, CTV, bladder, and rectum. CT images were registered to CBCT images with Rigid, Affine, B-spline, and Dense Anatomical Block Matching (DABM) algorithms. The performance of these registration methods was assessed quantitatively using Dice Score, Hausdorff Distances, Center-of-

Mass Shift metrics comparing the deformed and manually delineated contours. A physician evaluation also included the measurement of the time required to correct the deformed contours – the less time it took a physician to make necessary corrections the better the performance of DIR method. The topology of the vector field was analysed with the Jacobian determinant metric, while the accuracy of the inverted DVFs was examined with the Inverse Consistency Error metric. The comparison of DIR algorithms showed that the DABM registration outperformed the other methods in terms of both quantitative and physician evaluations. Additionally, this study demonstrated that the results for Dense Anatomical Block Matching algorithm are generally at a clinically acceptable level. This makes DABM a promising alternative to the existing registration methods when it comes to the challenging problem of CT-CBCT deformable image registration and its application in adaptive radiation therapy of the pelvic region.

The aim of the study presented in Chapter 4 was to evaluate eight adaptive radiation therapy approaches specifically for prostate cancer patients treated with hypofractionation. The ART strategies included online and offline methods as well as the application of dose feedback. The retrospective study included 20 prostate cancer patients treated with 40 Gy total dose over five fractions (8Gy/fraction) using volumetric modulated arc therapy (VMAT). Daily CBCT images were acquired before the delivery of every fraction and then, with the application of deformable image registration used for the estimation of the daily dose, contouring and plan re-optimization occurred. Dosimetric benefits of the various ART strategies were quantified by the comparison of dose and dose-volume metrics for the original treatment plan and the adapted plans with the consideration of target volumes (PTV and CTV) as well as several critical structures

(bladder, rectum, left and right femoral heads). Non-adapted plans unfavourably deviated from the intended delivery by 3% to 12% for PTV and CTV as well as by 1% to 7% for OARs in terms of $D_{min}$, $D_{max}$, $D_{mean}$, and other dose-volume metrics. The best performing online ART strategies resulted in a significant reduction of those deviations to a maximum of 5% for PTV and CTV as well as to a maximum of 2% for OARs. Additionally, most ART strategies improved the CTV coverage as measured with the homogeneity index by up to 30% relative to the intended plan. Using the same metric, PTV coverage improved by around 20% with respect to the non-adapted plan. The dosimetric impact of the application of several adaptive radiation therapy strategies for hypofractionated prostate cancer patients was quantified. Daily on-line adaptation approaches (ie. every fraction) were the most impactful, although strategies adapting every other fraction also were impactful while reducing relative workload as well. Offline treatment adaptations were shown to be less beneficial due to increased dose delivered to bladder and rectum, as compared to other ART strategies.

Chapter 5 described the study of a deep learning network to automatically contour critical organs on the Computed Tomography (CT) scans of head and neck cancer patients who underwent radiation therapy treatment. This study included 25 critical organs that were delineated by expert radiation oncologists. Contoured medical images of 964 patients were sourced from a publicly available TCIA database. The proportion of training, validation, and testing samples for deep learning model development was 65%, 25%, and 10% respectively. The CT scans and segmentation masks were augmented with shift, scale, and rotate transformations. Additionally, medical images were pre-processed using contrast limited adaptive histogram equalization (CLAHE) to enhance soft tissue contrast

while contours were subjected to morphological operations to ensure their structural integrity. The segmentation model was based on the U-Net architecture with embedded Inception-ResNet-v2 blocks and was trained over 100 epochs with a batch size of 32 and an adaptive learning rate optimizer. The loss function combined the Jaccard Index and Binary Cross Entropy. The model performance was evaluated with Dice Score, Jaccard Index, and Hausdorff Distances. The interpretability of the model was analyzed with Guided Gradient-weighted Class Activation Mapping. The Dice Score, Jaccard Index, and mean Hausdorff Distance averaged over all structures and patients were 0.79±0.10, 0.68±0.12, and 1.84±1.51mm respectively on the testing data sets. The average model training time was 8h per anatomical structure. Once the model was trained, the full segmentation of head and neck anatomy required only 6.8s per patient. Satisfactory accuracy (as compared to clinical studies in the literature), short contouring time, and realistic prediction reasoning make the model proposed in this work a feasible solution for head and neck CT scan segmentation in a clinical environment.

The last research chapter (Chapter 6) presents an investigation into the application of machine learning methods to build a decision support system that classified previously delivered VMAT plans of brain tumor patients into two categories created in collaboration with expert radiation oncologists. The first category contained plans that met the PTV treatment planning objectives. The second category included plans for which PTV objectives were not met due to the priority given to one or more OARs (i.e. a trade-off was required); those plans, however, were still clinically accepted and delivered. Once trained, the ML algorithm would be able to predict which new plans would require a compromise. The study included 79 plans for brain tumor patients that were prescribed a

total dose of 60 Gy delivered in 30 fractions (2 Gy per fraction) with volumetric modulated arc therapy. Structures considered for analysis included the planning target volume (PTV), brainstem, cochleae, and optic chiasm. The ML model was trained based on the ten features that were derived from historical radiotherapy plans. Several models were evaluated (including Support Vector Machine, Elastic Net, Logistic Regression, and Random Forest) using double-nested cross-validation and an area-under-the-curve (AUC) metric. The models with hyperparameters tuned after nested cross-validation were evaluated on the test data that accounted for 20% (15 plans) of all the data (79 plans) using single 5-fold cross-validation. Shapely additive explanation (SHAP) interaction values were utilized to provide explanations for model predictions. The Logistic Regression model received the highest score and showed the lowest standard deviation for the inner loop of the cross-validation metric. The accuracy and the AUC measure on the testing data for the Logistic Regression model were 93.8±4.1% and 0.98±0.02 respectively. The SHAP analysis indicated that the $\Delta D_{99\%}$ metric for PTV had the greatest influence on the model predictions and is approximately 3.5 times larger (i.e. more important) than the second most important feature – $\Delta D_{1\%}$ for optic chiasm. The least important features were $\Delta D_{MAX}$ for the left and right cochlea. The trained model achieved satisfactory accuracy and could be used by medical physicists in a data-driven quality assurance program as well as by radiation oncologists to support their decision-making process in terms of treatment plan approval and potential plan modifications. Model explanation analysis showed that the model relies on clinically valid logic when making predictions.

It is also worth noting that there are several challenges associated with techniques described in four research chapters of this thesis (Chapter 3-6) in terms of their application to adaptive

radiation therapy. Namely, CT-CBT deformable image registration still poses a difficulty due to the poor quality of CBCT images, that negatively affects the accuracy of the registration. Additionally, often lack of the "gold standard" for the deformable vector field makes evaluation of DIR methods difficult and time consuming. When it comes to ART strategies, the implementation of recent technological advancements in a clinic is required to utilize ART in timely manner, which poses a barrier for smaller medical centers. Additionally, the broad application of deep learning segmentation models in adaptive radiation therapy will remain challenging as long as contouring protocols will not be unified across all the medical facilities and high quality training data will not be shared across all institutions. Last but not least the application of machine learning for treatment plan management (for example quality assurance, approval, modifications, general plan analysis and interpretation) require close collaboration with radiation oncologists as well as collection of large amounts of high quality data and a significant time commitment, which remains to be challenging in the busy clinical environment.

## 7.2 Future Work

This section outlines the future work that could add value to the research projects completed in the four studies that were described in Chapters 3-6, as well as larger scale opportunities in these areas.

The Dense Anatomical Block Matching investigated in Chapter 3 provided very promising results for the challenging CT to CBCT image registration problem. However, it requires significant computational time (~7 min.) to complete the image alignment process. Therefore, future work could include the parallelization of the software code in order to

improve the time efficiency of the DABM algorithm. Although the image registration algorithms that the DABM was tested against are commonly-implemented in medical applications, additional algorithms could be examined including commercially available deformable image registration solutions. The performance of the DABM algorithm could be improved by automatically tuning the algorithmic parameters for individual registrations, eliminating the need for an empirical search of optimal parameters. Furthermore, the performance bias resulting from interobserver variability in contouring a patient's anatomy could be reduced by using data from more than one physician.

In regard to the evaluation of various ART strategies in Chapter 4, this study was limited to a quantitative dosimetric assessment only. Therefore it could potentially be valuable to perform a clinical evaluation of the adapted plans in which an experienced radiation oncologist scores each ART approach.

The work in Chapter 5 focussed on the application of a deep learning method for automatic segmentation task. In future work the segmentation could involve the use of many dependent and independent DL networks, that could be assembled to deliver predictions of the segmentation masks. Also, the imaging training data could be first clustered into groups with reduced variances, which could result in the creation of several, more robust DL contouring models. Another possibility for future work may include an introduction of artificial contour perturbations (many contours delineating the same structure on the same imaging data) to each segmentation, to model interobserver variability. In this case, the DL algorithm will not learn to segment a given part of a patient's anatomy according to the one contour created by one physician, but rather will learn how to ensure that automatically generated segmentation will be at an acceptable

level of clinically encountered interobserver variabilities. Also, future work could consider the utilization of only those data that were acquired and thoroughly examined by all involved physicians at a single clinic. The application of local, high-quality, and curated data would most likely result in developing a more reliable model that would have a significant and immediate impact on the time required to create the radiotherapy treatment plan for a specific group of patients. Beyond this, if national or international consensus on contouring methods could be fully and consistently adopted (for example, contouring guidelines published by ASTRO[1] and ESTRO[2]), huge amounts of data could be pooled for even more robust model development.

The clinical decision-making support system presented in Chapter 6 can also be improved. The creation of the relevant treatment plan classifications can depend on and vary among radiation oncology professionals and institutions, due to variability in experience, expertise, and clinical protocols. Therefore, to minimize intrinsic biases, future work could involve collaboration across several institutions in designing and applying specific plan classifications to maximize the potential of ML models and their impact on day-to-day clinical practice. Moreover, although the accuracy of the model trained in this research was close to 100%, the long-term model robustness could potentially benefit from data augmentation.

In addition to the preceding focussed suggestions, there are certain broader opportunities for future work that are common across all four studies. The main improvement that could be made in future studies is an increase in available relevant data. This includes the number of patients, the expansion of clinical cases covered, and the incorporation of various radiotherapy methods used to deliver radiation (i.e. irradiation techniques and

fractionation schemes). This could be achieved by establishing a national and/or an international sharable database built using a decentralized (i.e. where all network nodes contain a copy of all the data) system on the blockchain for increased security of the data. Immediate access to such a large amount of patient data, including large size files of medical imaging data, could be facilitated with 5G technology that allows for very fast wireless transfer of data over the internet network (up to 20Gb/s of peak data rate compared to current LTE-Advanced (4G) with up to 1Gb/s).

Artificial intelligence could be investigated for the applications studied in Chapters 3 and 4. Aspects of that work, including deformable image registration, CBCT image enhancement quality, and delivered dose estimation, lend themselves to the application of AI methods and published studies are already showing community interest in these applications. Indeed, the abundance of machine learning and deep learning algorithms and the seemingly unlimited flexibility in terms of network architecture modifications (in the case of DL) that can be performed, provide many research opportunities targeted at solving the complex problems encountered in radiation therapy.

## 7.3 References

1. The American Society of Radiotherapy and Oncology (ASTRO). https://www.astro.org/Patient-Care-and-Research/Clinical-Practice-Statements/Contouring-Consensus-Guidance-Document. 2019.
2. The European Society of Radiotherapy and Oncology (ESTRO). https://www.estro.org/Science/Guidelines. 2020.

# Chapter 8

# Appendix

## 8.1 Dose-Volume Metrics & Statistical Significance for ART Strategies

This section contains supplementary material for the research work described in Chapter 4 *Adaptive radiation therapy strategies in the treatment of prostate cancer patients using hypofractionated VMAT*

**Table 8.1.** Maximum, Mean, and Minimum radiation doses with corresponding standard deviations for the reference and ART strategies.

| | | | | Adaptive Radiation Therapy Strategies | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Structure** | **Metric** | **Planned** | **non-ART** | **Cont.+DF** | **DF 2-4** | **Online+DF** | **Online** | **Online 1-3-5** | **Offline+DF** | **Offline +Online** | **Offline** |
| **CTV** | Max. Dose [%] | 105.28 ± 0.71 | 99.81 ± 0.25 | 105.69 ± 2.31 | 106.10 ± 2.07 | 105.98 ± 2.53 | 105.84 ± 2.35 | 105.81 ± 1.65 | 106.89 ± 2.40 | 107.83 ± 4.42 | 108.76 ± 3.64 |
| | Mean Dose [%] | 102.12 ± 0.50 | 97.48 ± 0.31 | 103.32 ± 1.96 | 103.80 ± 1.88 | 103.49 ± 1.92 | 103.84 ± 2.37 | 103.93 ± 1.65 | 104.43 ± 2.10 | 105.65 ± 4.36 | 106.71 ± 3.39 |
| | Min. Dose [%] | 99.32 ± 0.59 | 94.47 ± 1.56 | 100.89 ± 2.30 | 101.49 ± 1.51 | 101.49 ± 1.94 | 101.68 ± 1.96 | 101.87 ± 1.34 | 100.96 ± 2.39 | 101.71 ± 2.09 | 101.10 ± 2.20 |
| **PTV** | Max. Dose [%] | 105.91 ± 0.64 | 100.00 ± 0.00 | 106.67 ± 2.32 | 106.77 ± 20 | 106.82 ± 2.47 | 106.07 ± 2.38 | 105.95 ± 1.60 | 107.59 ± 2.29 | 108.10 ± 4.38 | 108.89 ± 3.60 |
| | Mean Dose [%] | 101.37 ± 0.38 | 96.17 ± 0.58 | 102.26 ± 1.44 | 102.51 ± 1.28 | 102.39 ± 1.43 | 102.54 ± 1.76 | 102.60 ± 1.19 | 103.07 ± 1.60 | 103.98 ± 3.41 | 104.72 ± 2.63 |
| | Min. Dose [%] | 90.15 ± 0.70 | 74.77 ± 8.81 | 86.17 ± 6.97 | 85.99 ± 4.71 | 85.01 ± 9.18 | 84.73 ± 10.14 | 84.88 ± 6.03 | 84.51 ± 7.19 | 81.86 ± 10.06 | 82.92 ± 6.65 |
| **Bladder** | Max. Dose [%] | 105.44 ± 0.78 | 99.17 ± 0.61 | 106.25 ± 2.39 | 105.94 ± 1.85 | 106.08 ± 2.41 | 105.36 ± 2.38 | 105.27 ± 1.53 | 106.78 ± 2.54 | 106.93 ± 4.13 | 107.67 ± 3.20 |
| | Mean Dose [%] | 15.67 ± 6.52 | 14.24 ± 5.63 | 16.59 ± 7.86 | 15.68 ± 6.90 | 15.82 ± 7.38 | 15.63 ± 7.09 | 15.49 ± 6.47 | 16.06 ± 7.08 | 15.81 ± 6.87 | 15.74 ± 6.59 |
| | Min. Dose [%] | 1.32 ± 1.06 | 0.65 ± 1.23 | 0.75 ± 1.63 | 0.74 ± 1.48 | 0.74 ± 1.54 | 0.74 ± 1.48 | 0.72 ± 1.42 | 0.75 ± 1.51 | 0.73 ± 1.44 | 0.73 ± 1.43 |
| **Rectum** | Max. Dose [%] | 103.79 ± 1.03 | 98.36 ± 1.02 | 104.86 ± 2.29 | 104.28 ± 1.22 | 103.97 ± 1.84 | 104.29 ± 2.17 | 104.29 ± 1.22 | 104.75 ± 1.98 | 105.78 ± 4.31 | 106.51 ± 3.58 |
| | Mean Dose [%] | 38.60 ± 7.67 | 38.89 ± 7.77 | 38.86 ± 7.63 | 39.69 ± 7.21 | 39.10 ± 7.90 | 39.54 ± 7.77 | 40.26 ± 7.71 | 39.49 ± 7.72 | 40.20 ± 8.08 | 41.06 ± 8.18 |
| | Min. Dose [%] | 2.51 ± 0.98 | 2.20± 1.14 | 2.37 ± 1.30 | 2.36 ± 1.21 | 2.37 ± 1.20 | 2.35 ± 1.19 | 2.35 ± 1.20 | 2.40 ± 1.22 | 2.38 ± 1.23 | 2.42 ± 1.24 |
| **Femur-LT** | Max. Dose [%] | 33.45 ± 5.24 | 31.23 ± 4.97 | 35.57 ± 5.65 | 34.26 ± 5.39 | 34.35 ± 5.28 | 34.11 ± 5.06 | 32.44 ± 4.58 | 34.80 ± 5.50 | 34.43 ± 5.16 | 34.41 ± 5.37 |
| | Mean Dose [%] | 13.74 ± 2.46 | 13.16 ± 2.23 | 15.16 ± 2.76 | 14.55 ± 2.54 | 13.96 ± 2.52 | 13.62 ± 2.43 | 13.76 ± 2.25 | 14.40 ± 2.57 | 13.77 ± 2.39 | 14.03 ± 2.41 |
| | Min. Dose [%] | 1.10 ± 0.34 | 1.01 ± 0.43 | 1.17 ± 0.53 | 1.13 ± 0.50 | 1.11 ± 0.53 | 1.11 ± 0.52 | 1.09 ± 0.49 | 1.15 ± 0.53 | 1.13 ± 0.53 | 1.13 ± 0.52 |
| **Femur-RT** | Max. Dose [%] | 35.36 ± 6.75 | 33.48 ± 6.26 | 35.86 ± 5.54 | 35.17 ± 5.45 | 33.88 ± 4.91 | 32.66 ± 4.64 | 33.06 ± 4.92 | 34.26 ± 5.56 | 33.74 ± 4.59 | 33.80 ± 5.36 |
| | Mean Dose [%] | 13.65 ± 2.23 | 12.96 ± 2.07 | 15.51 ± 2.48 | 14.54 ± 2.28 | 14.67 ± 2.53 | 14.09 ± 2.47 | 14.04 ± 2.30 | 14.67 ± 2.54 | 14.36 ± 2.51 | 14.43 ± 2.64 |
| | Min. Dose [%] | 1.11 ± 0.36 | 1.01 ± 0.46 | 1.16 ± 0.60 | 1.11 ± 0.52 | 1.09 ± 0.53 | 1.06 ± 0.52 | 1.08 ± 0.51 | 1.12 ± 0.54 | 1.07 ± 0.52 | 1.10 ± 0.53 |

**Table 8. 2.** Dose and dose-volume metrics with corresponding standard deviations for the reference and ART strategies.

| | | | Adaptive Radiation Therapy Strategies | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Structure** | **Metric** | **Planned** | **non-ART** | **Cont.+DF** | **DF 2-4** | **Online+DF** | **Online** | **Online 1-3-5** | **Offline+DF** | **Offline+Online** | **Offline** |
| **CTV** | D1% [cGy] | 4149 ± 23 | 3953 ± 10 | 4184 ± 83 | 4201 ± 83 | 4188 ± 83 | 4195 ± 96 | 4198 ± 67 | 4230 ± 90 | 4274 ± 180 | 4314 ± 144 |
| | D2% [cGy] | 4142 ± 23 | 3946 ± 10 | 4178 ± 83 | 4195 ± 83 | 4181 ± 81 | 4189 ± 97 | 4192 ± 67 | 4222 ± 91 | 4268 ± 180 | 4309 ± 143 |
| | D5% [cGy] | 4131 ± 22 | 3937 ± 11 | 4169 ± 82 | 4185 ± 83 | 4171 ± 80 | 4181 ± 96 | 4184 ± 67 | 4212 ± 91 | 4259 ± 179 | 4301 ± 143 |
| | D50% [cGy] | 4084 ± 20 | 3898 ± 13 | 4132 ± 78 | 4152 ± 78 | 4139 ± 77 | 4153 ± 94 | 4156 ± 66 | 4180 ± 90 | 4227 ± 177 | 4273 ± 139 |
| | D95% [cGy] | 4042 ± 21 | 3864 ± 16 | 4099 ± 77 | 4119 ± 62 | 4112 ± 75 | 4125 ± 94 | 4131 ± 65 | 4132 ± 77 | 4185 ± 167 | 4220 ± 133 |
| | D98% [cGy] | 4033 ± 21 | 3854 ± 18 | 4090 ± 79 | 4110 ± 60 | 4105 ± 76 | 4116 ± 95 | 4123 ± 65 | 4117 ± 80 | 4164 ± 147 | 4192 ± 117 |
| | D99% [cGy] | 4027 ± 21 | 3846 ± 23 | 4083 ± 81 | 4105 ± 60 | 4099 ± 76 | 4110 ± 95 | 4118 ± 65 | 4106 ± 83 | 4146 ± 124 | 4166 ± 94 |
| | V100% [% of volume] | 99.87 ± 0.17 | 0.00 ± 0.00 | 99.87 ± 0.36 | 99.99 ± 0.03 | 99.99 ± 0.03 | 99.97 ± 0.10 | 100 ± 0.00 | 99.78 ± 0.75 | 99.93 ± 0.33 | 99.88 ± 0.36 |
| | V105% [% of volume] | 0.15 ± 0.49 | 0.00 ± 0.00 | 11.37 ± 30.55 | 18.60 ± 36.63 | 15.19 ± 33.66 | 15.52 ± 35.04 | 14.88 ± 32.66 | 28.99 ± 44.35 | 40.15 ± 46 | 61.29 ± 42.73 |
| | HI [%] | 0.03 ± 0.00 | 0.02 ± 0.00 | 0.02 ± 0.01 | 0.02 ± 0.01 | 0.02 ± 0.00 | 0.02 ± 0.00 | 0.02 ± 0.00 | 0.02 ± 0.02 | 0.02 ± 0.01 | 0.03 ± 0.02 |
| **PTV** | D1% [cGy] | 4158 ± 21 | 3954 ± 8 | 4198 ± 81 | 4207 ± 80 | 4198 ± 80 | 4197 ± 95 | 4200 ± 66 | 4238 ± 88 | 4275 ± 179 | 4313 ± 143 |
| | D2% [cGy] | 4149 ± 20 | 3947 ± 9 | 4188 ± 79 | 4199 ± 80 | 4188 ± 79 | 4191 ± 96 | 4193 ± 66 | 4229 ± 88 | 4268 ± 179 | 4307 ± 143 |
| | D5% [cGy] | 4135 ± 19 | 3936 ± 10 | 4175 ± 79 | 4187 ± 80 | 4175 ± 78 | 4181 ± 95 | 4184 ± 66 | 4215 ± 89 | 4258 ± 178 | 4298 ± 142 |
| | D50% [cGy] | 4075 ± 17 | 3887 ± 13 | 4123 ± 74 | 4138 ± 64 | 4131 ± 75 | 4143 ± 92 | 4146 ± 63 | 4168 ± 83 | 4212 ± 169 | 4254 ± 129 |
| | D95% [cGy] | 3880 ± 8 | 3591 ± 110 | 3880 ± 8 | 3880 ± 8 | 3880 ± 8 | 3880 ± 8 | 3880 ± 8 | 3880 ± 8 | 3880 ± 8 | 3880 ± 8 |
| | D98% [cGy] | 3828 ± 3 | 3459 ± 162 | 3799 ± 60 | 3800 ± 37 | 3788 ± 98 | 3786 ± 115 | 3788 ± 63 | 3783 ± 68 | 3742 ± 148 | 3745 ± 119 |
| | D99% [cGy] | 3801 ± 1 | 3372 ± 203 | 3740 ± 116 | 3743 ± 71 | 3722 ± 179 | 3717 ± 206 | 3723 ± 115 | 3714 ± 127 | 3652 ± 228 | 3659 ± 179 |
| | V100% [% of volume] | 83.02 ± 2.33 | 0.00 ± 0.00 | 85.12 ± 2.60 | 85.36 ± 2.12 | 85.57 ± 2.80 | 85.13 ± 3.03 | 85.89 ± 1.93 | 86.09 ± 2.58 | 87.25 ± 3.06 | 87.64 ± 3.1 |
| | V105% [% of volume] | 0.14 ± 0.30 | 0.00 ± 0.00 | 9.11 ± 22.70 | 12.72 ± 24.46 | 11.77 ± 25.09 | 11.83 ± 26.59 | 10.56 ± 22.87 | 21.79 ± 32.42 | 30.14 ± 35.16 | 44.86 ± 32.99 |
| | V95% [% of volume] | 99.02 ± 0.04 | 81.99 ± 4.62 | 98.32 ± 0.80 | 98.14 ± 0.70 | 98.22 ± 0.83 | 98.36 ± 0.89 | 98 ± 0.67 | 97.92 ± 0.71 | 97.67 ± 0.98 | 97.60 ± 10 |
| | CI [%] | 0.96 ± 0.01 | 0.86 ± 0.21 | 0.93 ± 0.03 | 0.92 ± 0.04 | 0.94 ± 0.03 | 0.94 ± 0.03 | 0.93 ± 0.03 | 0.93 ± 0.08 | 0.91 ± 0.06 | 0.90 ± 0.05 |
| | HI [%] | 0.08 ± 0.00 | 0.13 ± 0.04 | 0.09 ± 0.03 | 0.10 ± 0.02 | 0.10 ± 0.04 | 0.10 ± 0.04 | 0.10 ± 0.03 | 0.11 ± 0.03 | 0.12 ± 0.07 | 0.13 ± 0.05 |
| **Bladder** | D1% [cGy] | 3925 ± 226 | 3598 ± 290 | 3938 ± 235 | 3893 ± 243 | 3932 ± 233 | 3914 ± 244 | 3883 ± 256 | 3959 ± 192 | 3927 ± 254 | 3911 ± 196 |
| | D1cc [cGy] | 4099 ± 40 | 3832 ± 108 | 4135 ± 71 | 4115 ± 75 | 4135 ± 72 | 4126 ± 73 | 4107 ± 80 | 4162 ± 92 | 4164 ± 109 | 4181 ± 109 |
| | D2% [cGy] | 3577 ± 405 | 3266 ± 457 | 3616 ± 423 | 3540 ± 411 | 3580 ± 433 | 3563 ± 439 | 3528 ± 425 | 3611 ± 390 | 3587 ± 440 | 3538 ± 371 |
| | V15% [% of volume] | 25.73 ± 13.55 | 24.12 ± 11.59 | 28.55 ± 16.38 | 26.34 ± 14.42 | 26.53 ± 15.55 | 26.09 ± 14.86 | 25.98 ± 13.77 | 26.94 ± 15.10 | 26.23 ± 14.30 | 26.39 ± 13.72 |
| | V20% [% of volume] | 22.61 ± 12.05 | 20.86 ± 10.15 | 24.56 ± 14.69 | 22.72 ± 12.76 | 22.93 ± 13.92 | 22.55 ± 13.3 | 22.41 ± 12.12 | 23.26 ± 13.42 | 22.72 ± 12.86 | 22.78 ± 12.23 |
| | V50% [% of volume] | 10.12 ± 5.09 | 8.79 ± 4.25 | 10.63 ± 6.33 | 10.02 ± 5.44 | 10.13 ± 5.84 | 10.02 ± 5.53 | 9.89 ± 4.91 | 10.34 ± 5.51 | 10.22 ± 5.40 | 10.12 ± 5.13 |
| | V80% [% of volume] | 3.60 ± 1.84 | 2.75 ± 1.56 | 4.07 ± 2.68 | 3.61 ± 2.03 | 3.78 ± 2.31 | 3.67 ± 2.14 | 3.50 ± 1.80 | 3.87 ± 2.15 | 3.75 ± 2.07 | 3.6 ± 1.91 |
| | V95% [% of volume] | 1.86 ± 1.02 | 0.79 ± 0.70 | 2.11 ± 1.66 | 1.74 ± 1.10 | 1.99 ± 1.36 | 1.88 ± 1.18 | 1.68 ± 0.96 | 1.99 ± 1.20 | 1.90 ± 1.16 | 1.68 ± 1.00 |
| **Rectum** | D1% [cGy] | 3979 ± 71 | 3818 ± 101 | 4014 ± 128 | 4010 ± 69 | 3999 ± 96 | 4020 ± 111 | 4038 ± 75 | 3996 ± 102 | 4065 ± 181 | 4084 ± 138 |
| | D1cc [cGy] | 3942 ± 82 | 3798 ± 121 | 3973 ± 142 | 3975 ± 79 | 3958 ± 104 | 3978 ± 111 | 4003 ± 73 | 3953 ± 115 | 4025 ± 178 | 4047 ± 146 |
| | D2% [cGy] | 3883 ± 114 | 3744 ± 149 | 3903 ± 177 | 3914 ± 108 | 3888 ± 133 | 3914 ± 145 | 3944 ± 115 | 3889 ± 130 | 3953 ± 199 | 3976 ± 143 |
| | V15% [% of volume] | 66.68 ± 15.64 | 67.8 ± 15.55 | 70.02 ± 14.49 | 68.92 ± 14.64 | 68.99 ± 14.49 | 68.75 ± 14.44 | 68.75 ± 14.76 | 69.11 ± 14.64 | 68.94 ± 14.28 | 69.10 ± 14.58 |
| | V20% [% of volume] | 64.22 ± 15.44 | 65.15 ± 15.43 | 67.01 ± 14.33 | 66.15 ± 14.51 | 66.22 ± 14.37 | 65.96 ± 14.27 | 65.96 ± 14.62 | 66.34 ± 14.52 | 66.09 ± 14.12 | 66.27 ± 14.49 |
| | V50% [% of volume] | 38.29 ± 9.37 | 37.41 ± 9.96 | 35.67 ± 9.69 | 37.91 ± 9.07 | 37.91 ± 10.55 | 39.28 ± 10.48 | 40.39 ± 10.22 | 38.11 ± 9.83 | 40.54 ± 11.22 | 42.04 ± 10.92 |
| | V80% [% of volume] | 8.84 ± 2.45 | 10.04 ± 4.49 | 8.91 ± 2.65 | 10.25 ± 3.13 | 8.79 ± 2.60 | 9.09 ± 2.53 | 10.37 ± 3.10 | 9.36 ± 2.73 | 9.80 ± 2.96 | 11.05 ± 2.93 |
| | V95% [% of volume] | 3.19 ± 1.36 | 2.43 ± 2.01 | 3.32 ± 1.51 | 3.67 ± 1.50 | 3.12 ± 1.19 | 3.41 ± 1.40 | 3.94 ± 1.68 | 3.13 ± 1.26 | 3.76 ± 1.91 | 4.04 ± 1.67 |
| **Femur-LT** | D1% [cGy] | 1153 ± 164 | 1090 ± 152 | 1270 ± 187 | 1209 ± 176 | 1211 ± 169 | 1198 ± 161 | 1158 ± 145 | 1232 ± 182 | 1210 ± 161 | 1211 ± 168 |
| | D1cc [cGy] | 1190 ± 174 | 1123 ± 162 | 1305 ± 191 | 1243 ± 182 | 1247 ± 176 | 1236 ± 167 | 1190 ± 151 | 1268 ± 186 | 1248 ± 169 | 1247 ± 176 |
| | D2% [cGy] | 1100 ± 155 | 1042 ± 142 | 1216 ± 177 | 1157 ± 164 | 1156 ± 160 | 1142 ± 151 | 1108 ± 135 | 1179 ± 172 | 1154 ± 150 | 1157 ± 157 |
| | D5% [cGy] | 1009 ± 140 | 957 ± 125 | 1118 ± 162 | 1063 ± 146 | 1058 ± 146 | 1041 ± 139 | 1017 ± 120 | 1082 ± 156 | 1051 ± 134 | 1058 ± 140 |
| **Femur-RT** | D1% [cGy] | 1214 ± 219 | 1155 ± 207 | 1286 ± 189 | 1237 ± 183 | 1203 ± 171 | 1157 ± 165 | 1172 ± 166 | 1216 ± 191 | 1190 ± 160 | 1194 ± 186 |
| | D1cc [cGy] | 1255 ± 226 | 1194 ± 213 | 1320 ± 194 | 1272 ± 186 | 1236 ± 175 | 1191 ± 169 | 1204 ± 170 | 1249 ± 195 | 1224 ± 165 | 1230 ± 190 |
| | D2% [cGy] | 1154 ± 204 | 1099 ± 192 | 1235 ± 179 | 1184 ± 173 | 1154 ± 162 | 1108 ± 157 | 1124 ± 158 | 1165 ± 181 | 1139 ± 153 | 1143 ± 178 |
| | D5% [cGy] | 1049 ± 172 | 998 ± 161 | 1144 ± 163 | 1090 ± 158 | 1067 ± 148 | 1021 ± 145 | 1036 ± 145 | 1075 ± 164 | 1048 ± 141 | 1052 ± 162 |

**Table 8. 3.** The results of paired, two-tailed, T-test for $D_{max}$, $D_{min}$, and $D_{mean.}$ The fields for which $.01<p\leq.05$ were highlighted in green (i.e. significant), while those with $p\leq.01$ were highlighted in red (i.e. very significant).

| | | | Adaptive Radiation Therapy Strategies | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Structure** | **Metric** | **non-ART** | **Cont.+DF** | **DF 2-4** | **Online+DF** | **Online** | **Online 1-3-5** | **Offline+DF** | **Offline+Online** | **Ofline** |
| **CTV** | D1% [cGy] | 0.0000 | 0.0763 | 0.0197 | 0.0535 | 0.0550 | 0.0078 | 0.0011 | 0.0075 | 0.0001 |
| | D2% [cGy] | 0.0000 | 0.0654 | 0.0180 | 0.0488 | 0.0459 | 0.0057 | 0.0011 | 0.0068 | 0.0001 |
| | D5% [cGy] | 0.0000 | 0.0514 | 0.0146 | 0.0398 | 0.0349 | 0.0035 | 0.0011 | 0.0057 | 0.0001 |
| | D50% [cGy] | 0.0000 | 0.0146 | 0.0021 | 0.0061 | 0.0050 | 0.0002 | 0.0002 | 0.0022 | 0.0000 |
| | D95% [cGy] | 0.0000 | 0.0043 | 0.0001 | 0.0007 | 0.0011 | 0.0000 | 0.0001 | 0.0014 | 0.0000 |
| | D98% [cGy] | 0.0000 | 0.0047 | 0.0000 | 0.0006 | 0.0011 | 0.0000 | 0.0002 | 0.0009 | 0.0000 |
| | D99% [cGy] | 0.0000 | 0.0065 | 0.0000 | 0.0006 | 0.0011 | 0.0000 | 0.0004 | 0.0005 | 0.0000 |
| | V100% [% of volume] | 0.0000 | 0.9355 | 0.0047 | 0.0042 | 0.0332 | 0.0025 | 0.6075 | 0.4931 | 0.8273 |
| | V105% [% of volume] | 0.1734 | 0.1173 | 0.0367 | 0.0606 | 0.0650 | 0.0585 | 0.0090 | 0.0010 | 0.0000 |
| | HI [%] | 0.0001 | 0.0041 | 0.0071 | 0.0000 | 0.0000 | 0.0000 | 0.6951 | 0.5128 | 0.9334 |
| **PTV** | D1% [cGy] | 0.0000 | 0.0451 | 0.0204 | 0.0441 | 0.0932 | 0.0179 | 0.0010 | 0.0111 | 0.0002 |
| | D2% [cGy] | 0.0000 | 0.0411 | 0.0179 | 0.0422 | 0.0698 | 0.0108 | 0.0009 | 0.0094 | 0.0001 |
| | D5% [cGy] | 0.0000 | 0.0360 | 0.0140 | 0.0337 | 0.0456 | 0.0051 | 0.0009 | 0.0073 | 0.0001 |
| | D50% [cGy] | 0.0000 | 0.0092 | 0.0007 | 0.0039 | 0.0043 | 0.0001 | 0.0001 | 0.0021 | 0.0000 |
| | D95% [cGy] | 0.0000 | 0.2591 | 0.3535 | 0.1183 | 0.9169 | 0.2698 | 0.5916 | 0.0912 | 0.1224 |
| | D98% [cGy] | 0.0000 | 0.0380 | 0.0017 | 0.0749 | 0.1145 | 0.0091 | 0.0065 | 0.0161 | 0.0049 |
| | D99% [cGy] | 0.0000 | 0.0316 | 0.0018 | 0.0635 | 0.0840 | 0.0072 | 0.0065 | 0.0089 | 0.0022 |
| | V100% [% of volume] | 0.0000 | 0.0056 | 0.0067 | 0.0024 | 0.0119 | 0.0007 | 0.0015 | 0.0001 | 0.0000 |
| | V105% [% of volume] | 0.0474 | 0.0935 | 0.0334 | 0.0524 | 0.0644 | 0.0563 | 0.0076 | 0.0012 | 0.0000 |
| | V95% [% of volume] | 0.0000 | 0.0010 | 0.0000 | 0.0004 | 0.0035 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | CI [%] | 0.0606 | 0.0048 | 0.0018 | 0.0100 | 0.1022 | 0.0001 | 0.2266 | 0.0015 | 0.0003 |
| | HI [%] | 0.0001 | 0.0340 | 0.0079 | 0.0520 | 0.0896 | 0.0088 | 0.0011 | 0.0081 | 0.0004 |
| **Bladder** | D1% [cGy] | 0.0000 | 0.4900 | 0.0239 | 0.6020 | 0.3387 | 0.0106 | 0.0914 | 0.9240 | 0.4712 |
| | D1cc [cGy] | 0.0000 | 0.0110 | 0.2207 | 0.0240 | 0.1030 | 0.5861 | 0.0061 | 0.0136 | 0.0025 |
| | D2% [cGy] | 0.0001 | 0.2156 | 0.1071 | 0.9276 | 0.6059 | 0.0393 | 0.2351 | 0.7476 | 0.0942 |
| | V15% [% of volume] | 0.0296 | 0.0311 | 0.4975 | 0.4878 | 0.7160 | 0.7063 | 0.2763 | 0.5604 | 0.4115 |
| | V20% [% of volume] | 0.0159 | 0.1097 | 0.8936 | 0.7677 | 0.9521 | 0.7493 | 0.5163 | 0.8967 | 0.8167 |
| | V50% [% of volume] | 0.0017 | 0.3624 | 0.7531 | 0.9740 | 0.7937 | 0.2712 | 0.5346 | 0.7781 | 0.9954 |
| | V80% [% of volume] | 0.0015 | 0.1048 | 0.9235 | 0.3488 | 0.6072 | 0.2370 | 0.1117 | 0.2749 | 0.9960 |
| | V95% [% of volume] | 0.0000 | 0.2292 | 0.0769 | 0.3171 | 0.8027 | 0.0079 | 0.2408 | 0.5762 | 0.0184 |
| **Rectum** | D1% [cGy] | 0.0000 | 0.1138 | 0.0258 | 0.2372 | 0.0527 | 0.0000 | 0.3613 | 0.0316 | 0.0029 |
| | D1cc [cGy] | 0.0000 | 0.1959 | 0.0328 | 0.3520 | 0.0683 | 0.0000 | 0.5468 | 0.0326 | 0.0020 |
| | D2% [cGy] | 0.0002 | 0.4637 | 0.0688 | 0.8081 | 0.1469 | 0.0000 | 0.7848 | 0.0680 | 0.0045 |
| | V15% [% of volume] | 0.0126 | 0.0001 | 0.0002 | 0.0009 | 0.0016 | 0.0002 | 0.0001 | 0.0012 | 0.0001 |
| | V20% [% of volume] | 0.0266 | 0.0002 | 0.0002 | 0.0018 | 0.0047 | 0.0008 | 0.0001 | 0.0052 | 0.0002 |
| | V50% [% of volume] | 0.2669 | 0.0065 | 0.5856 | 0.6954 | 0.3971 | 0.0352 | 0.8517 | 0.1745 | 0.0041 |
| | V80% [% of volume] | 0.1341 | 0.7808 | 0.0097 | 0.8991 | 0.3480 | 0.0005 | 0.1393 | 0.0518 | 0.0000 |
| | V95% [% of volume] | 0.0518 | 0.5697 | 0.0972 | 0.7096 | 0.1924 | 0.0001 | 0.7728 | 0.0747 | 0.0093 |
| **Femur-LT** | D1% [cGy] | 0.0000 | 0.0004 | 0.0049 | 0.0214 | 0.0643 | 0.7559 | 0.0032 | 0.0442 | 0.0121 |
| | D1cc [cGy] | 0.0000 | 0.0006 | 0.0079 | 0.0245 | 0.0665 | 0.9993 | 0.0038 | 0.0526 | 0.0146 |
| | D2% [cGy] | 0.0000 | 0.0004 | 0.0041 | 0.0229 | 0.0817 | 0.6301 | 0.0030 | 0.0485 | 0.0111 |
| | D5% [cGy] | 0.0000 | 0.0006 | 0.0058 | 0.0395 | 0.1770 | 0.5870 | 0.0045 | 0.0949 | 0.0169 |
| **Femur-RT** | D1% [cGy] | 0.0000 | 0.0067 | 0.1997 | 0.6714 | 0.0439 | 0.0461 | 0.9645 | 0.3893 | 0.4811 |
| | D1cc [cGy] | 0.0000 | 0.0155 | 0.3434 | 0.5038 | 0.0311 | 0.0240 | 0.8257 | 0.3048 | 0.3913 |
| | D2% [cGy] | 0.0000 | 0.0018 | 0.0741 | 0.9893 | 0.0763 | 0.1097 | 0.6797 | 0.5635 | 0.6704 |
| | D5% [cGy] | 0.0000 | 0.0001 | 0.0064 | 0.4183 | 0.1949 | 0.3828 | 0.2425 | 0.9733 | 0.8759 |
| **The percentage of statistically significant metrics at $.01< p\leq.05$** | | **88%** | **63%** | **66%** | **44%** | **29%** | **66%** | **49%** | **51%** | **68%** |
| **The percentage of statistically significant metrics at $p\leq.01$** | | **76%** | **41%** | **46%** | **24%** | **17%** | **49%** | **49%** | **34%** | **56%** |

## 8.2 Pre-processing, Training & Inference Times for U-net Model

This section contains a Part I of supplementary material for the research work described in Chapter 5 *U-net architecture with embedded Inception-ResNet-v2 image encoding modules for automatic segmentation of organs-at-risk based on computed tomography scans of head and neck cancer patients*

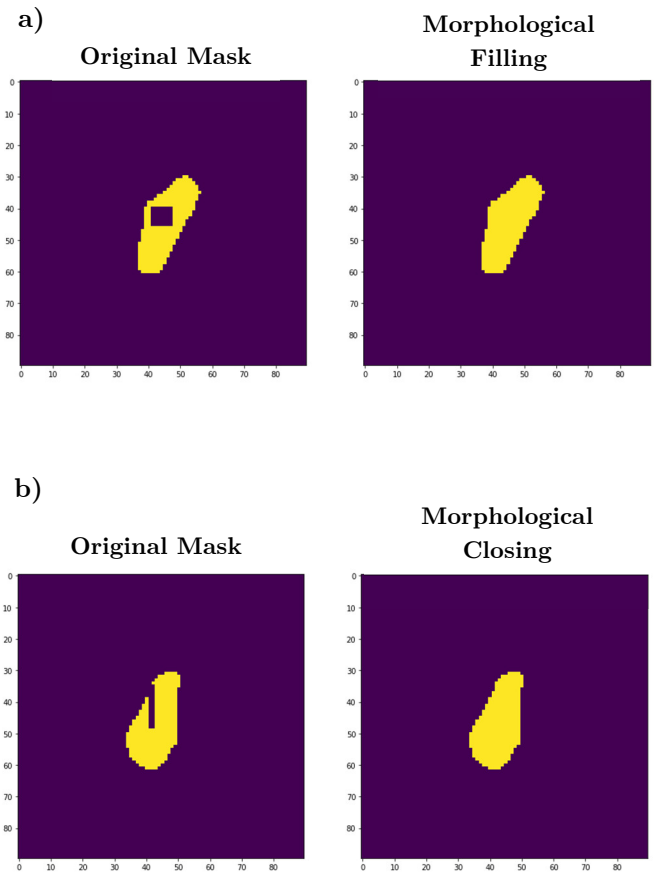



**Figure 8. 1.** The impact of morphological operations on the binary masks that were created by the conversion from DICOM contours; a) filling an empty region inside the mask; b) morphological closing of non-convex regions.

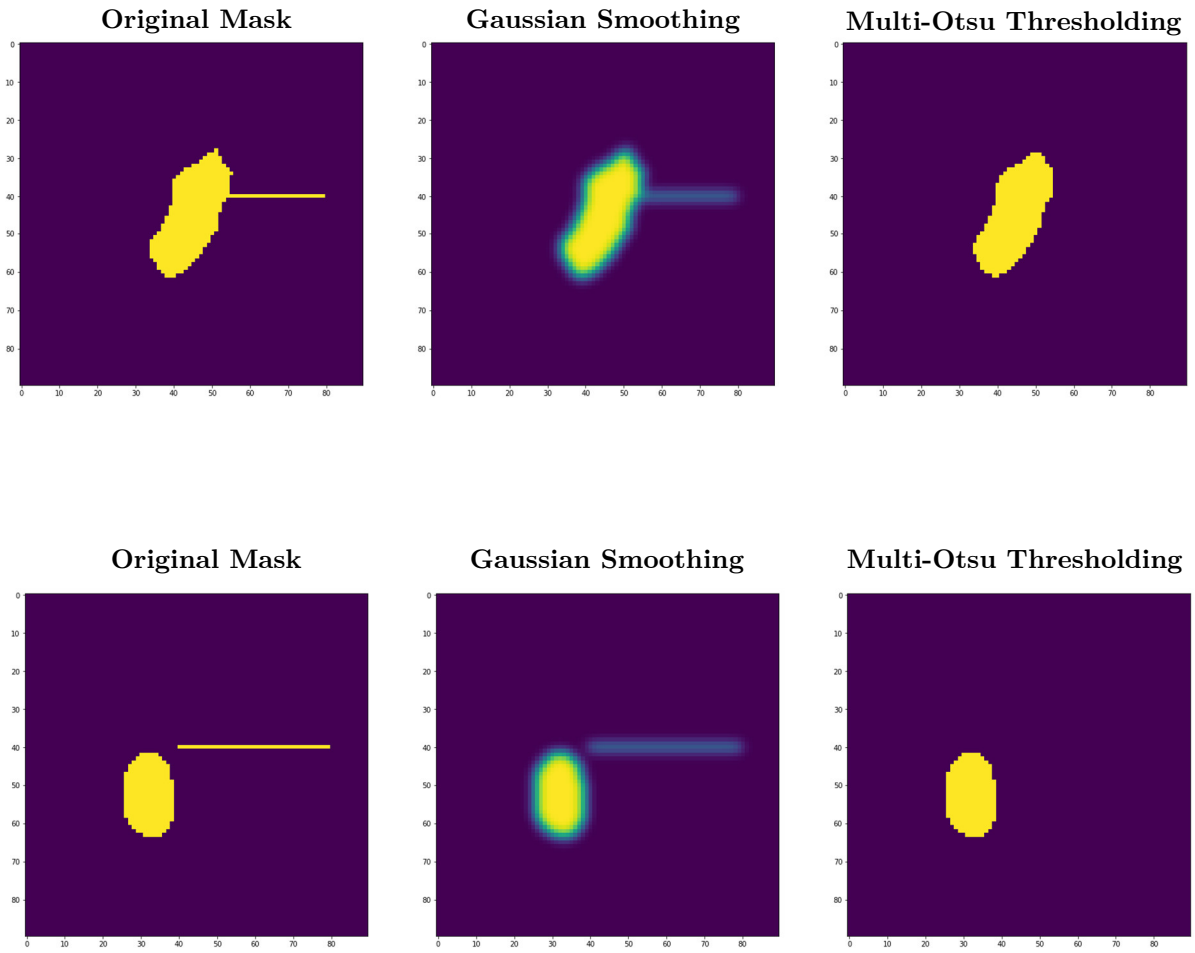



**Figure 8. 2.** The combination of Gaussian smoothing and Multi-Otsu Thresholding for outlier removal in the segmentation mask pre-processing steps.

**Table 8. 4.** Training and inference times for U-net models.

| Structure | GPU* | Training time [ms/CT slice] | Inference time [ms/CT slice] | Number of slices | Contouring time [s] | Training Time [h] |
|---|---|---|---|---|---|---|
| **Brain** | V | 21 | 8 | 54 | 0.432 | 12 |
| **Brainstem** | V | 21 | 9 | 25 | 0.225 | 12 |
| **Cochlea Left** | V | 21 | 106 | 3 | 0.318 | 2 |
| **Cochlea Right** | V | 21 | 100 | 3 | 0.300 | 2 |
| **Cord** | V | 7 | 3 | 77 | 0.231 | 7 |
| **Eye Left** | P | 34 | 16 | 12 | 0.192 | 9 |
| **Eye Right** | V | 21 | 13 | 12 | 0.156 | 6 |
| **Lacrimal Left** | P | 34 | 96 | 4 | 0.384 | 4 |
| **Lacrimal Right** | V | 21 | 92 | 3 | 0.276 | 2 |
| **Larynx** | V | 21 | 13 | 15 | 0.195 | 12 |
| **Lens Left** | V | 21 | 34 | 4 | 0.136 | 6 |
| **Lens Right** | V | 21 | 33 | 4 | 0.132 | 6 |
| **Lips** | P | 34 | 19 | 11 | 0.209 | 12 |
| **Mandible** | V | 21 | 8 | 34 | 0.272 | 12 |
| **Neck Left** | V | 21 | 14 | 44 | 0.616 | 8 |
| **Neck Right** | V | 21 | 14 | 44 | 0.616 | 8 |
| **Oesophagus** | V | 21 | 9 | 35 | 0.315 | 12 |
| **Optic Chiasm** | P | 34 | 46 | 4 | 0.184 | 5 |
| **Optic Nerve Left** | P | 34 | 32 | 4 | 0.128 | 9 |
| **Optic Nerve Right** | V | 21 | 35 | 4 | 0.140 | 6 |
| **Oral Cavity** | V | 21 | 21 | 18 | 0.378 | 6 |
| **Parotid Left** | V | 21 | 8 | 23 | 0.184 | 12 |
| **Parotid Right** | V | 21 | 8 | 24 | 0.192 | 12 |
| **Submandibular Left** | V | 21 | 23 | 14 | 0.322 | 6 |
| **Submandibular Right** | V | 28 | 20 | 14 | 0.280 | 8 |
| | **Mean** | **23** | **31** | **20** | **0.273** | **8** |
| | **Sum** | **583** | **780** | **489** | **6.813** | **193** |

* V - NVIDIA Tesla V-100-SXM2 GPU

P - NVIDIA Tesla P-100-PCIe GPU

## 8.3 Training, Testing & Explanation Visualizations for U-net Model

This section contains a Part II of supplementary material for the research work described in Chapter 5 *U-net architecture with embedded Inception-ResNet-v2 image encoding modules for automatic segmentation of organs-at-risk based on computed tomography scans of head and neck cancer patients*

To simplify the visualization of numerous charts and avoid the appearance of multiple identical descriptions, the figure descriptions were not provided in this part of the Appendix. However, descriptions can be found in Chapter 5.

For the proper data visualization, the supplementary material starts from the next page.

## Brainstem

**Figure 8. 3. Model training.**

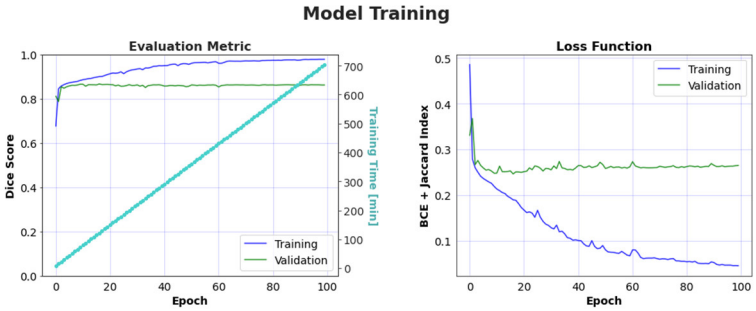



**Figure 8. 4. Model evaluation on testing data.**

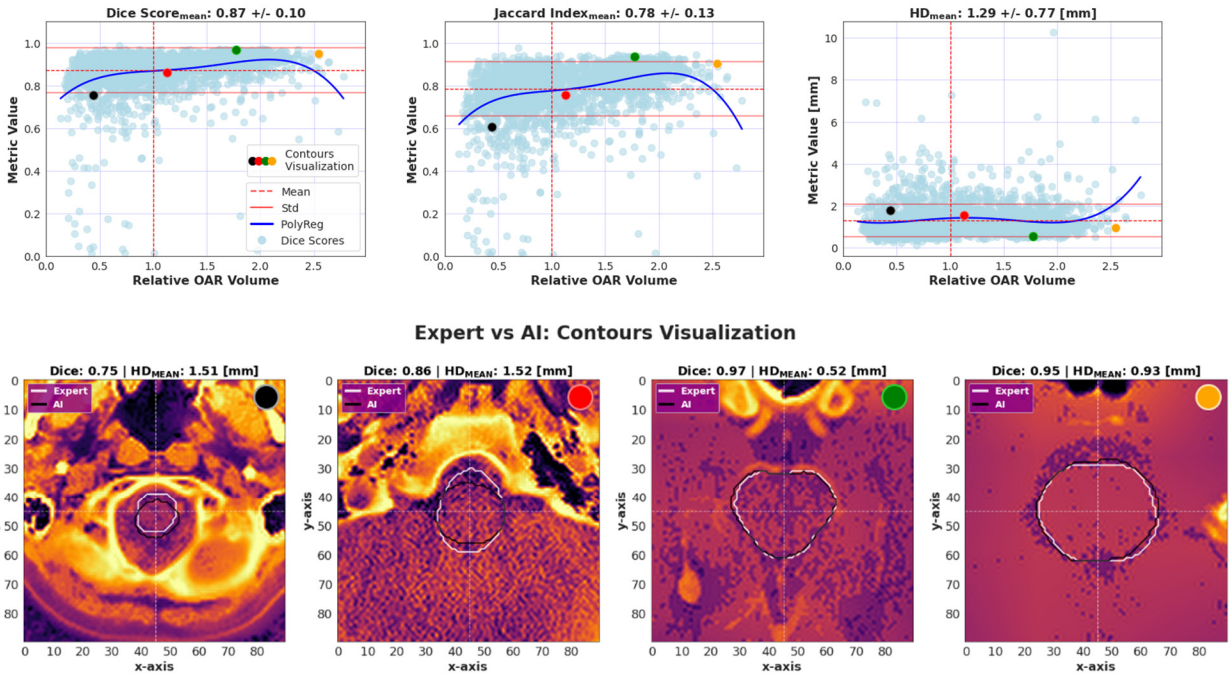
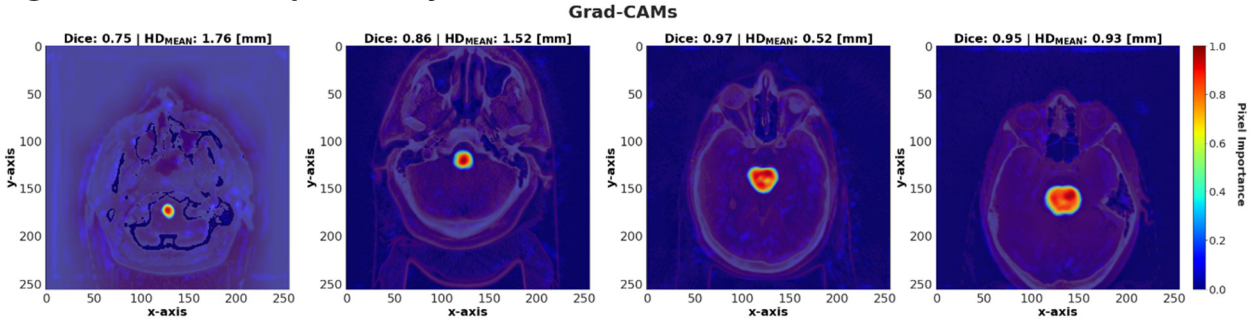



**Figure 8. 5. Model explainability.**



# Brain

**Figure 8. 6. Model training.**

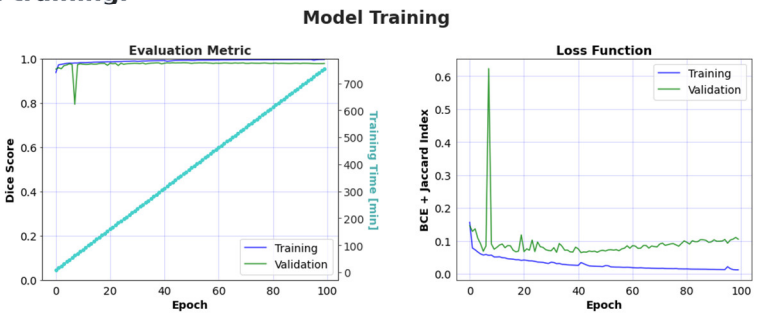



**Figure 8.7. Model evaluation on testing data.**

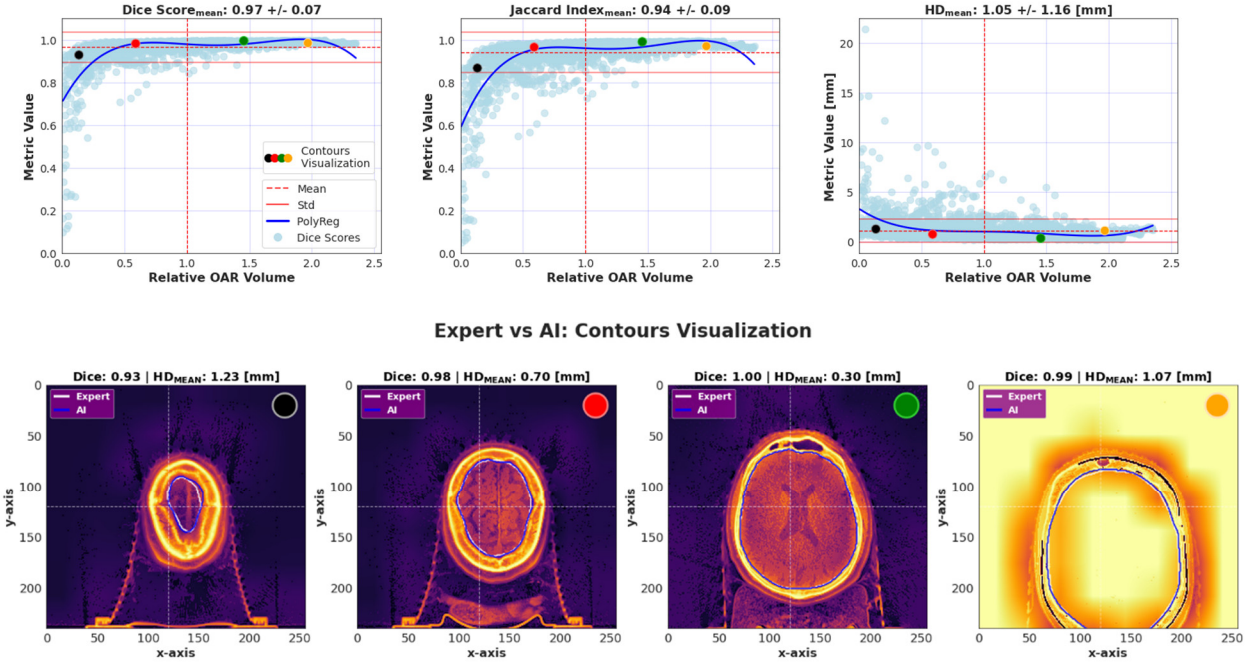
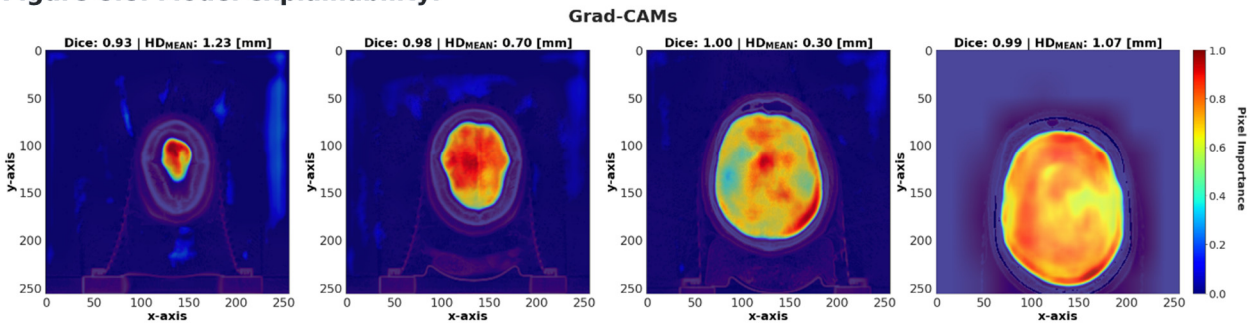



**Figure 8.8. Model explainability.**



# Cochlea Left

**Figure 8.9. Model training.**

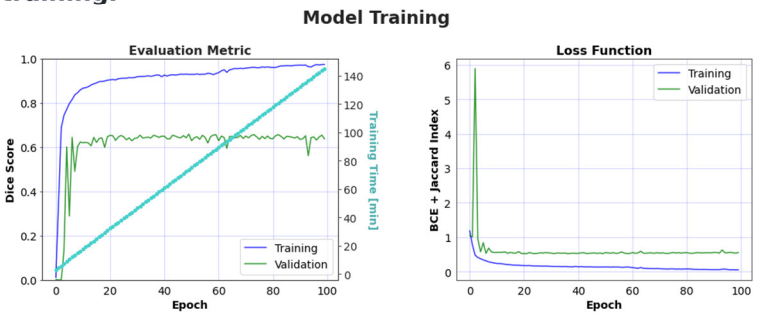



**Figure 8.10. Model evaluation on testing data.**

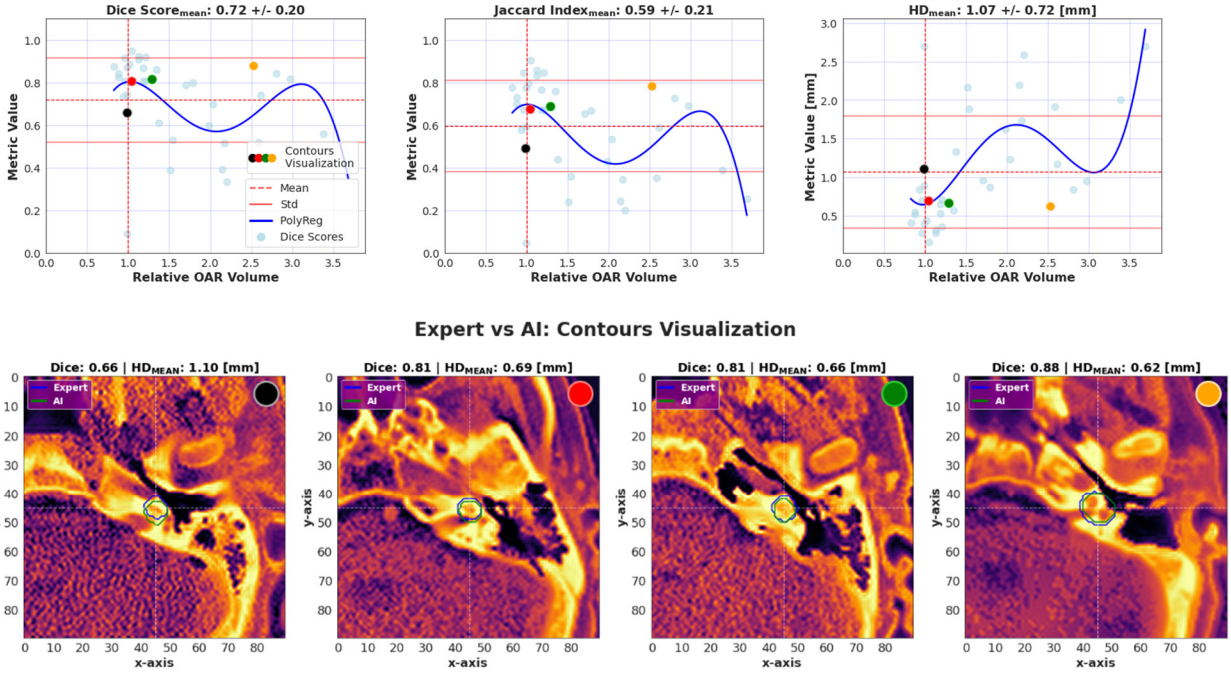
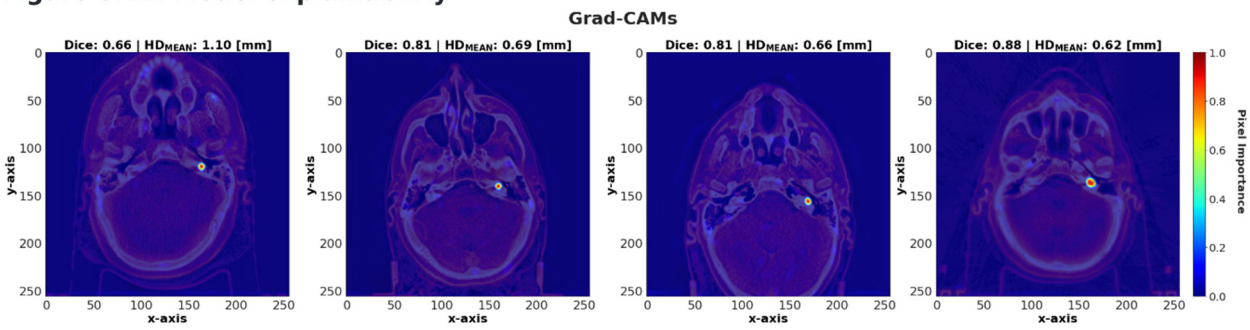



**Figure 8.11. Model explainability.**



## Cochlea Right

**Figure 8.12. Model training.**

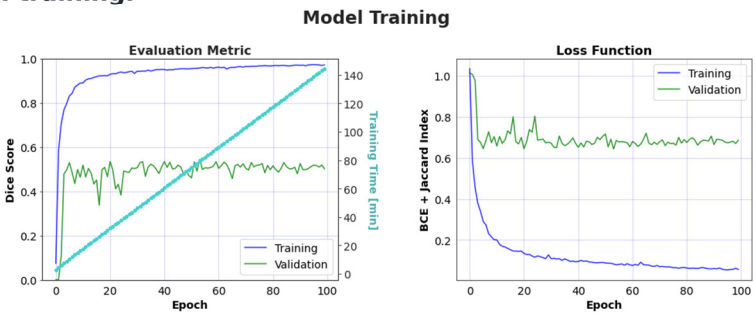



**Figure 8.13. Model evaluation on testing data.**

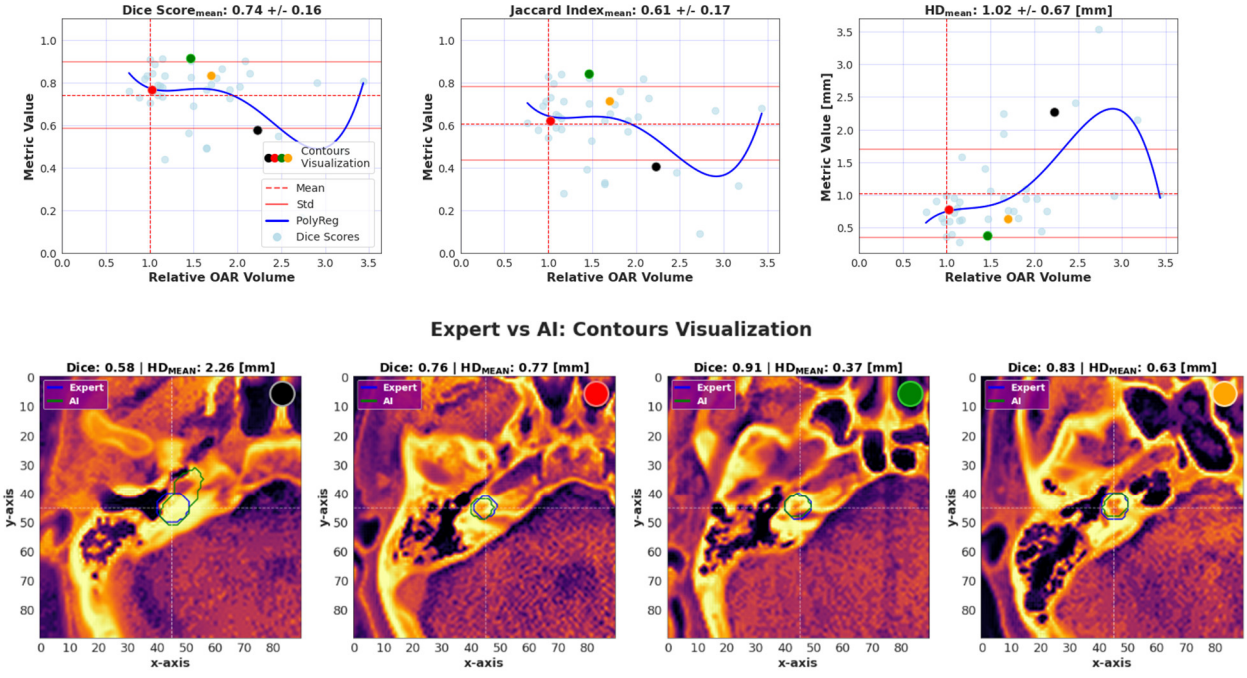
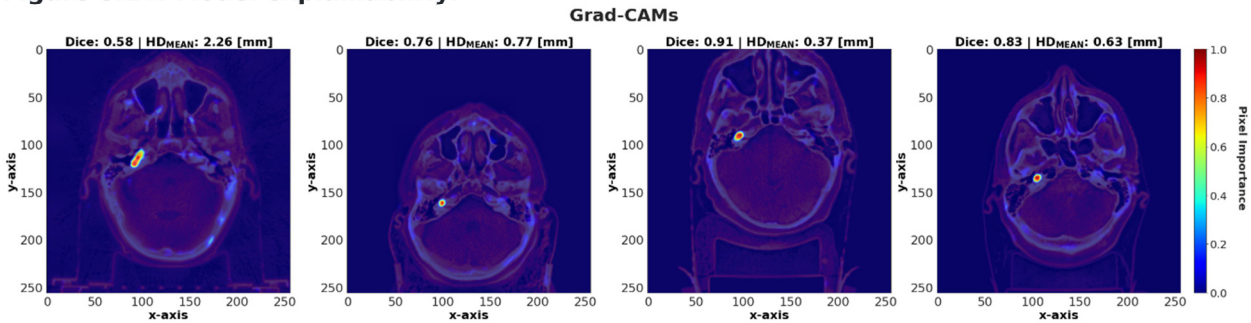



**Figure 8.14. Model explainability.**



## Cord

**Figure 8.15. Model training.**

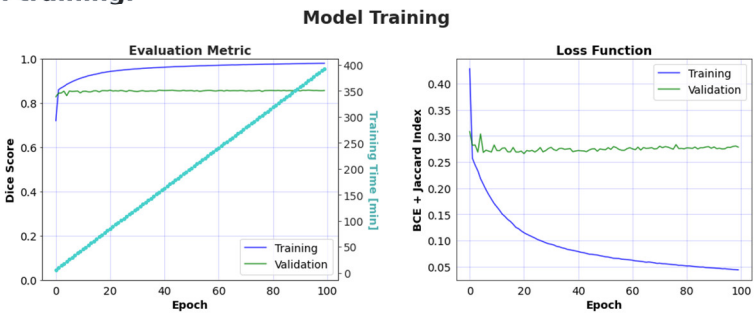



**Figure 8.16. Model evaluation on testing data.**

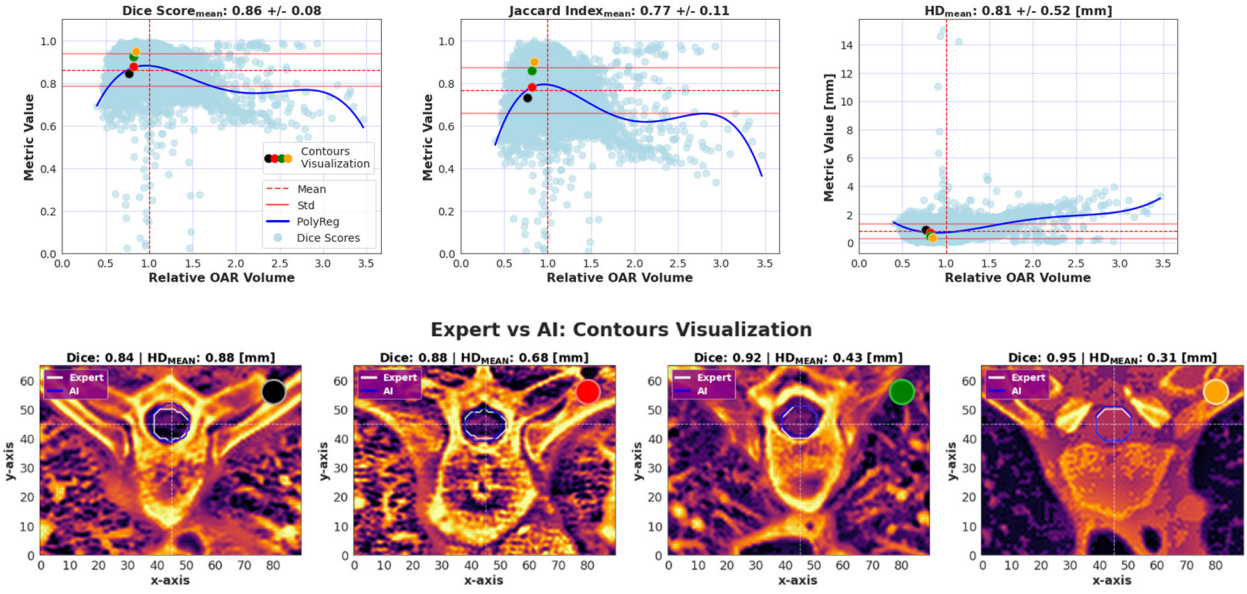
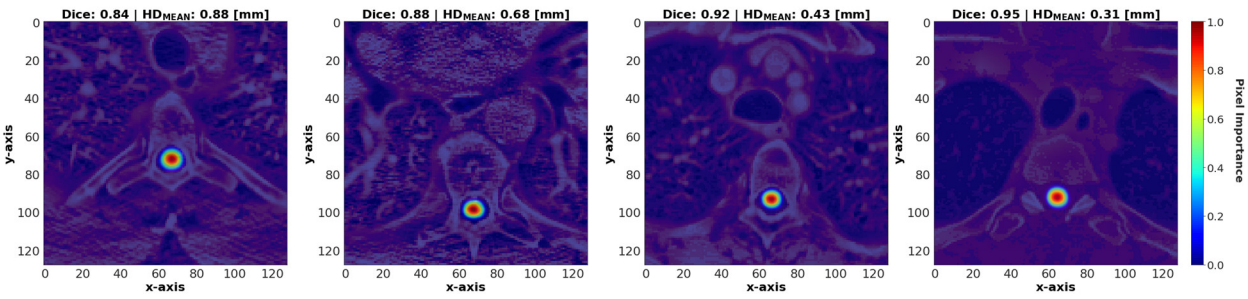



**Figure 8.17. Model explainability.**



## Eye Left

**Figure 8.18. Model training.**

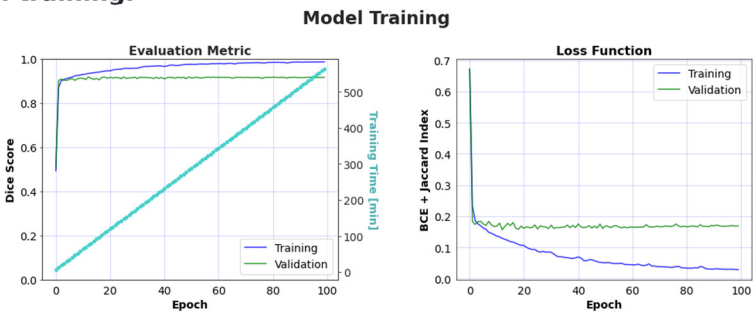



**Figure 8.19. Model evaluation on testing data.**

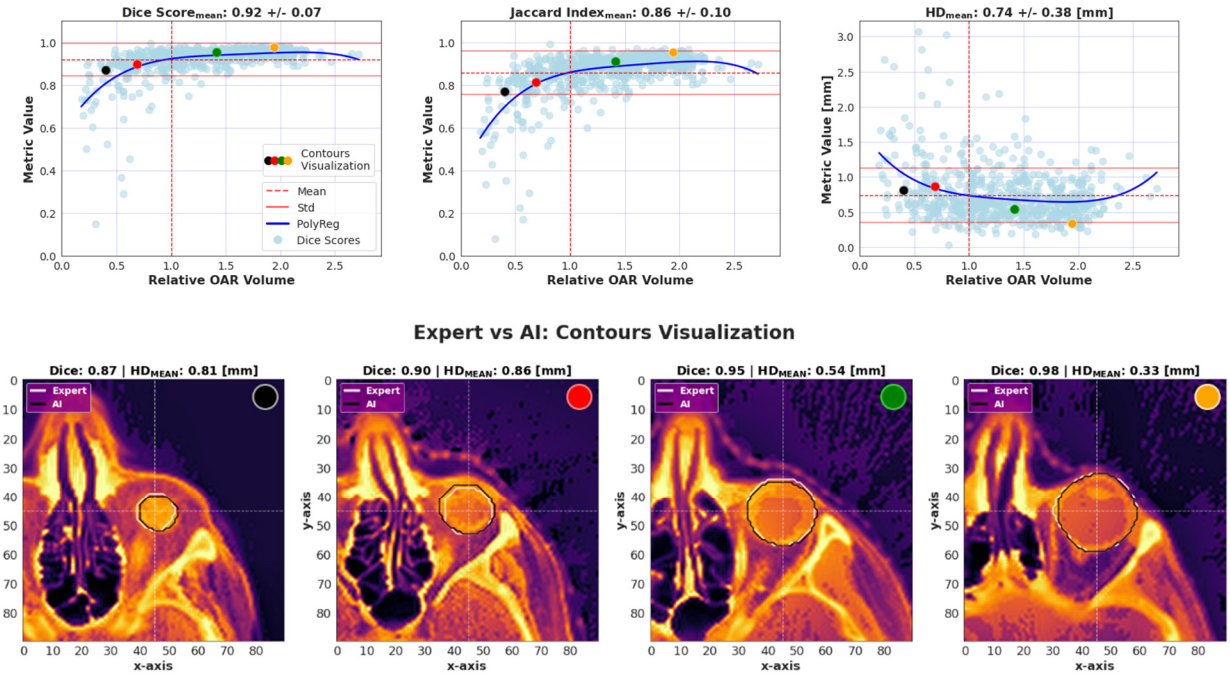
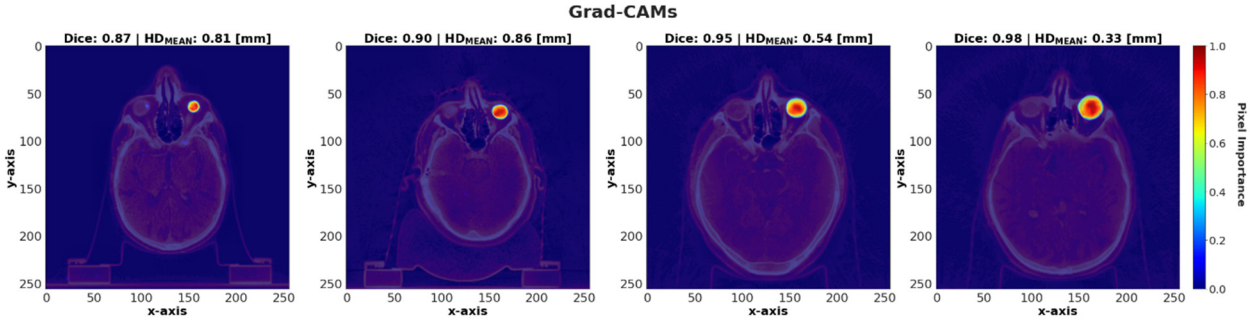



**Figure 8.20. Model explainability.**



# Eye Right

**Figure 8.21. Model training.**

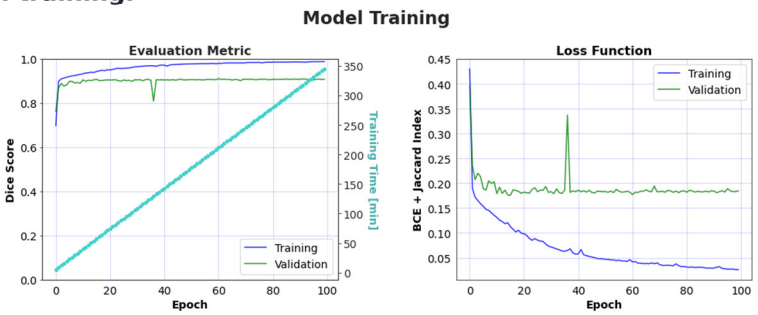
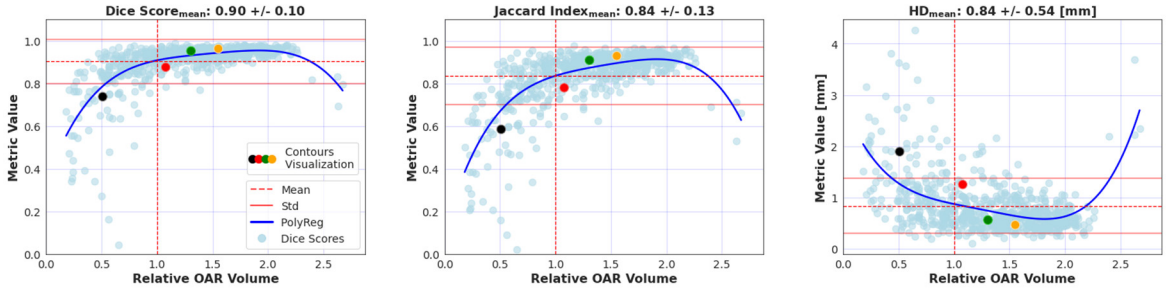



**Figure 8.22. Model evaluation on testing data.**



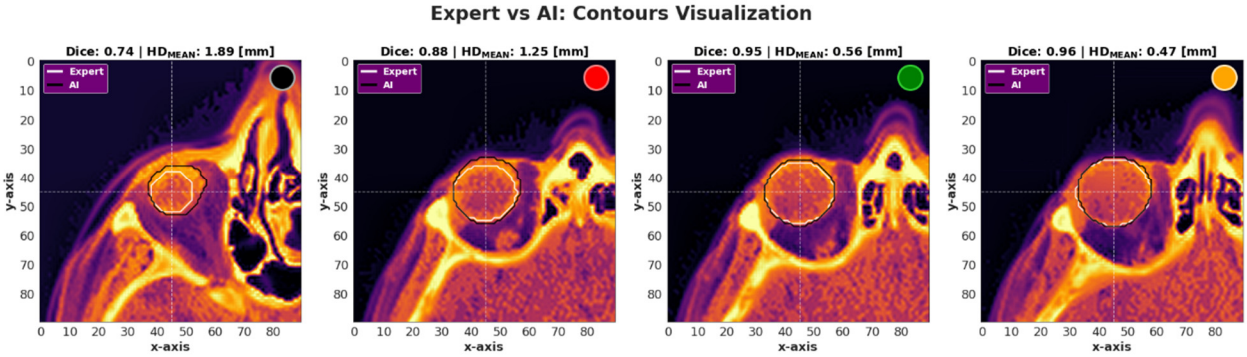
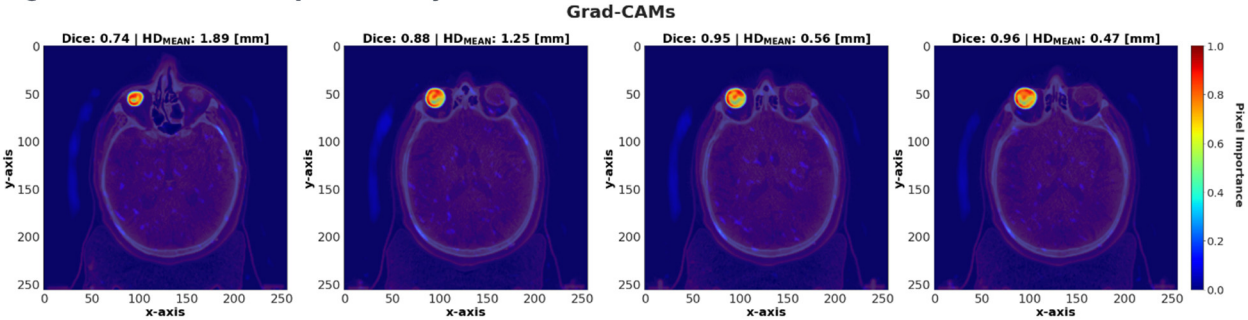




**Figure 8.23. Model explainability.**



## Lacrimal Left

**Figure 8.24. Model training.**

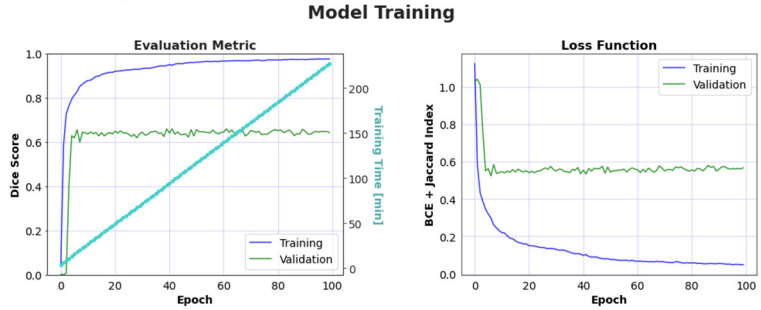
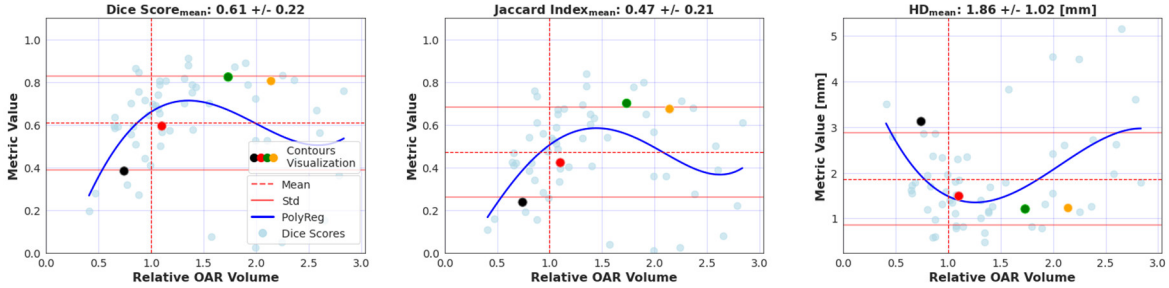



**Figure 8.25. Model evaluation on testing data.**



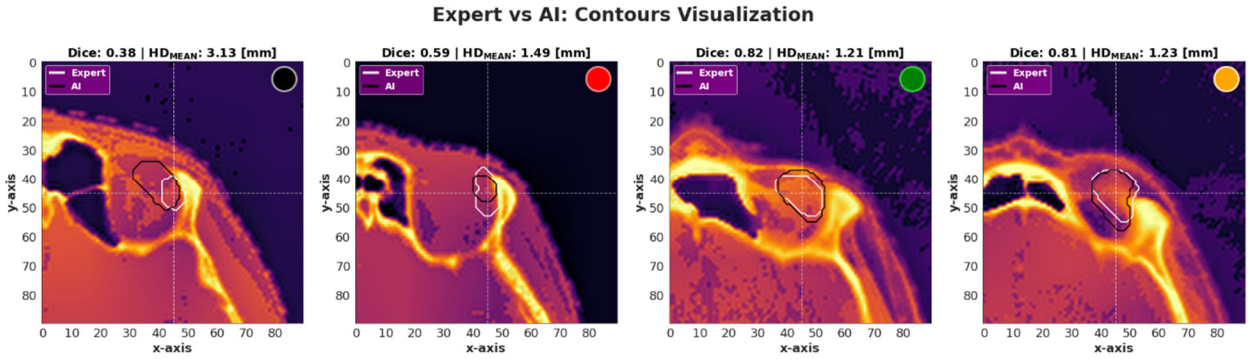
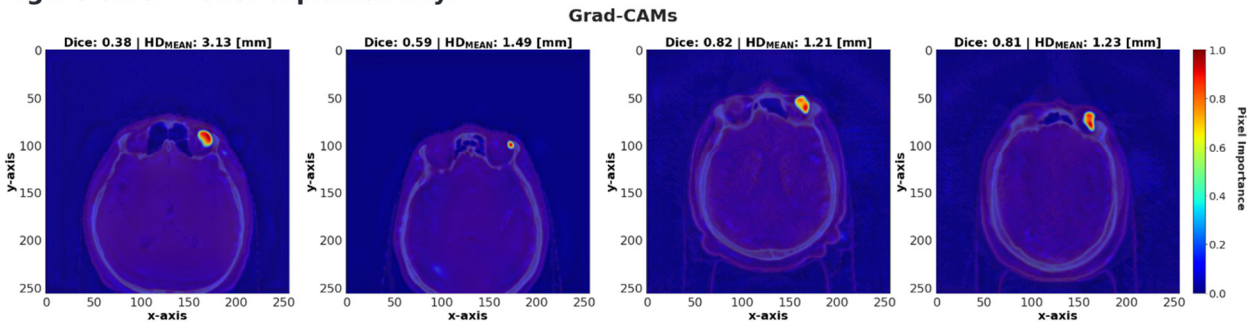




**Figure 8.26. Model explainability.**



# Lacrimal Right

**Figure 8.27. Model training.**

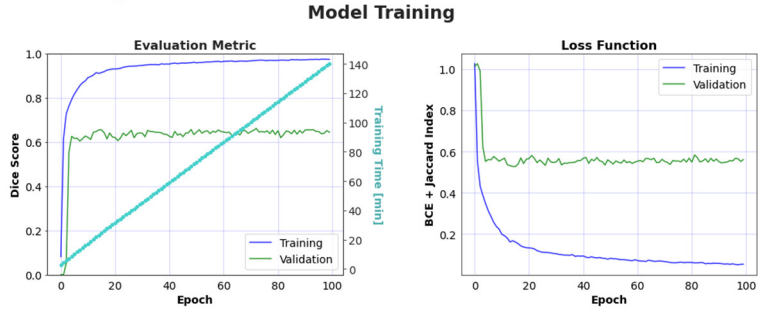



**Figure 8.28. Model evaluation on testing data.**

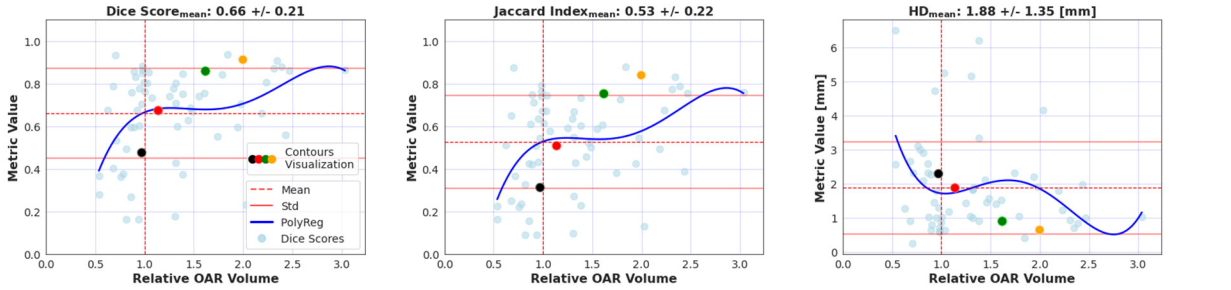
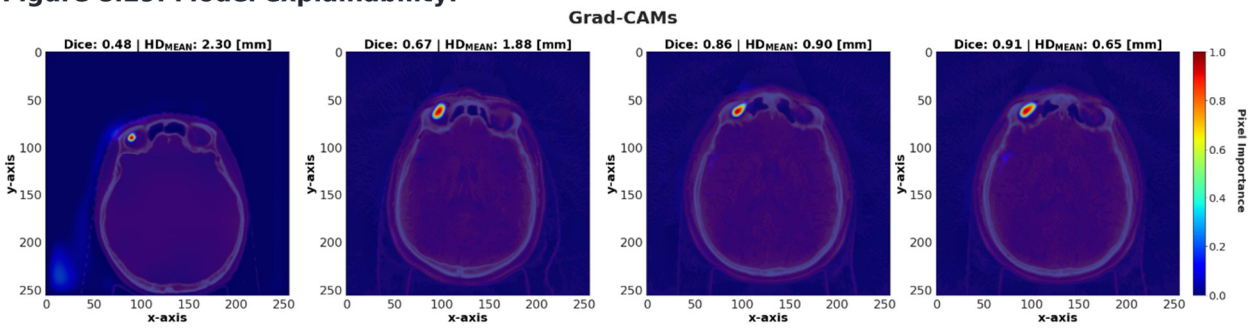



**Figure 8.29. Model explainability.**



## Larynx

**Figure 8.30. Model training.**

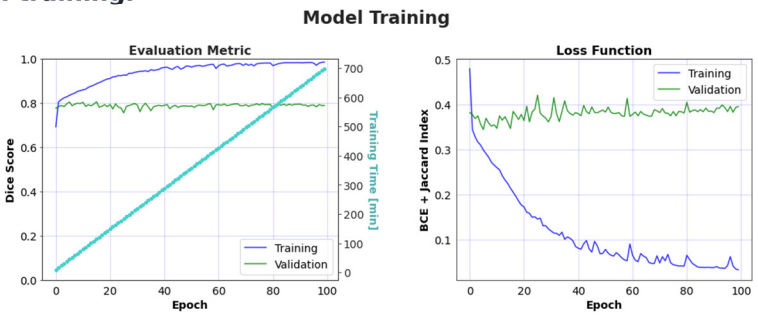
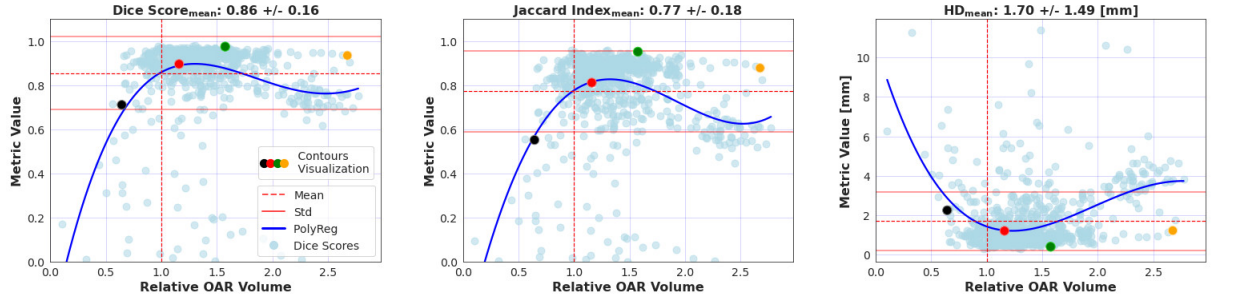



**Figure 8.31. Model evaluation on testing data.**



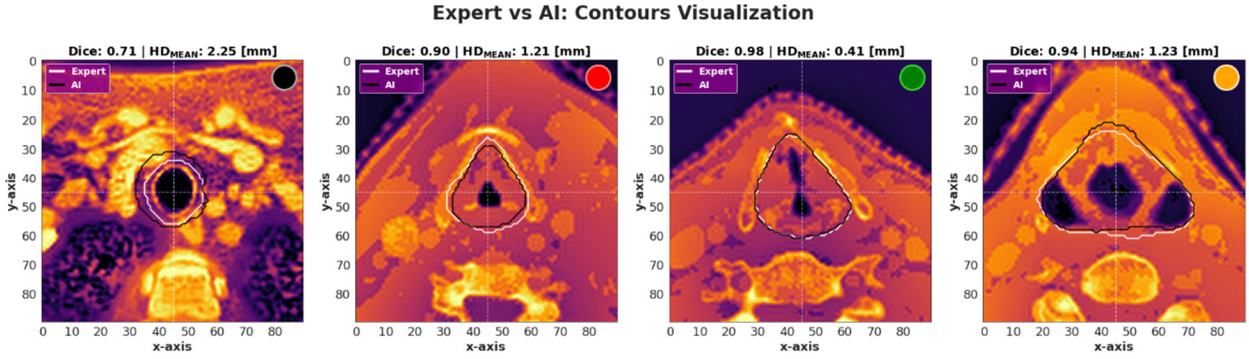
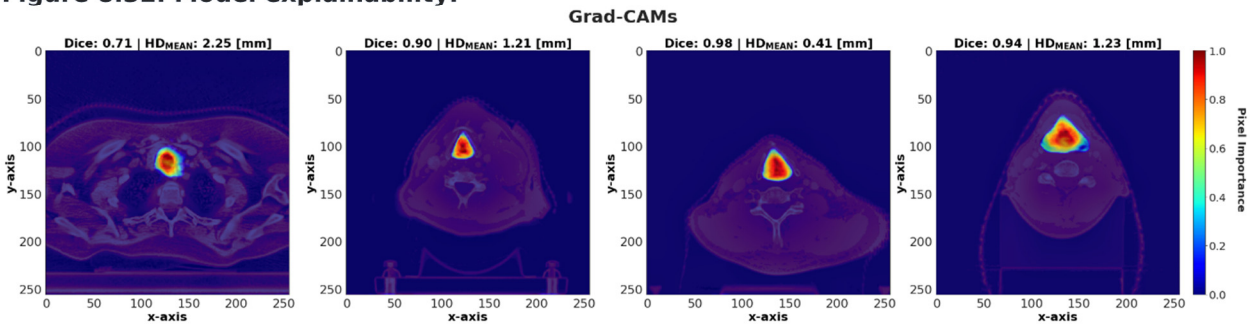




**Figure 8.32. Model explainability.**



## Lens Left

**Figure 8.33. Model training.**

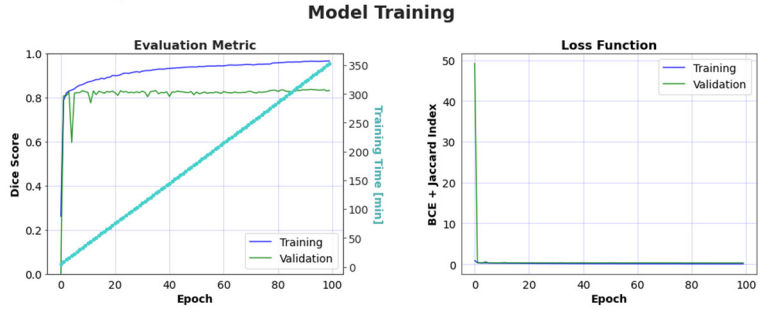
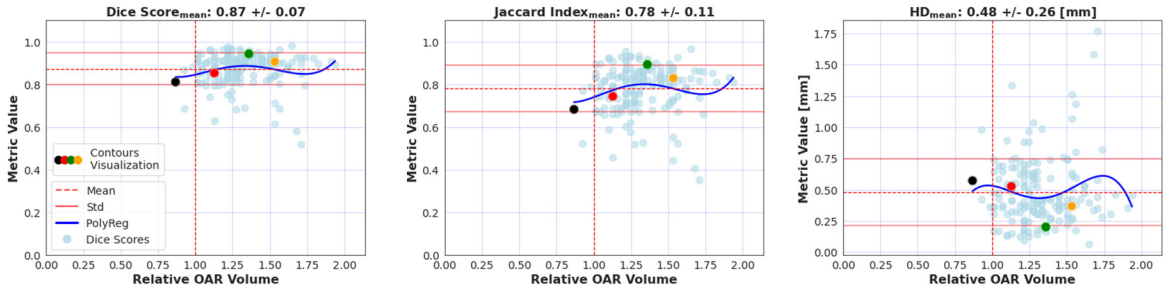



**Figure 8.34. Model evaluation on testing data.**



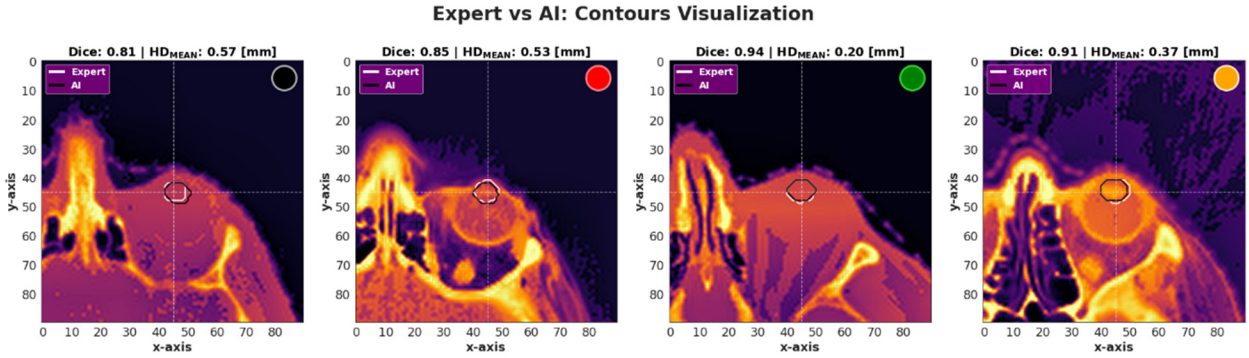
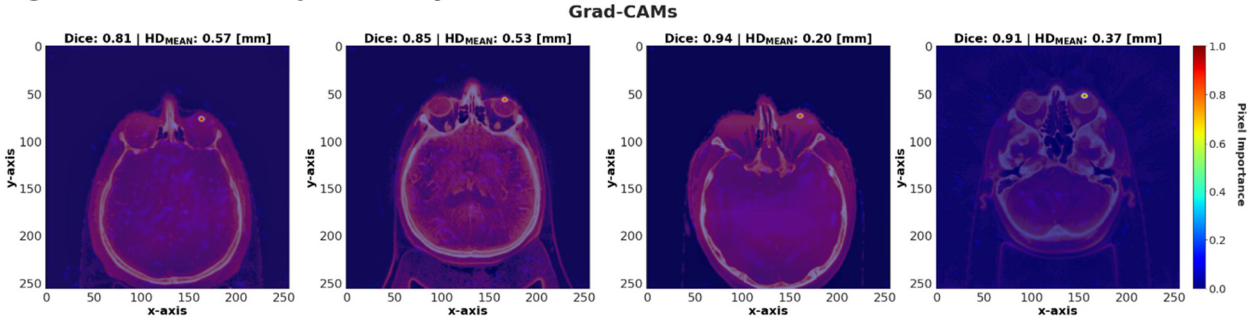




**Figure 8.35. Model explainability.**



## Lens Right

**Figure 8.36. Model training.**

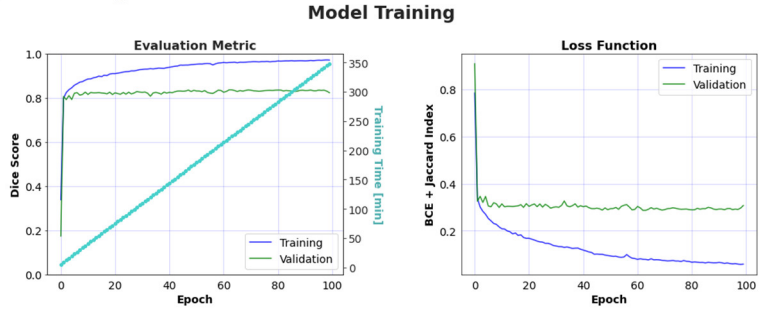
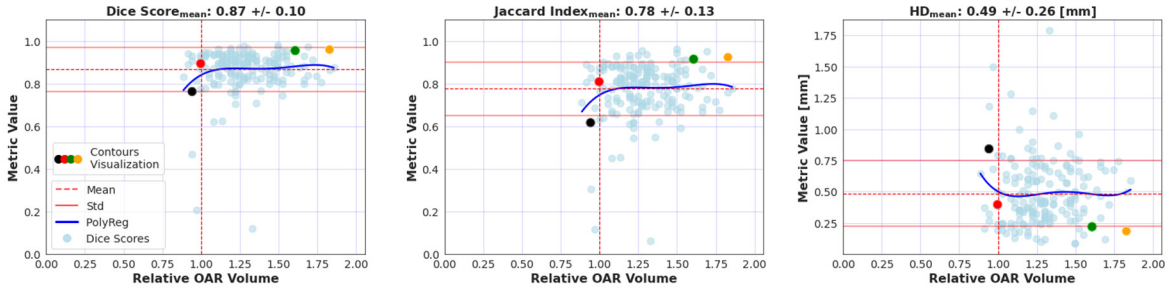



**Figure 8.37. Model evaluation on testing data.**



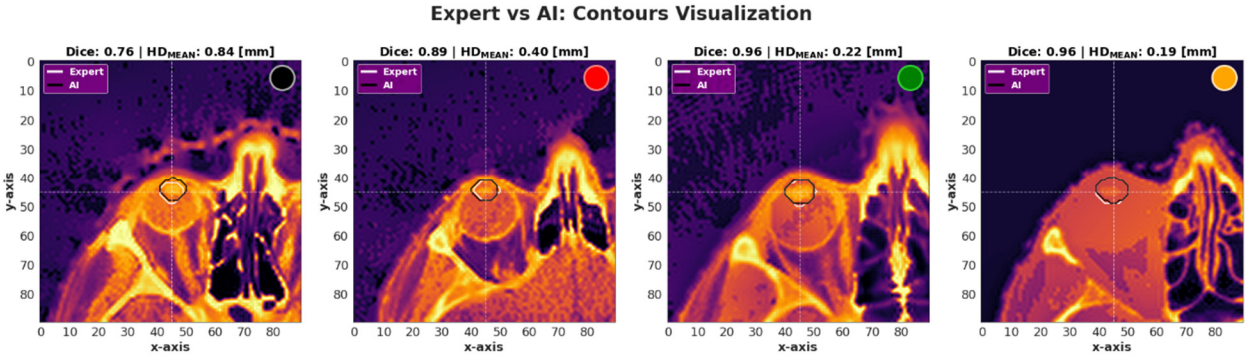
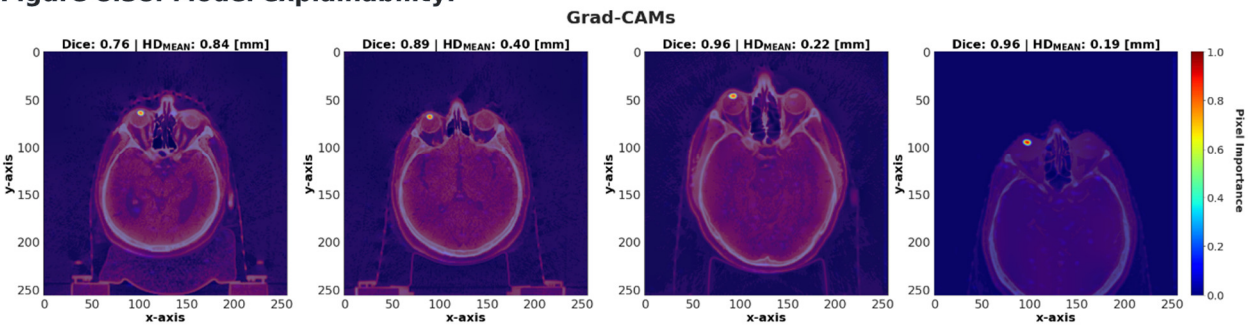




**Figure 8.38. Model explainability.**



# Lips

**Figure 8.39. Model training.**

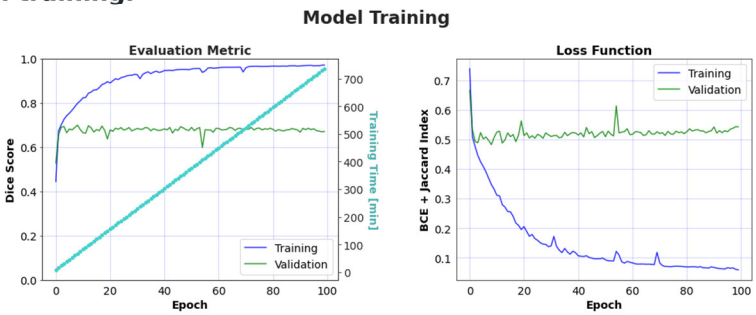



**Figure 8.40. Model evaluation on testing data.**

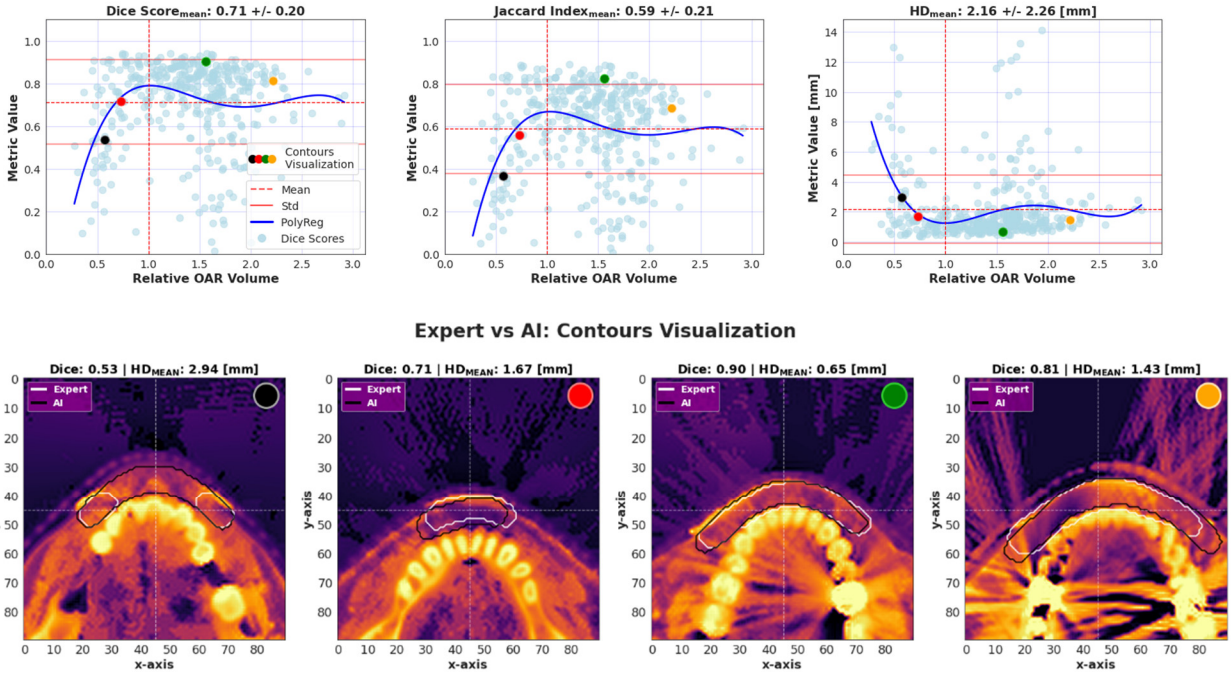
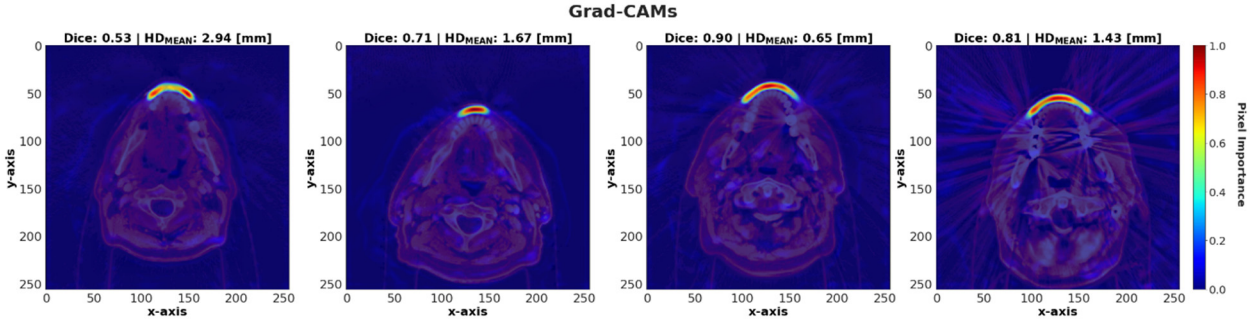



**Figure 8.41. Model explainability.**



## Mandible

**Figure 8.42. Model training.**

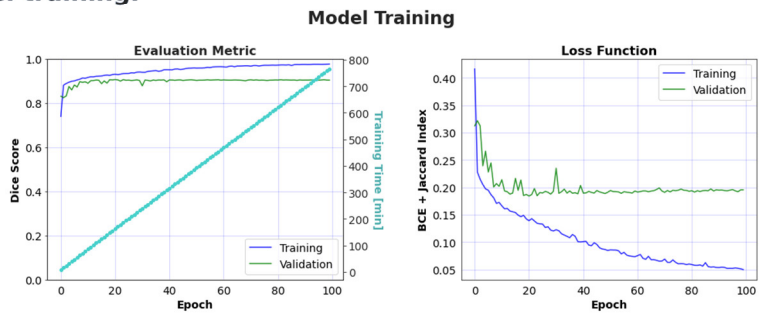



**Figure 8.43. Model evaluation on testing data.**

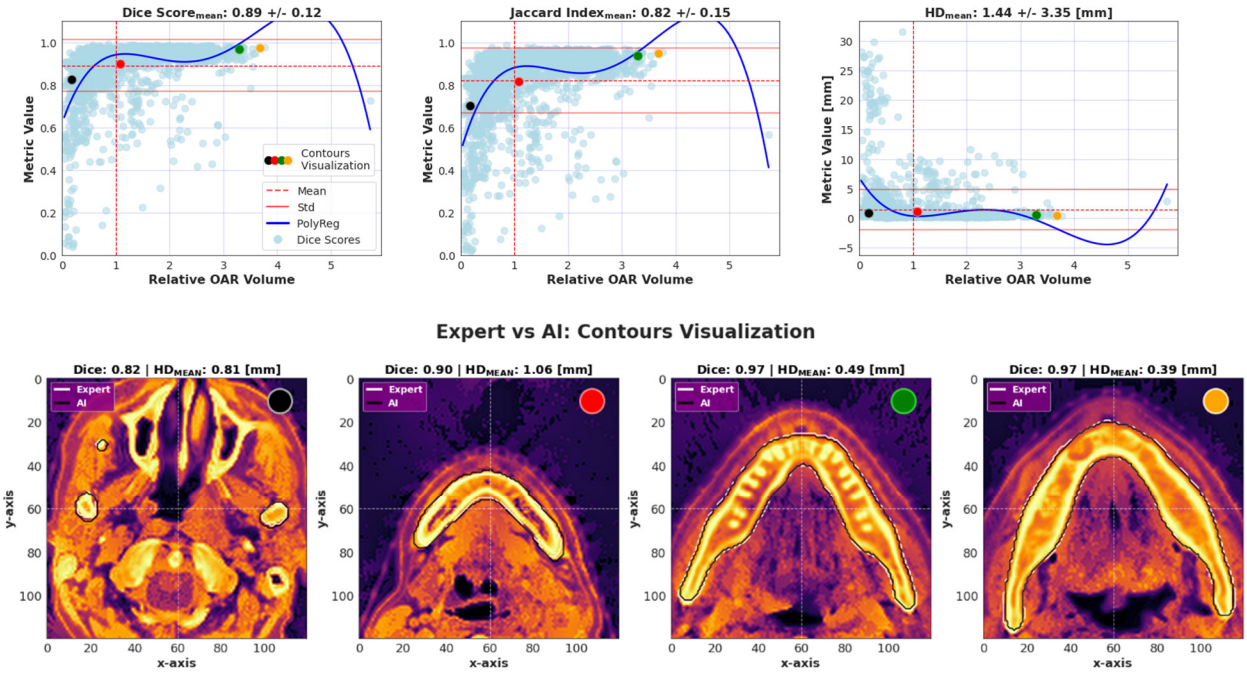
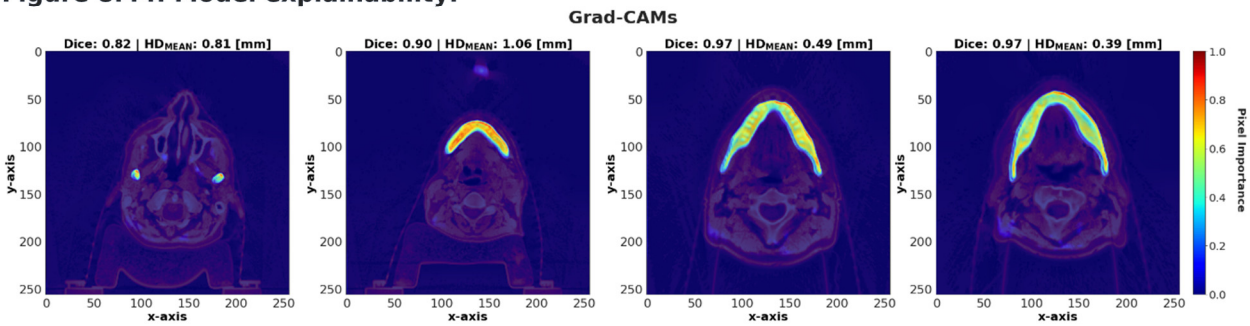



**Figure 8.44. Model explainability.**



# Neck Left

**Figure 8.45. Model training.**

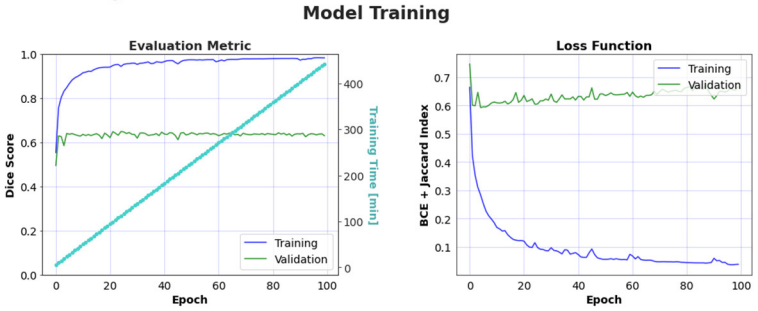



**Figure 8.46. Model evaluation on testing data.**

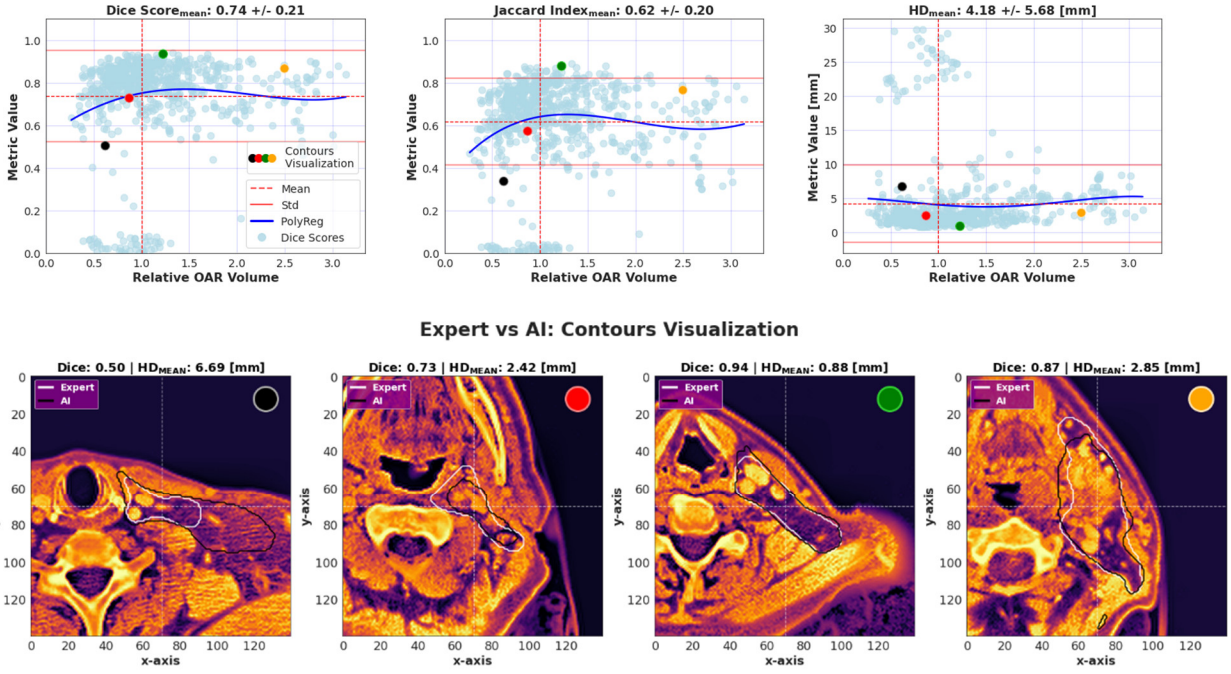
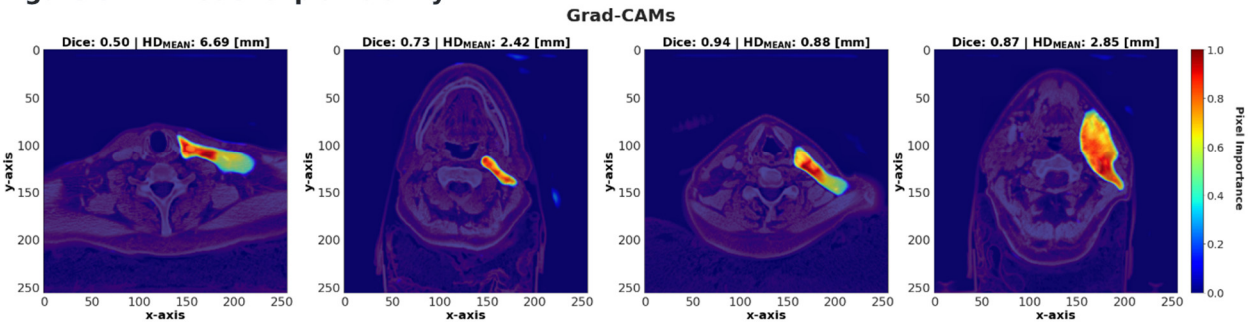



**Figure 8.47. Model explainability.**



# Neck Right

**Figure 8.48. Model training.**

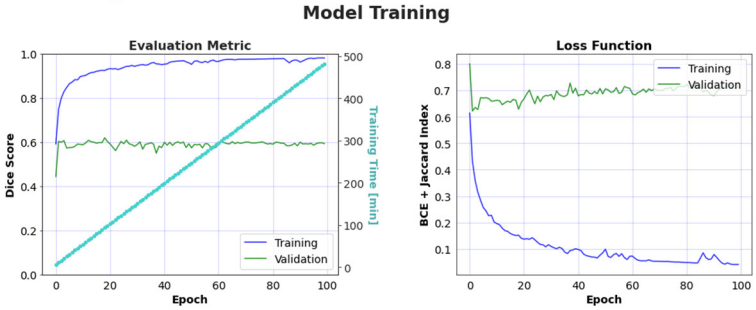
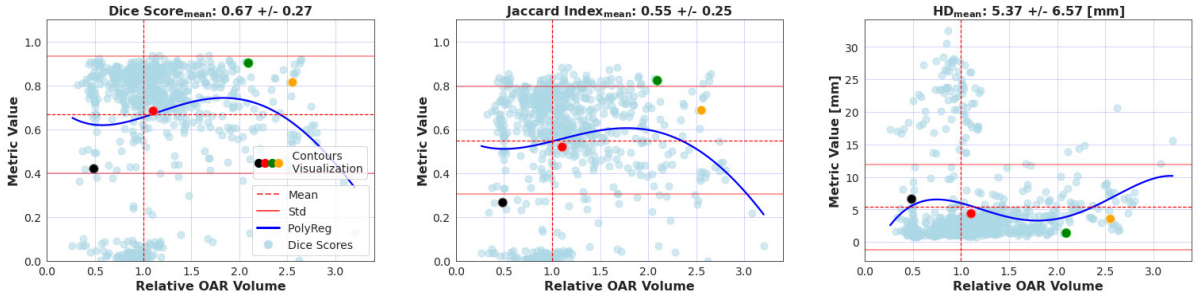



**Figure 8.49. Model evaluation on testing data.**



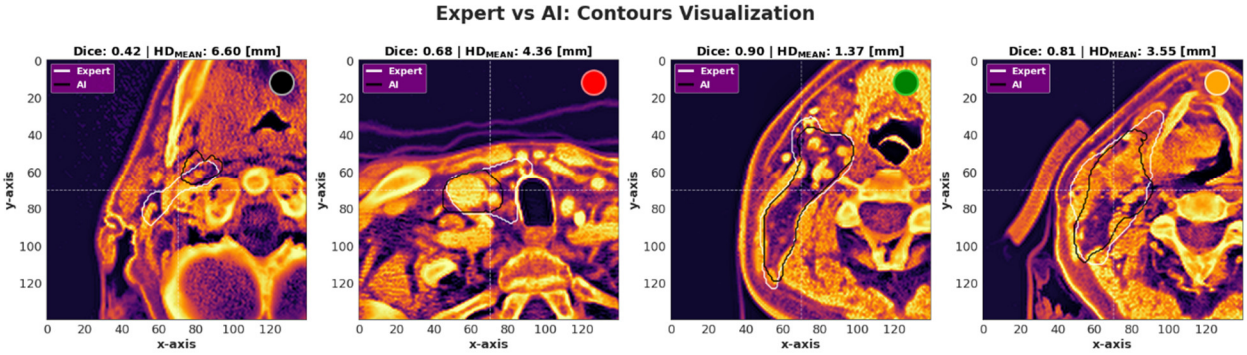
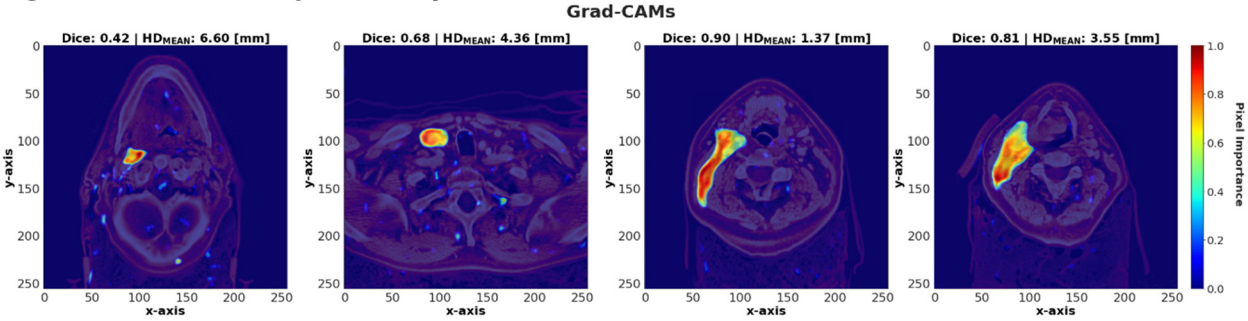




**Figure 8.50. Model explainability.**



# Oesophagus

**Figure 8.51. Model training.**

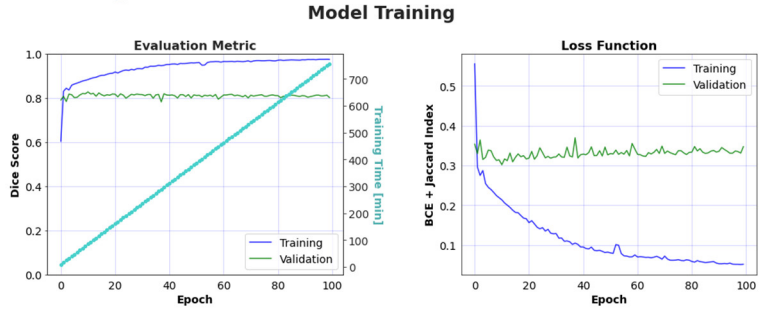
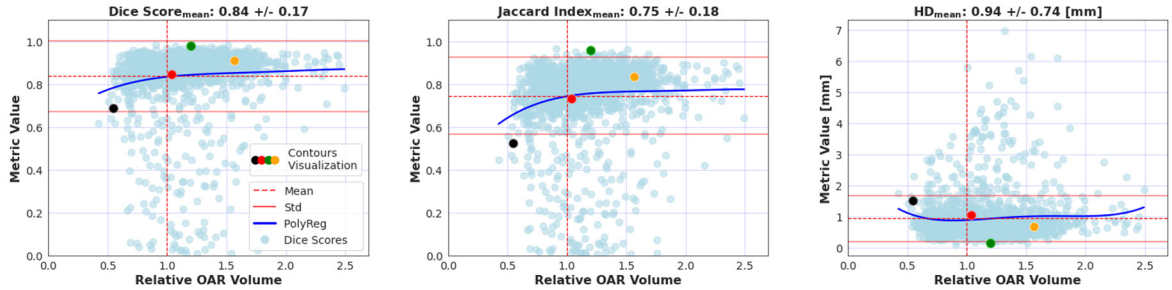



**Figure 8.52. Model evaluation on testing data.**



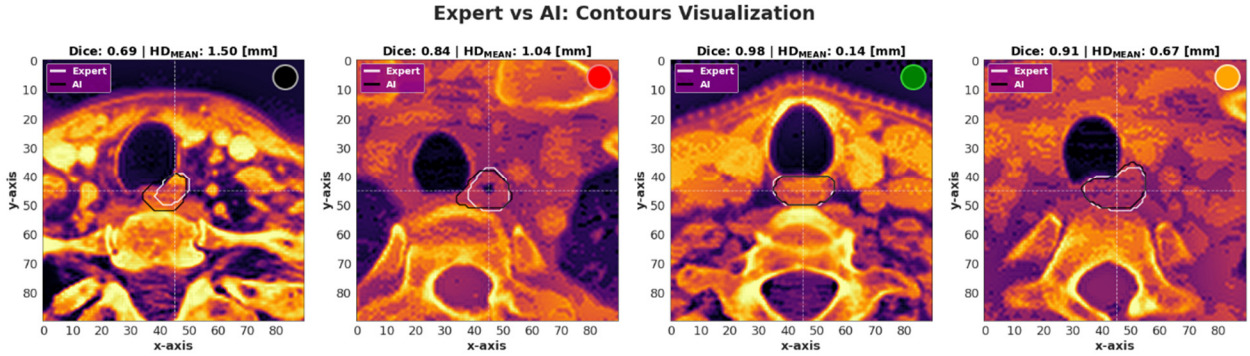
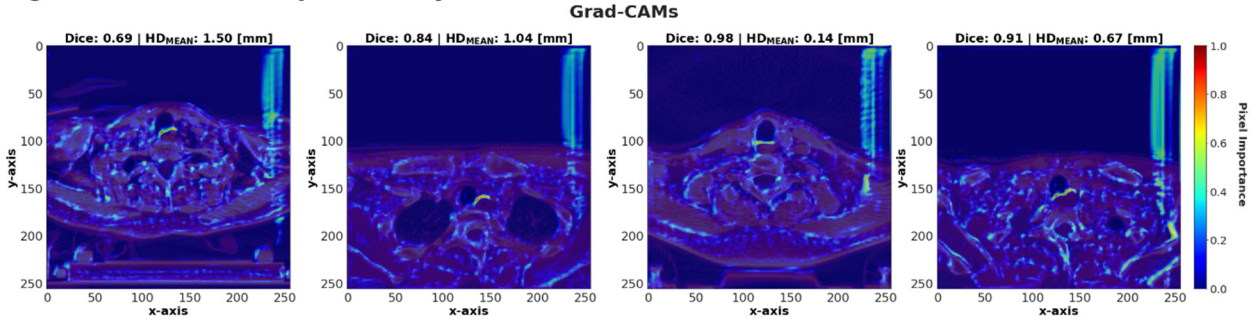




**Figure 8.53. Model explainability.**



## Optic Chiasm

**Figure 8.54. Model training.**

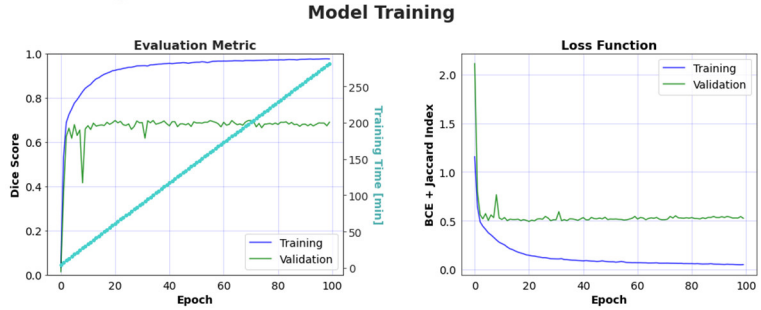



**Figure 8.55. Model evaluation on testing data.**

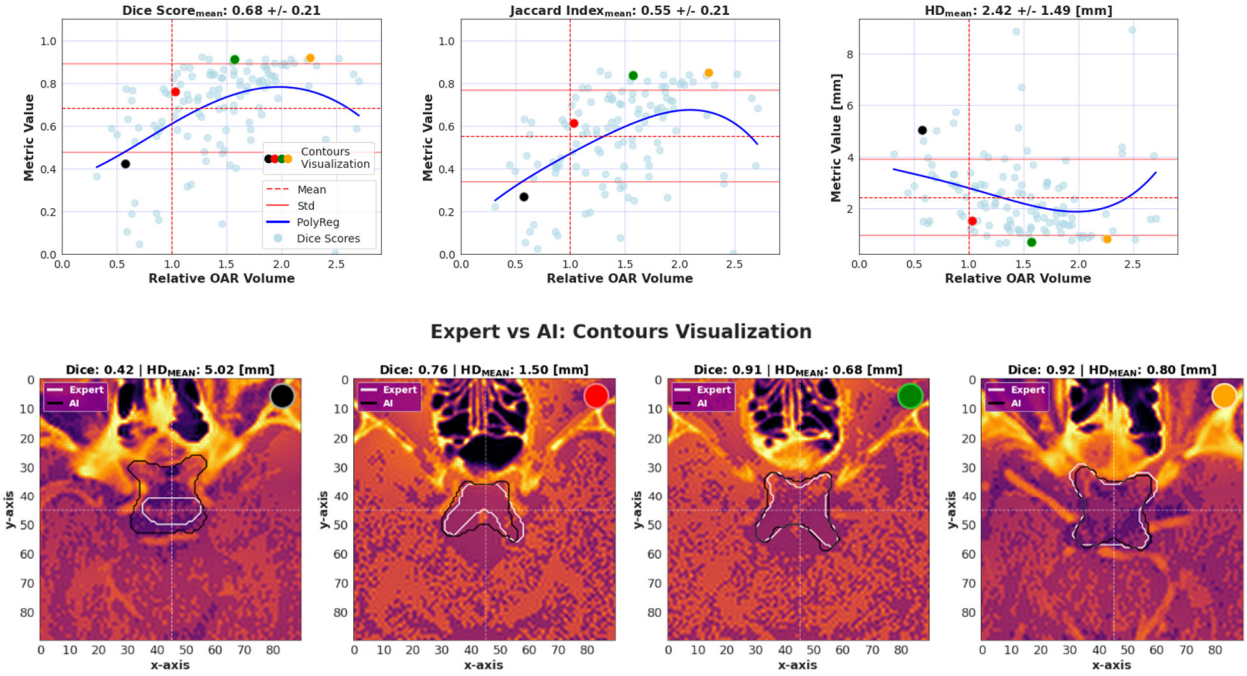
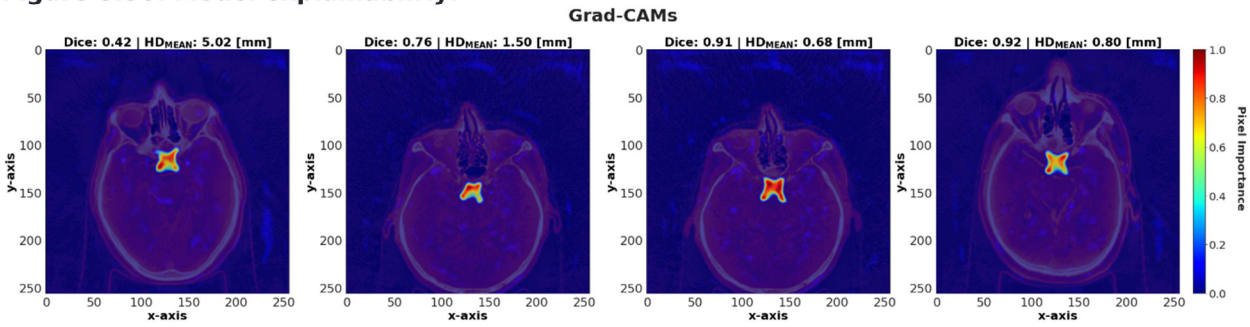



**Figure 8.56. Model explainability.**



## Optic Nerve Left

**Figure 8.57. Model training.**

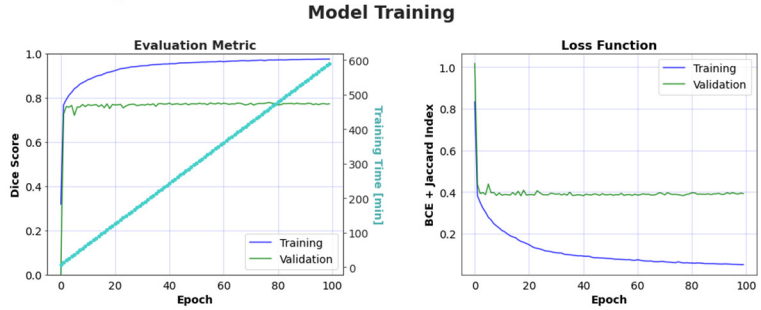
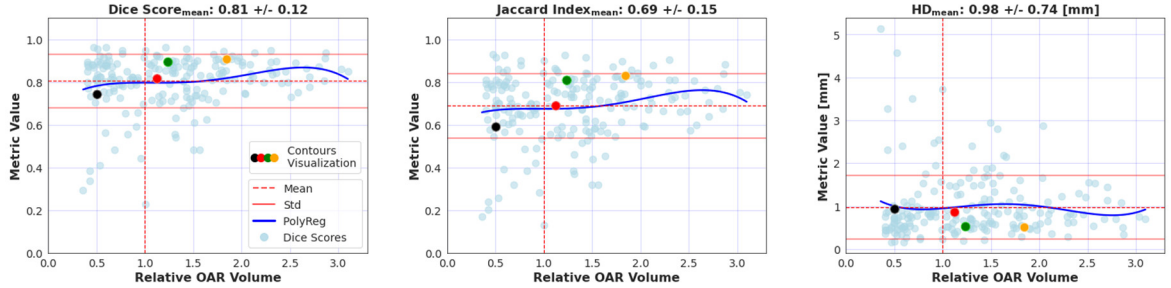



**Figure 8.58. Model evaluation on testing data.**



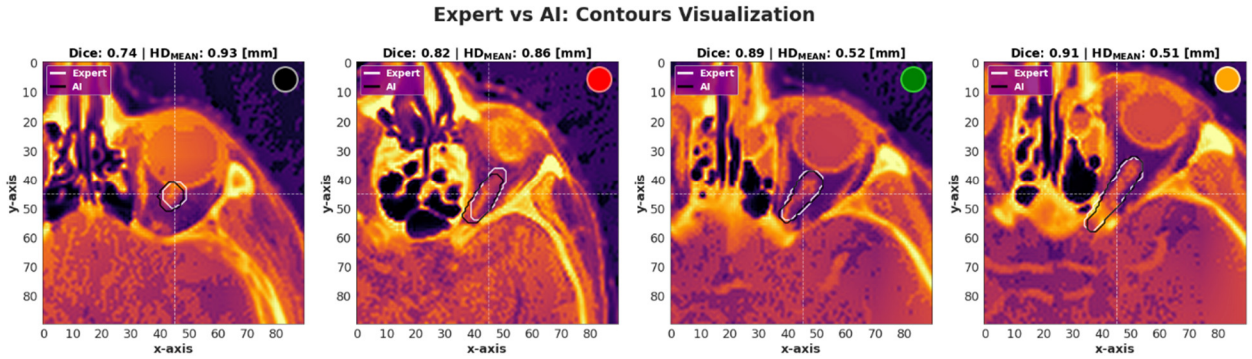
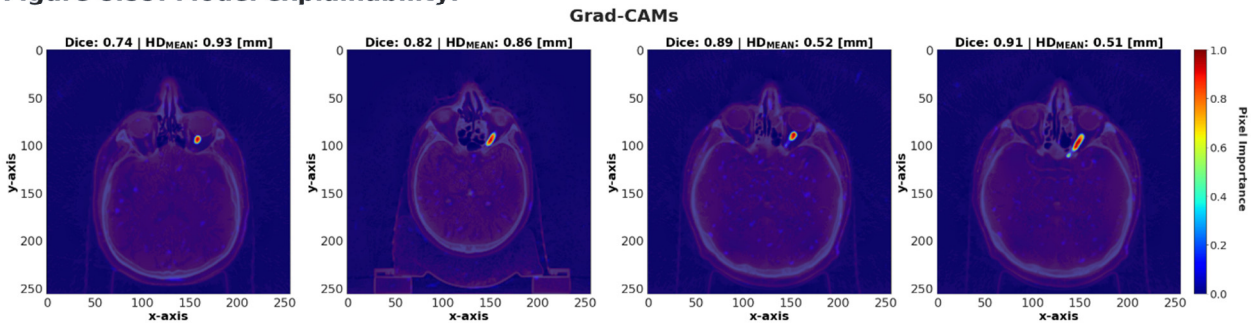




**Figure 8.59. Model explainability.**



## Optic Nerve Right

**Figure 8.60. Model training.**

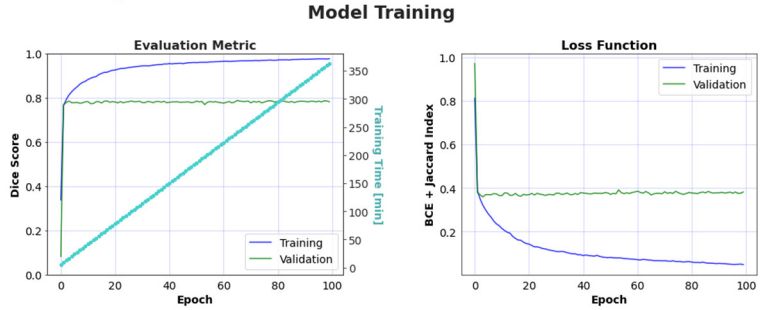
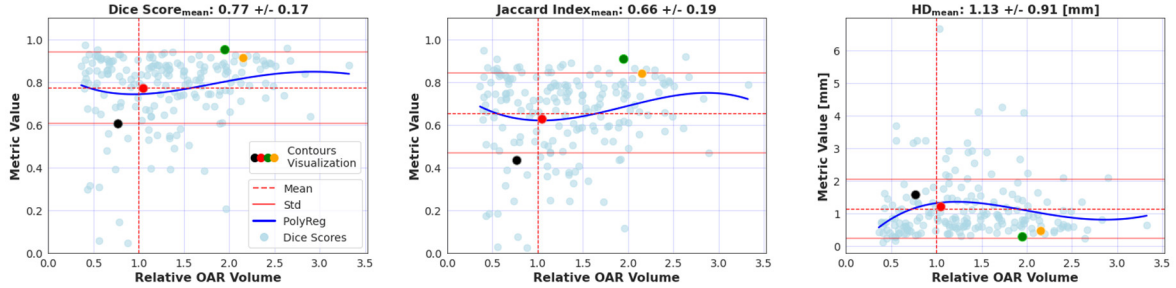



**Figure 8.61. Model evaluation on testing data.**



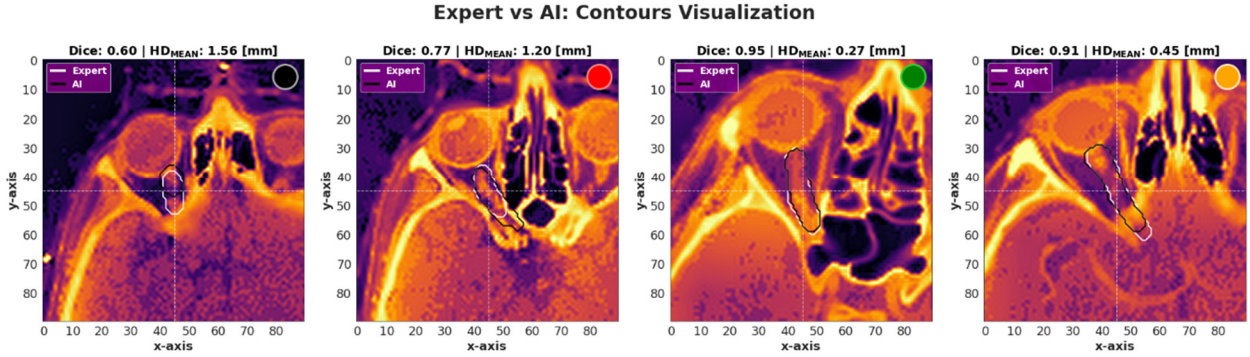
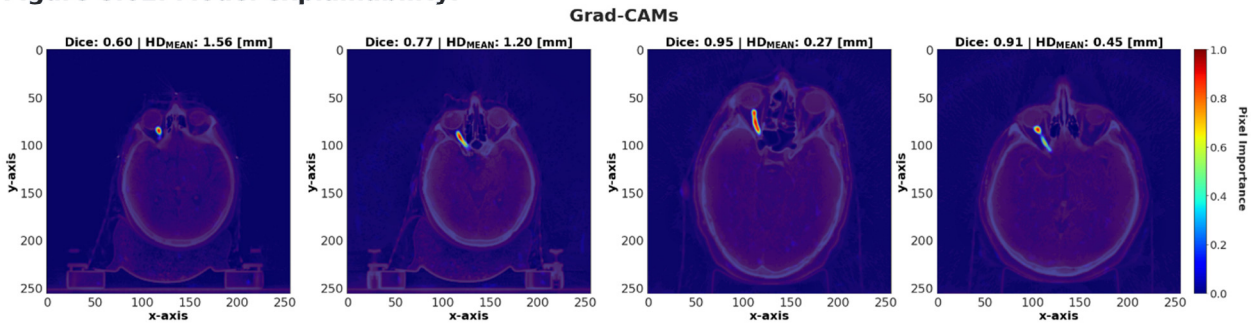




**Figure 8.62. Model explainability.**



# Oral Cavity

**Figure 8.63. Model training.**

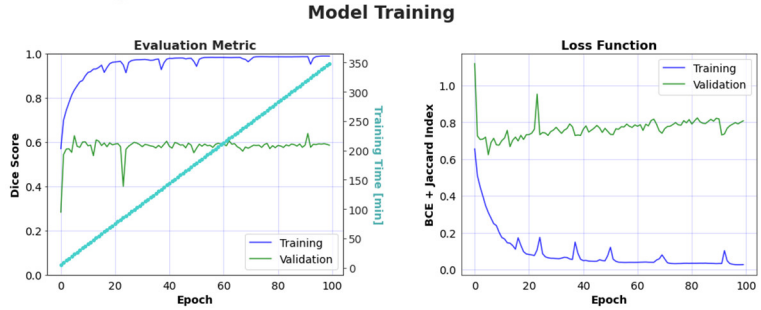
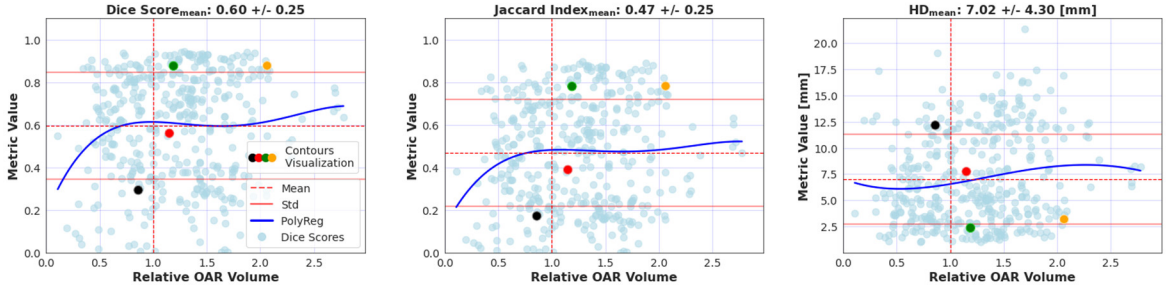



**Figure 8.64. Model evaluation on testing data.**



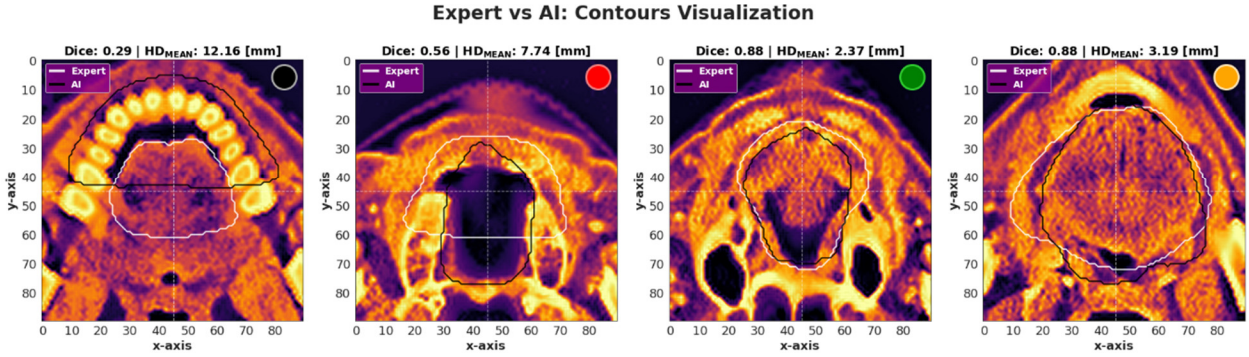
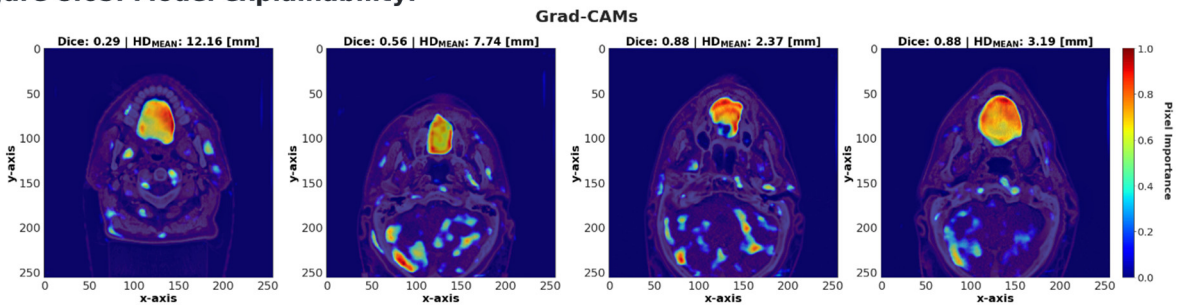




**Figure 8.65. Model explainability.**



# Parotid Left

**Figure 8.66. Model training.**

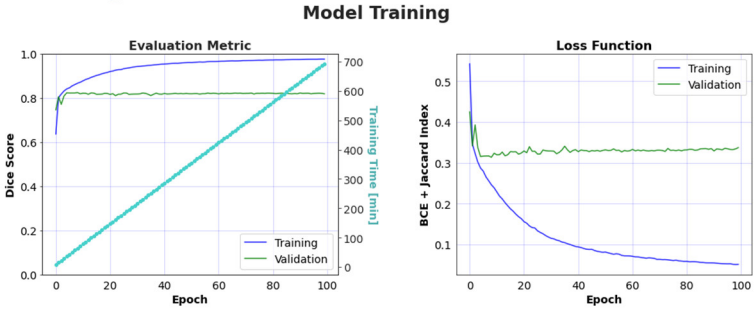



**Figure 8.67. Model evaluation on testing data.**

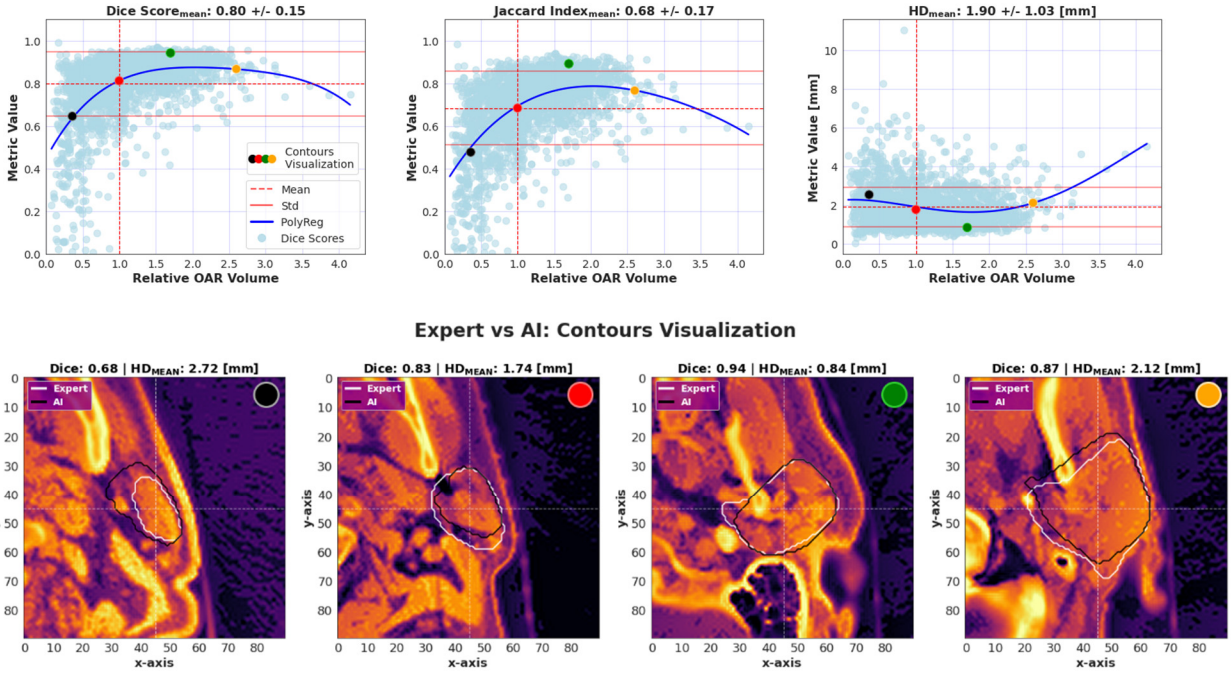
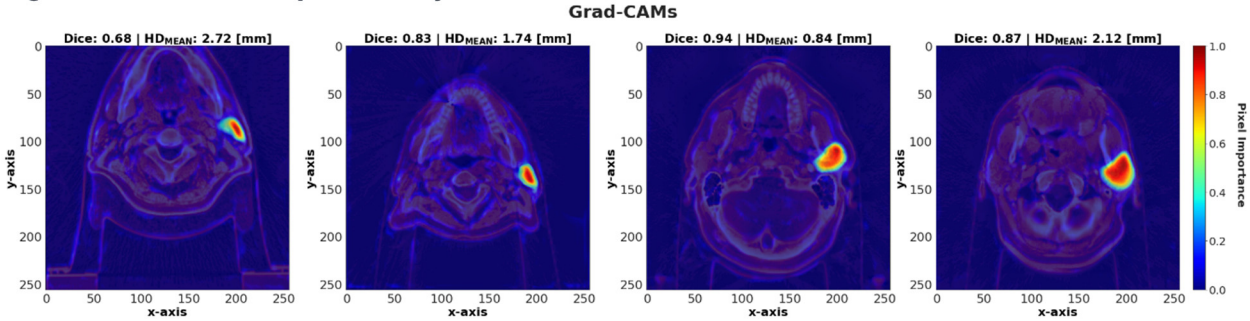



**Figure 8.68. Model explainability.**



## Parotid Right

**Figure 8.69. Model training.**

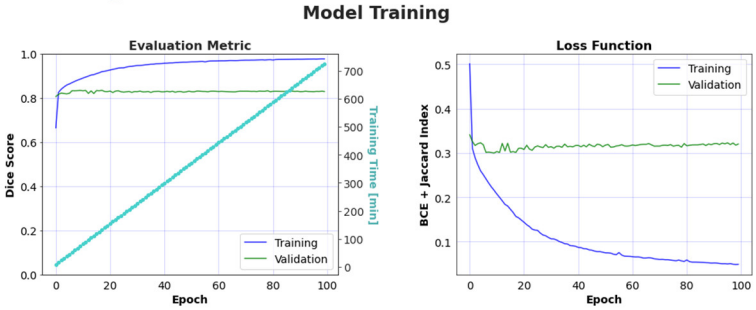
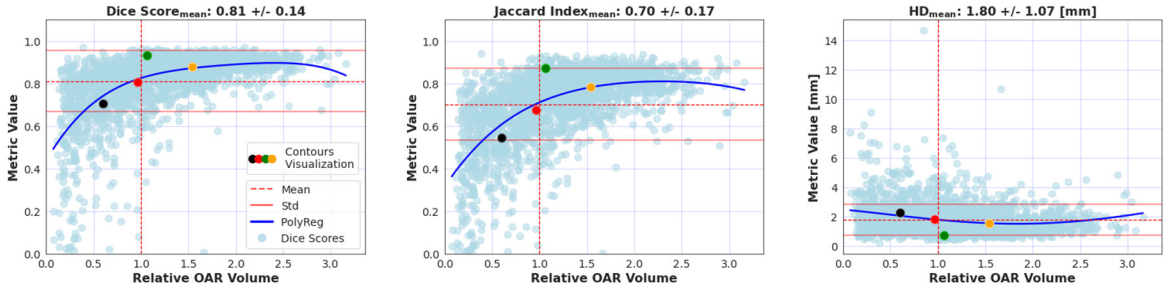



**Figure 8.70. Model evaluation on testing data.**



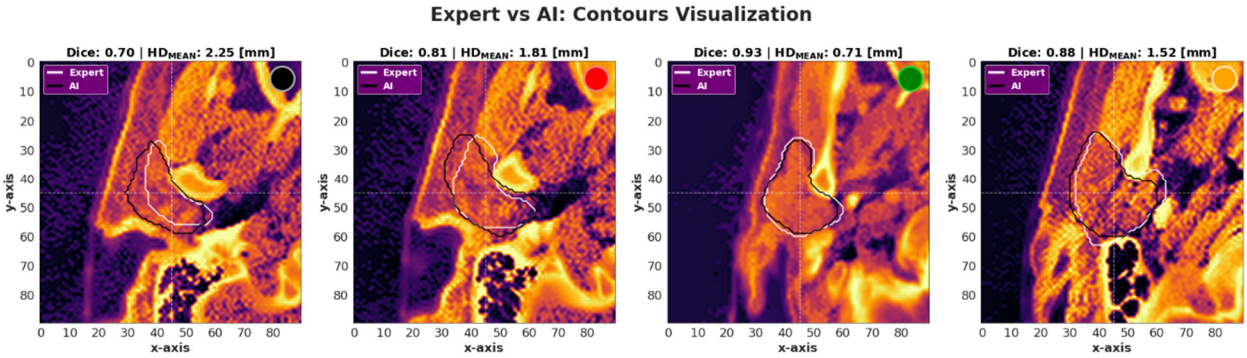
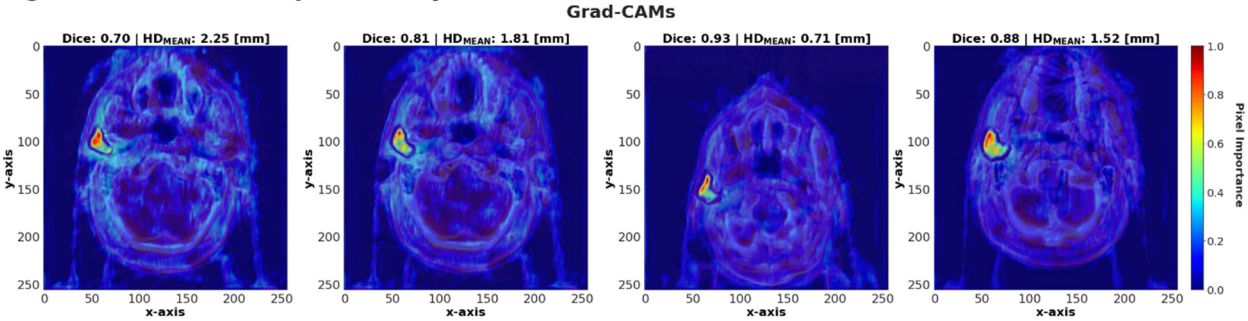




**Figure 8.71. Model explainability.**



## Submandibular Left

**Figure 8.72. Model training.**

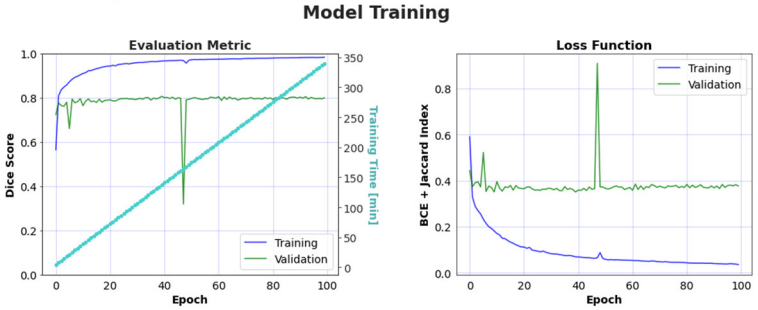
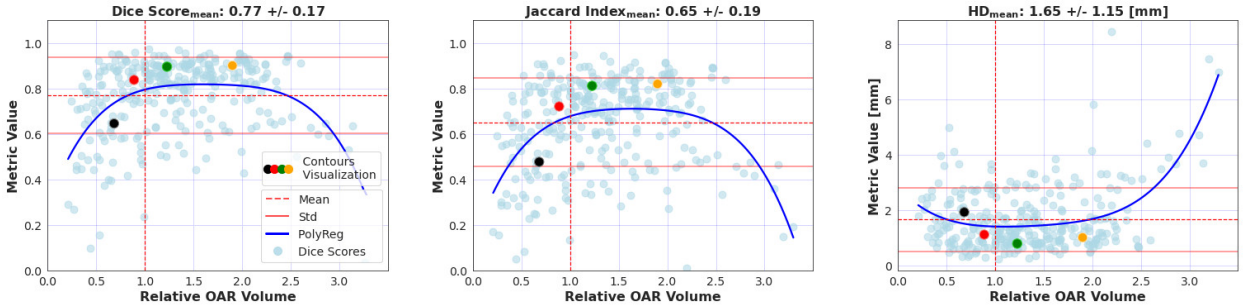



**Figure 8.73. Model evaluation on testing data.**



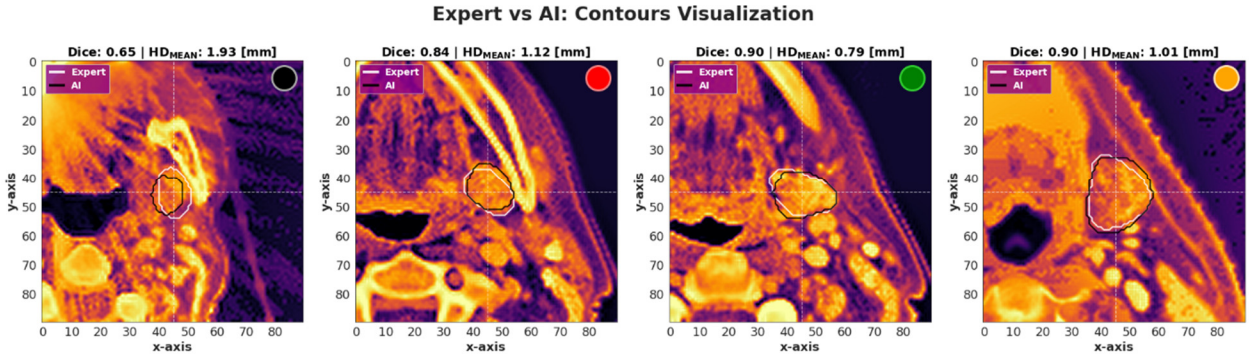
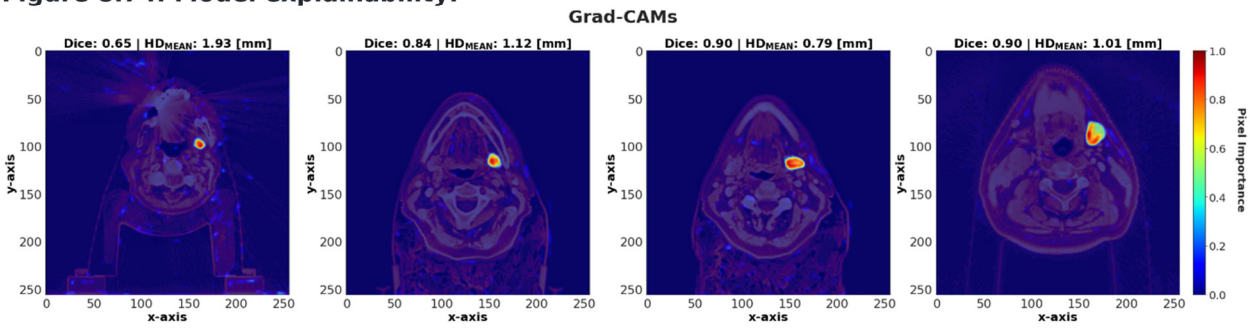




**Figure 8.74. Model explainability.**



## Submandibular Right

**Figure 8.75. Model training.**

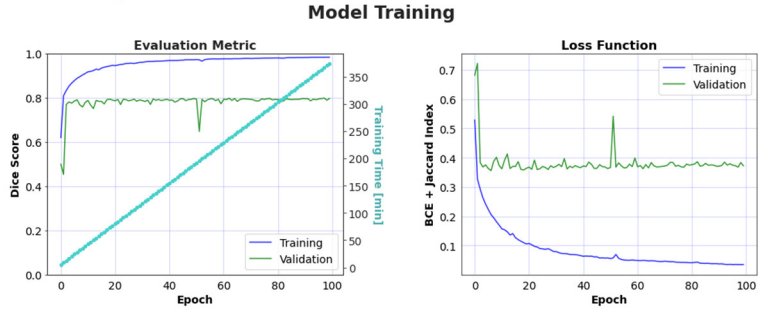
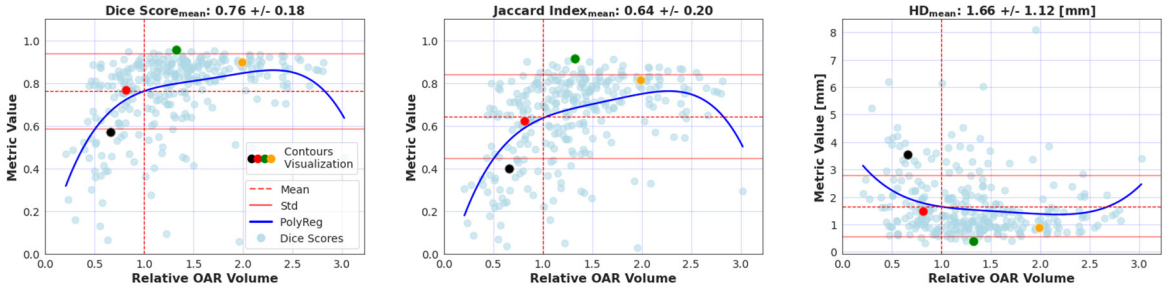



**Figure 8.76. Model evaluation on testing data.**



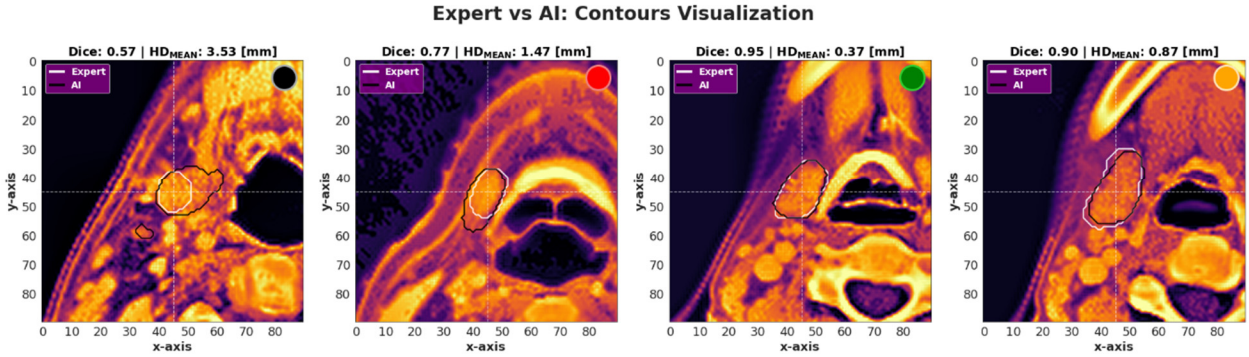
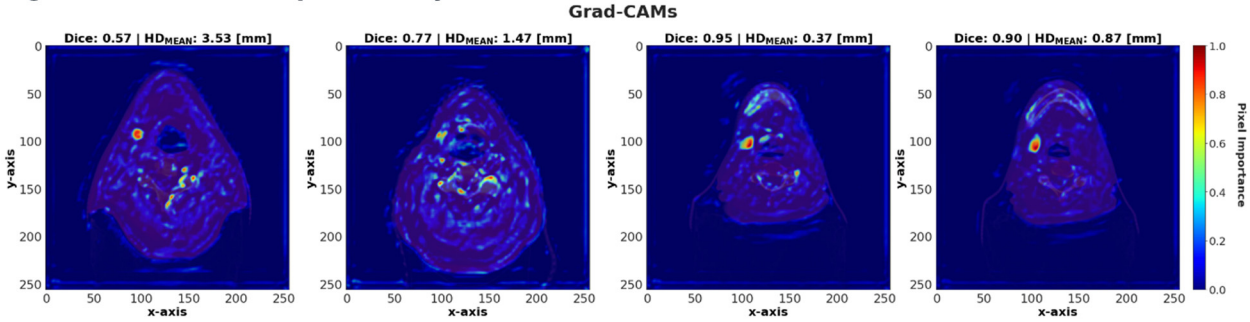




**Figure 8.77. Model explainability.**



## 8.4 Development & Performance of Logistic Regression Model

This section contains Part I of supplementary material for the research work described in Chapter 6 *Machine learning for dose-volume histogram based clinical decision-making support system in radiation therapy plans for brain tumors*

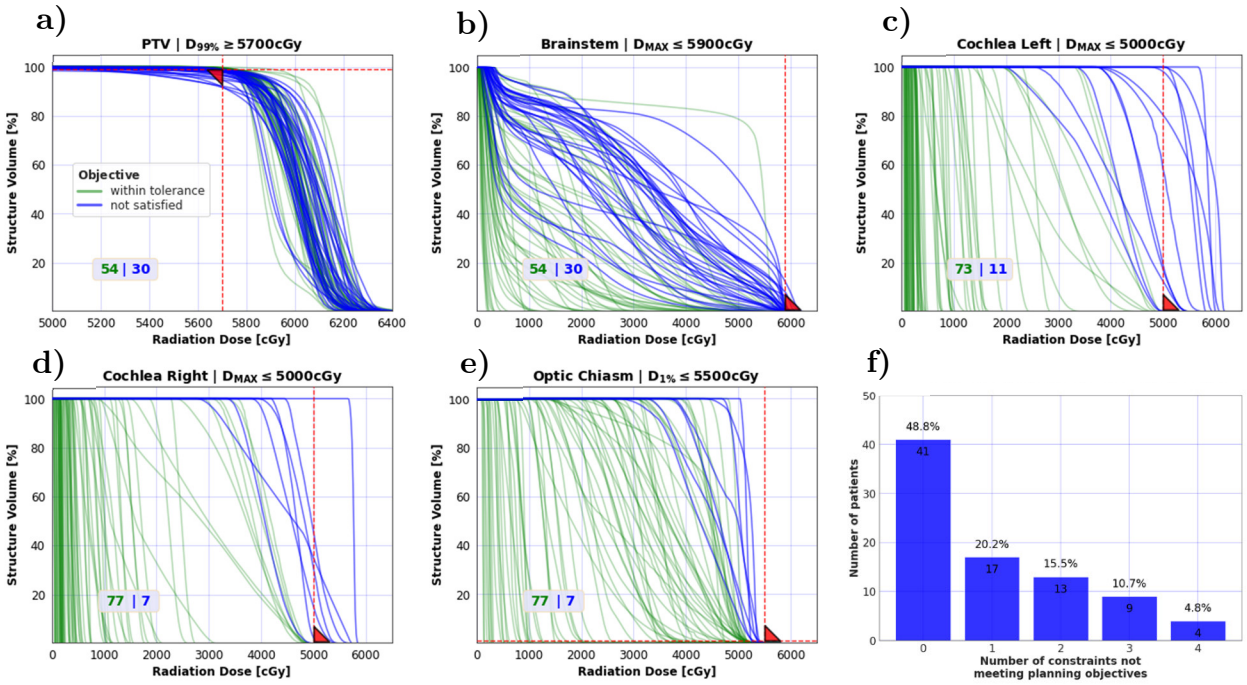



**Figure 8.78. a)-e)** Dose-volume histograms for a target and OARs structures together with the treatment planning constraints used as features in the machine learning model training. These charts include all patients initially considered in this study. Numbers separated by the vertical bar in the lower left part of each graph shows the number of patients who meet (green color corresponding to the green DVH curves) and do not meet the particular planning objective (blue color corresponding to the blue DVH curves); **f)** Number of patients for whom a specified number of treatment planning objectives were not met. The absolute number of patients indicated as bar annotations do not add up to the total number of patients as a particular patient may have more than one structure not meeting planning criteria.

**Table 8. 5.** Hyperparameters for tuning machine learning models.

| Model | Hyperparameters* |
|---|---|
| **Support Vector Machine** | **Kernels** \| Linear, RBF, Sigmoid<br>Kernels map the features extracted from the original data samples into a high-dimensional feature space, where those features can be well-separable using a linear function (Linear kernel), Gaussian radial basis function (RBF kernel), or Sigmoid function (Sigmoid kernel). |
| | **C (RBF kernel only)** \| logarithmic grid from $10^{-2}$ to $10^{2}$<br>Determines the tolerance of the classifier to classification errors. |
| | **Gamma (RBF kernel only)** \|logarithmic grid from $10^{-2}$ to $10^{2}$<br>Defines the inverse of the standard deviation of the RBF kernel. |
| **Elastic Net** | **L1_ratio** \| 0.10, 0.50, 0.70, 0.90, 0.95, 1.00<br>Regularization term. |
| | **Alpha** \| 0.1, 0.5, 1<br>$\alpha$ is the mixing parameter between ridge ($\alpha = 0$) and lasso ($\alpha = 1$) regularization. |
| **Logistic Regression** | **Penalty** \| L1, L2 (default values)<br>L1 – regularization using Lasso (Least Absolute Shrinkage and Selection Operator) regression<br>L2 – regularization using Rigid regression |
| | **C** \| logarithmic grid from $10^{-2}$ to $10^{2}$<br>The inverse of regularization strength. |
| | **Solver** \| liblinear (not hyperparameter)<br>Liblinear solver uses the coordinate descent algorithm that is based on minimizing a multivariate function by solving univariate optimization problems in a loop. |
| **Random Forest Classifier** | **Estimators** \| 5, 10, 15, 20<br>The number of decision trees. |
| | **Max. depth** \| 2, 3, 4, 5, 6<br>The number of levels in each decision tree. |

**Table 8. 6.** Interpretation of model evaluation metrics.

| Metric | Interpretation |
|---|---|
| **TP** – true positives | The number of correctly classified plans for which PTV objective was not met. |
| **TN** – true negatives | The number of correctly classified plans for which PTV objective was met. |
| **FP** – false positives | The number of incorrectly classified plans for which PTV objective was not met. |
| **FN** – false negatives | The number of incorrectly classified plans for which PTV objective was met. |
| **Accuracy**<br>= (TP+TN)/(TP+TN+FP+FN) | The overall accuracy of the model associated with both classes. In the other words, this metric reflects the ratio of all correct predictions over all the predictions. |
| **Precision**<br>= TP/(TP+FP) → class 1<br>= TN/(TN+FN) → class 0 | If the model predicts class 1 (class 0) then precision shows what is the probability that it is a correct prediction of class 1 (class 0). In the other words, precision reflects the accuracy associated with a single class. |
| **Recall**<br>= TP/(TP+FN) → class 1 (sensitivity)<br>= TN/(TN+FP) → class 0 (specificity) | Sensitivity – if the correct prediction is class 1 then sensitivity shows what is the probability that the model would predict class 1.<br><br>Specificity - if the correct prediction is class 0 then sensitivity shows what is the probability that the model would predict class 0. |
| **F1**<br>$= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$ | The harmonic mean of precision and recall where both measures are assumed to be equally important (weighted). |

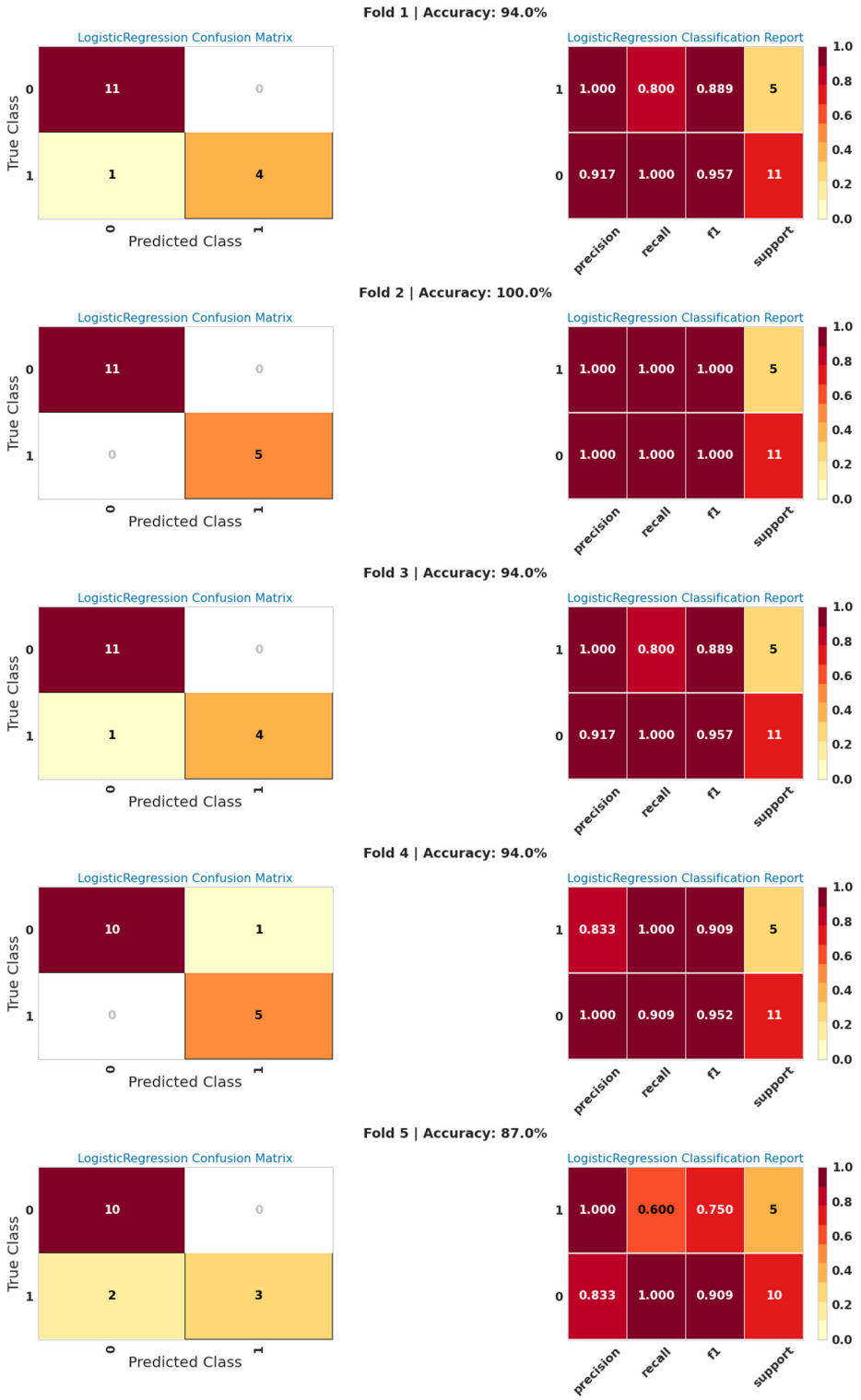



**Figure 8. 79.** Confusion matrices and Logistic Regression model evaluation metrics for five cross-validation folds.

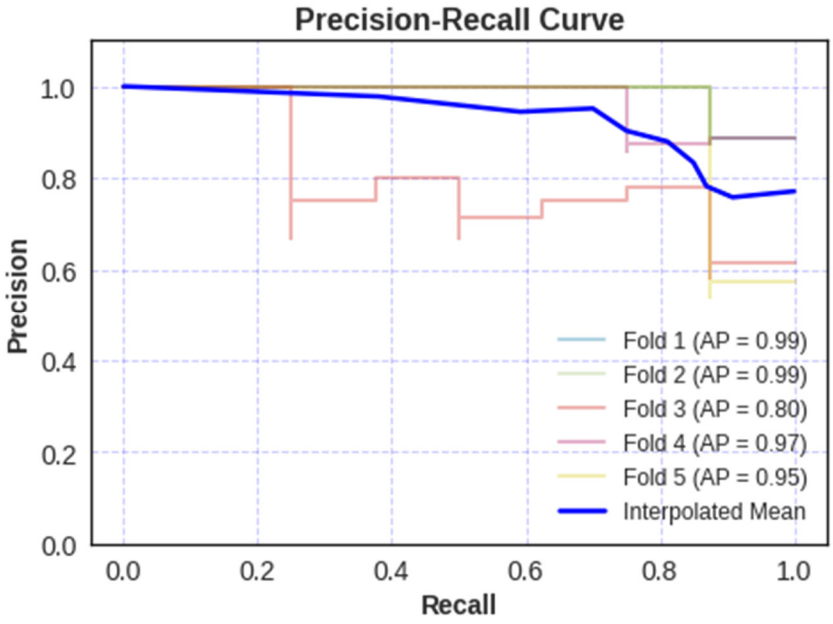



**Figure 8.80.** Precision Recall Curves for five cross-validation folds of Logistic Regression Model. The AP stands for the average precision and is approximately equal to the area under the curves. The AP for interpolated mean curve is 0.94.

## 8.5 Explainability Analysis for Logistic Regression Model

This section contains Part II of supplementary material for the research work described in Chapter 6 *Machine learning for dose-volume histogram based clinical decision-making support system in radiation therapy plans for brain tumors.*

To simplify the visualization of numerous charts the figure description was not provided in this Appendix. However, it can be found in Chapter 6.

For the proper data visualization, the supplementary material starts from the next page.

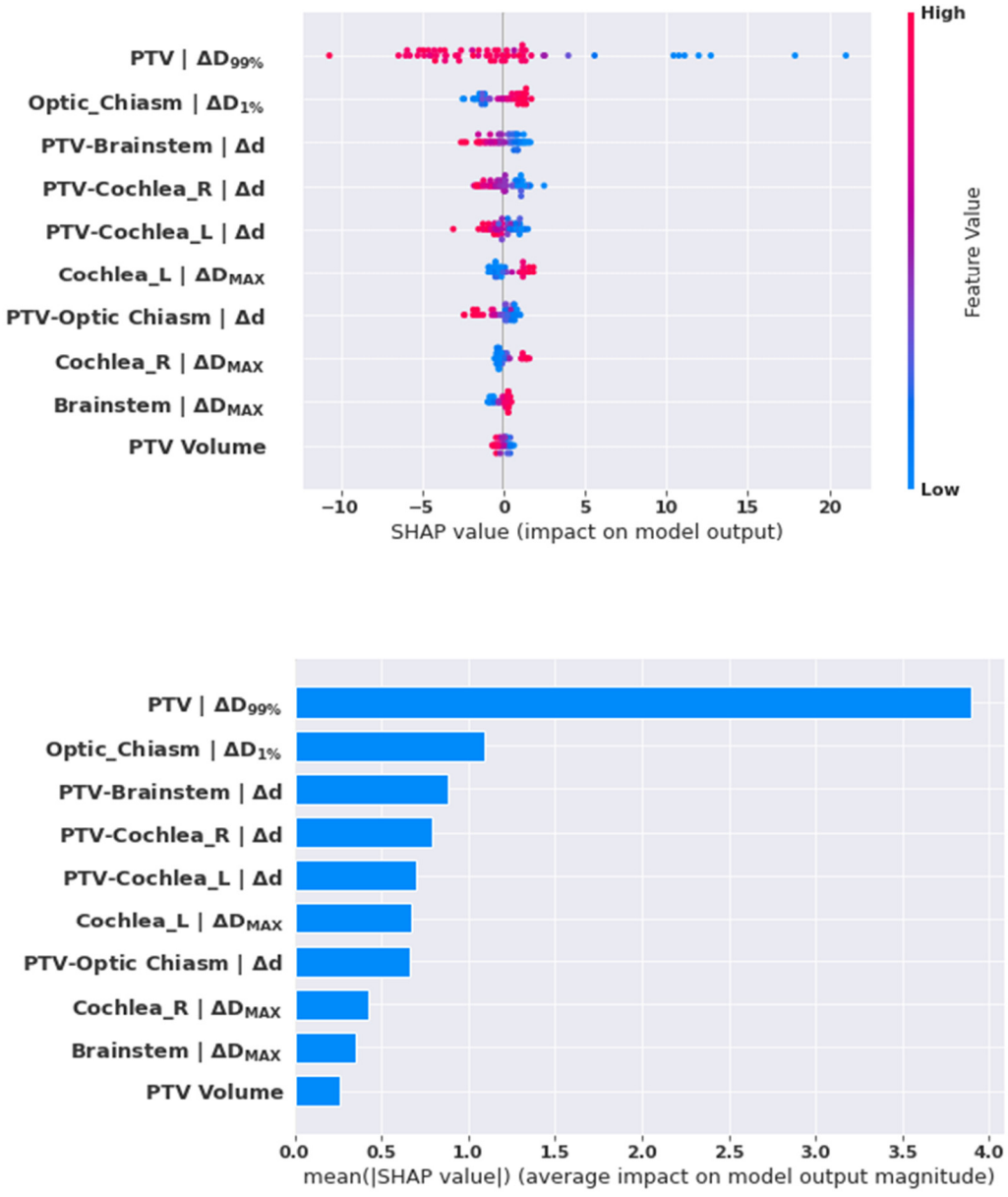



**Figure 8.81.** Cross-Validation Fold 1. Mean Feature Importance.

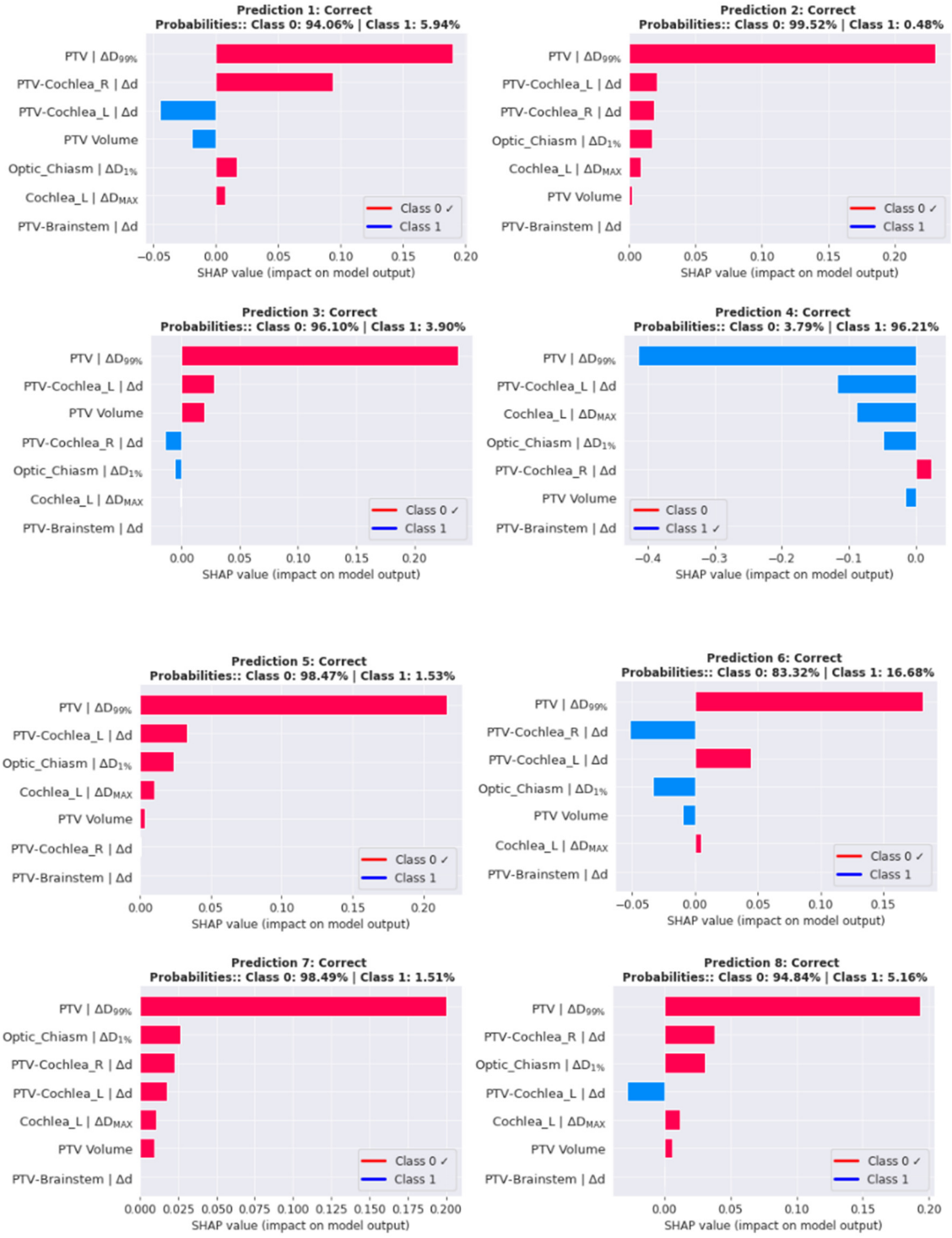



**Figure 8.82.** Cross-Validation Fold 1. Feature Importance for Single Predictions (1-8).

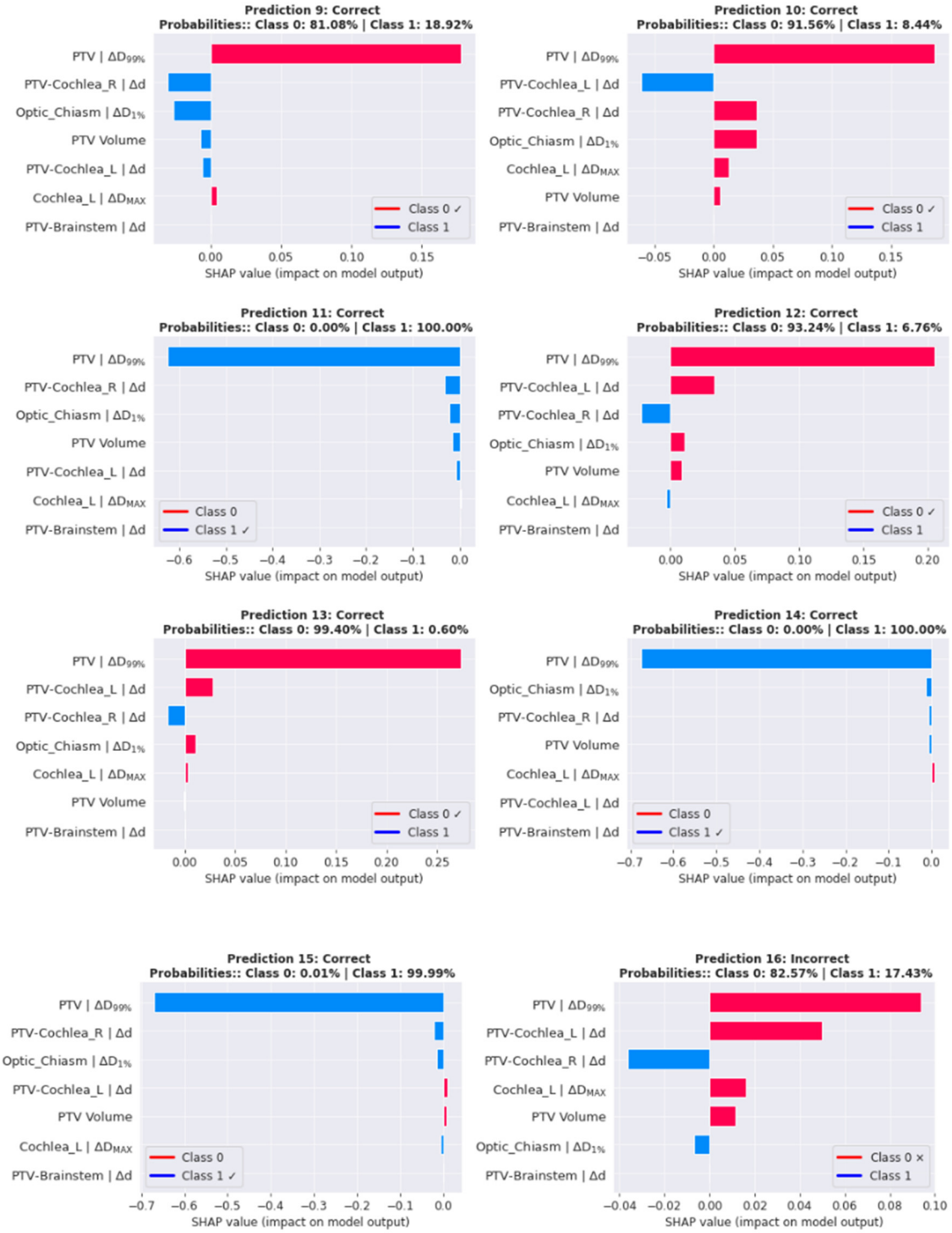
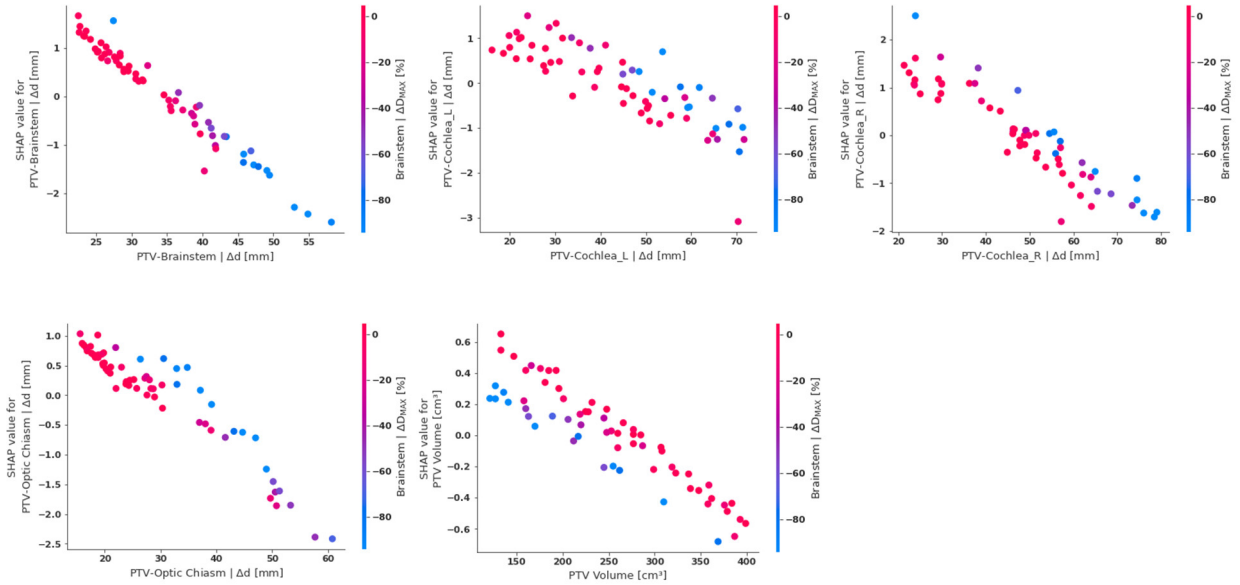



**Figure 8.83.** Cross-Validation Fold 1. Feature Importance for Single Predictions (9-16).

**Geometric Features**




**Dosimetric Features**

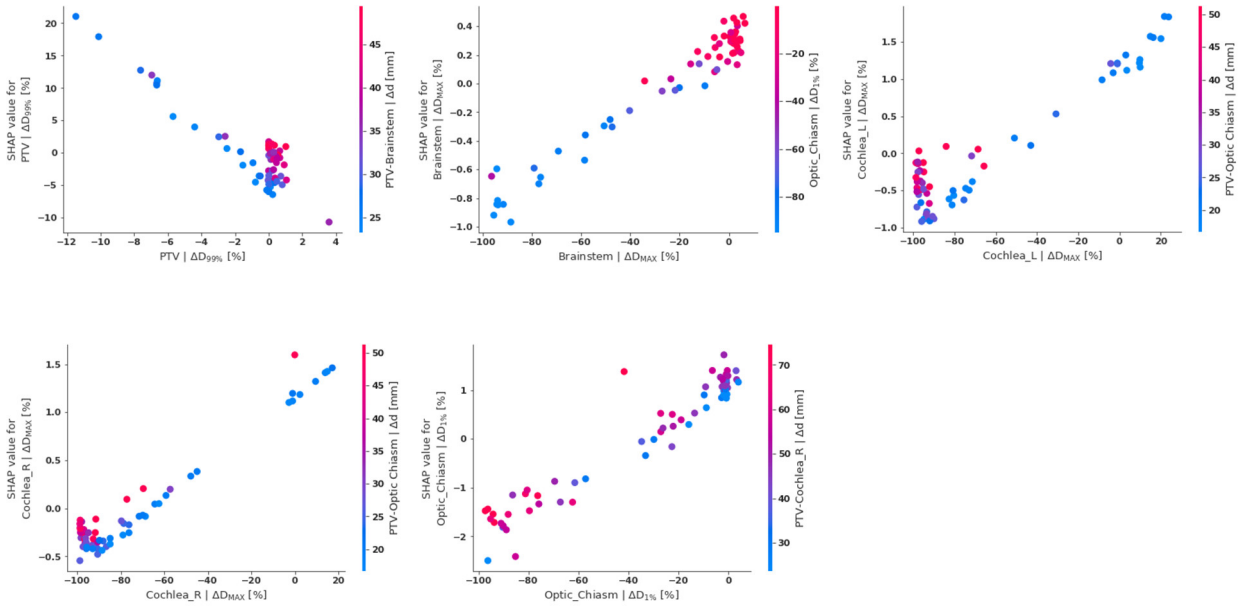



**Figure 8.84.** Cross-Validation Fold 1. Partial Dependency Charts.

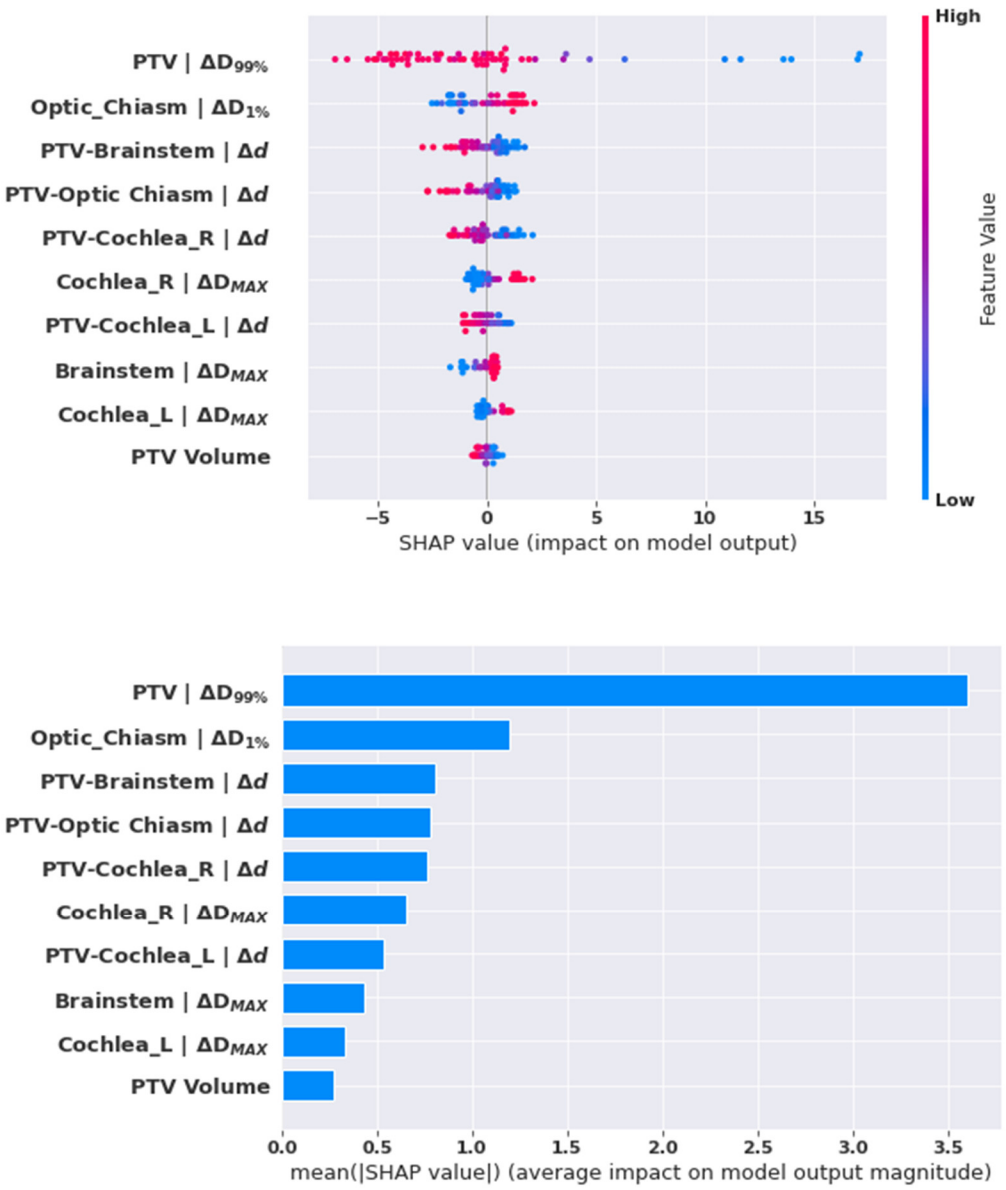



**Figure 8.85.** Cross-Validation Fold 2. Mean Feature Importance.

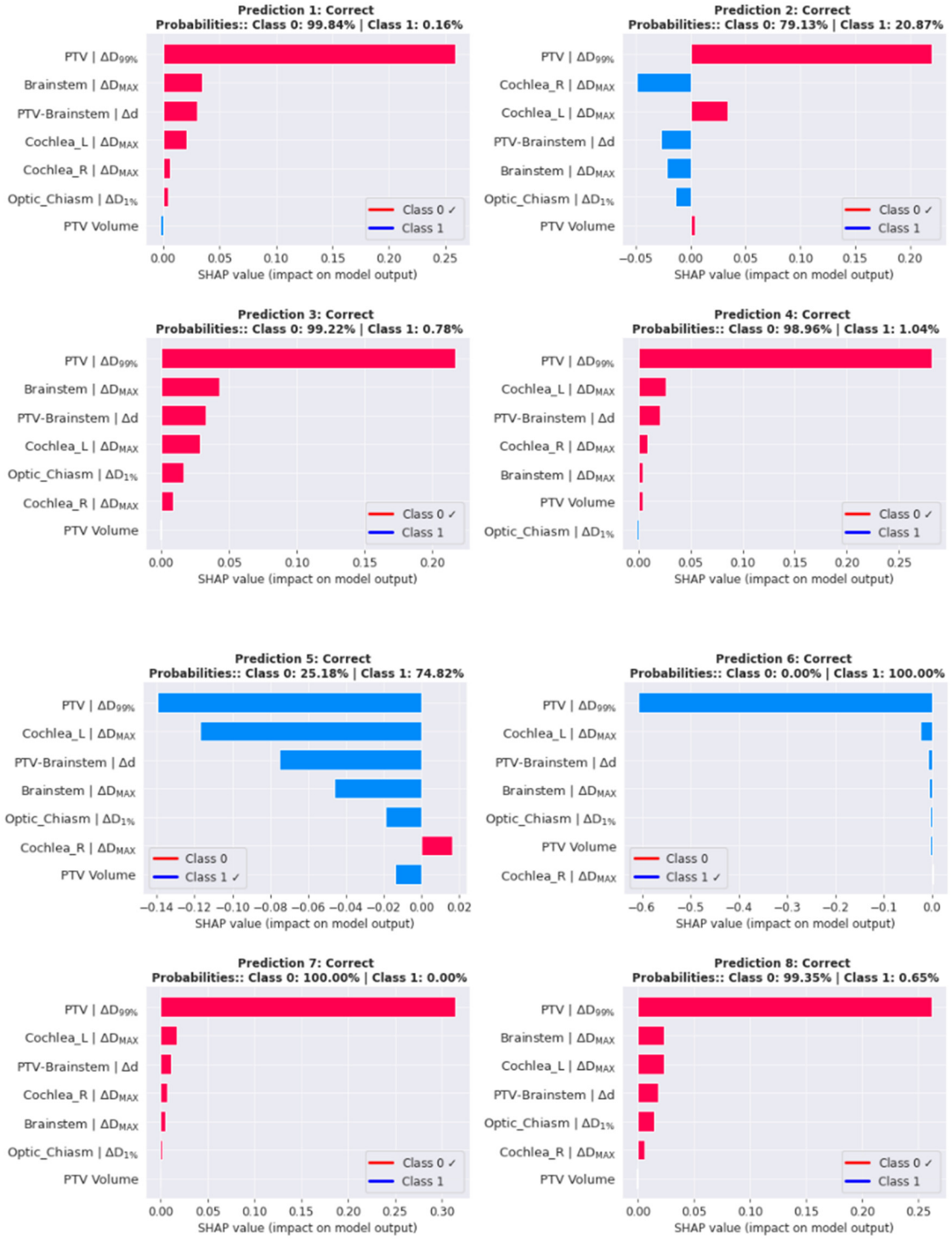



**Figure 8.86.** Cross-Validation Fold 2. Feature Importance for Single Predictions (1-8).

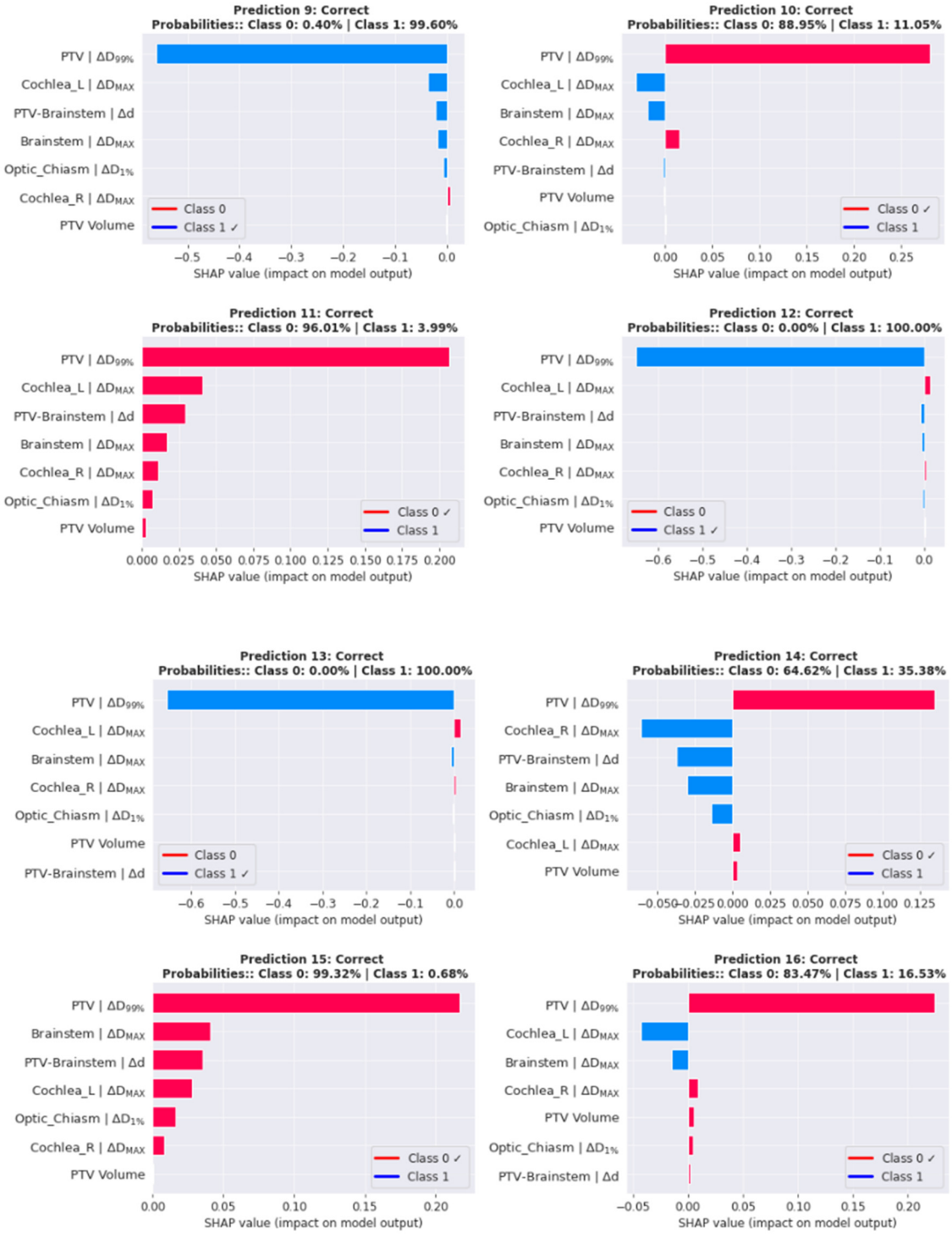
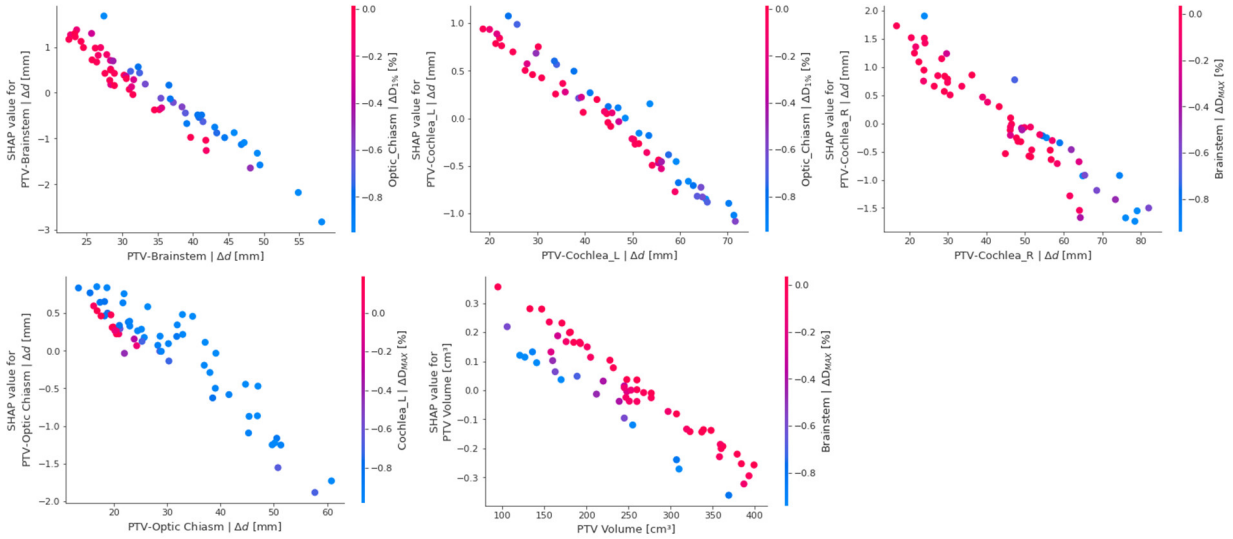



**Figure 8.87.** Cross-Validation Fold 2. Feature Importance for Single Predictions (9-16).

**Geometric Features**




**Dosimetric Features**

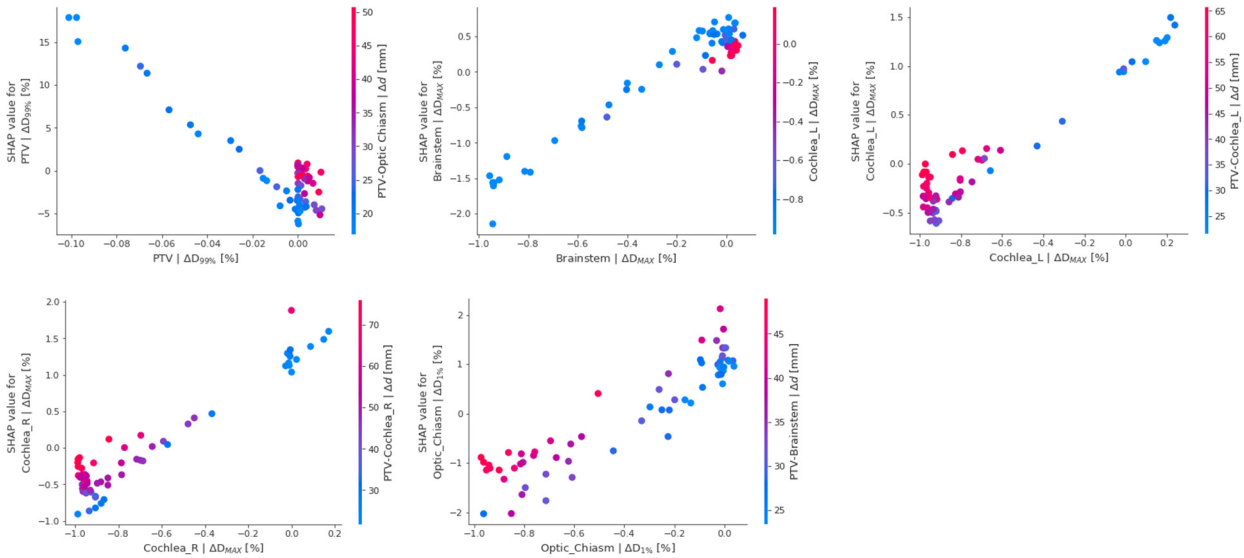



**Figure 8.88.** Cross-Validation Fold 2. Partial Dependency Charts.

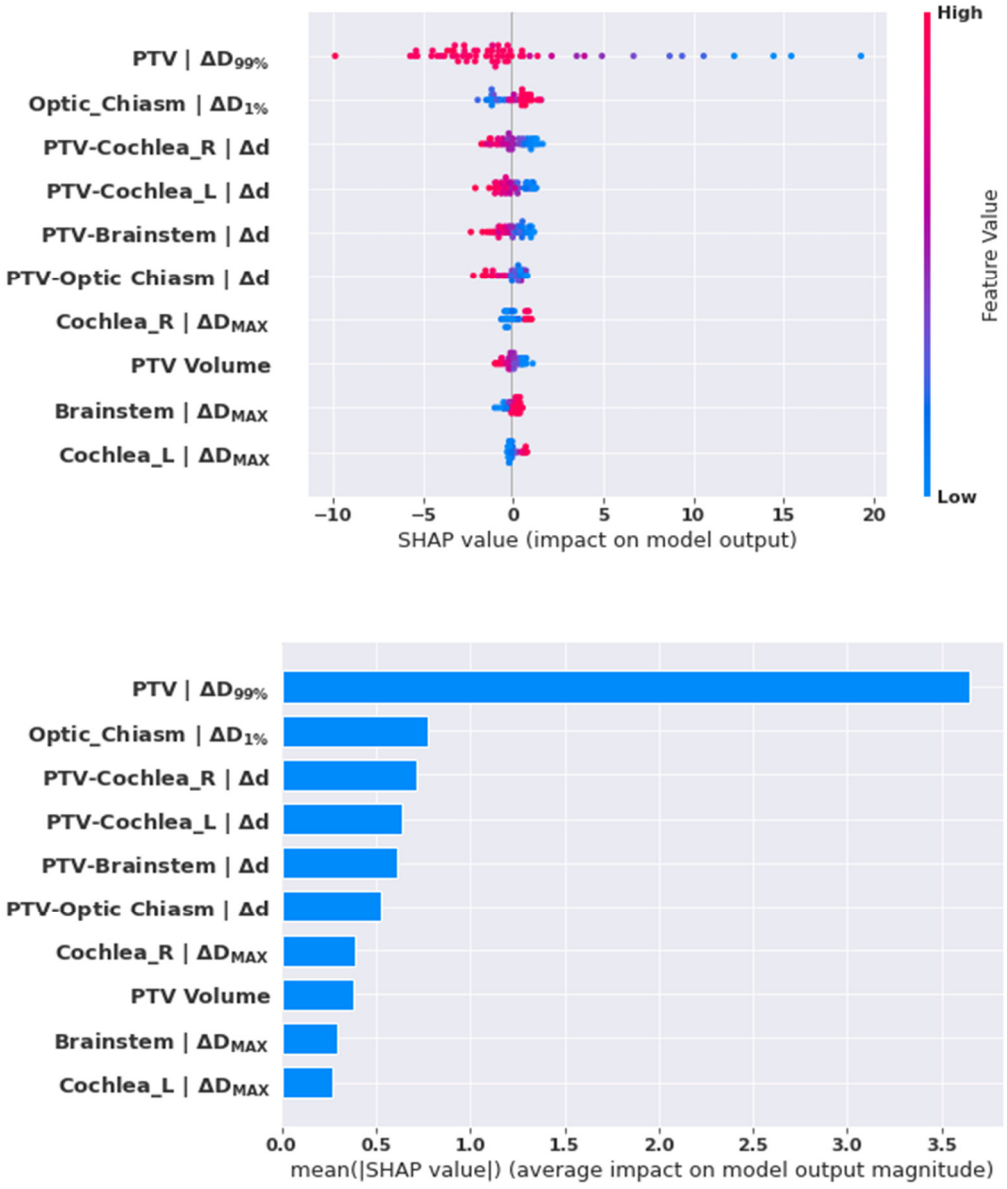



**Figure 8.89.** Cross-Validation Fold 3. Mean Feature Importance.

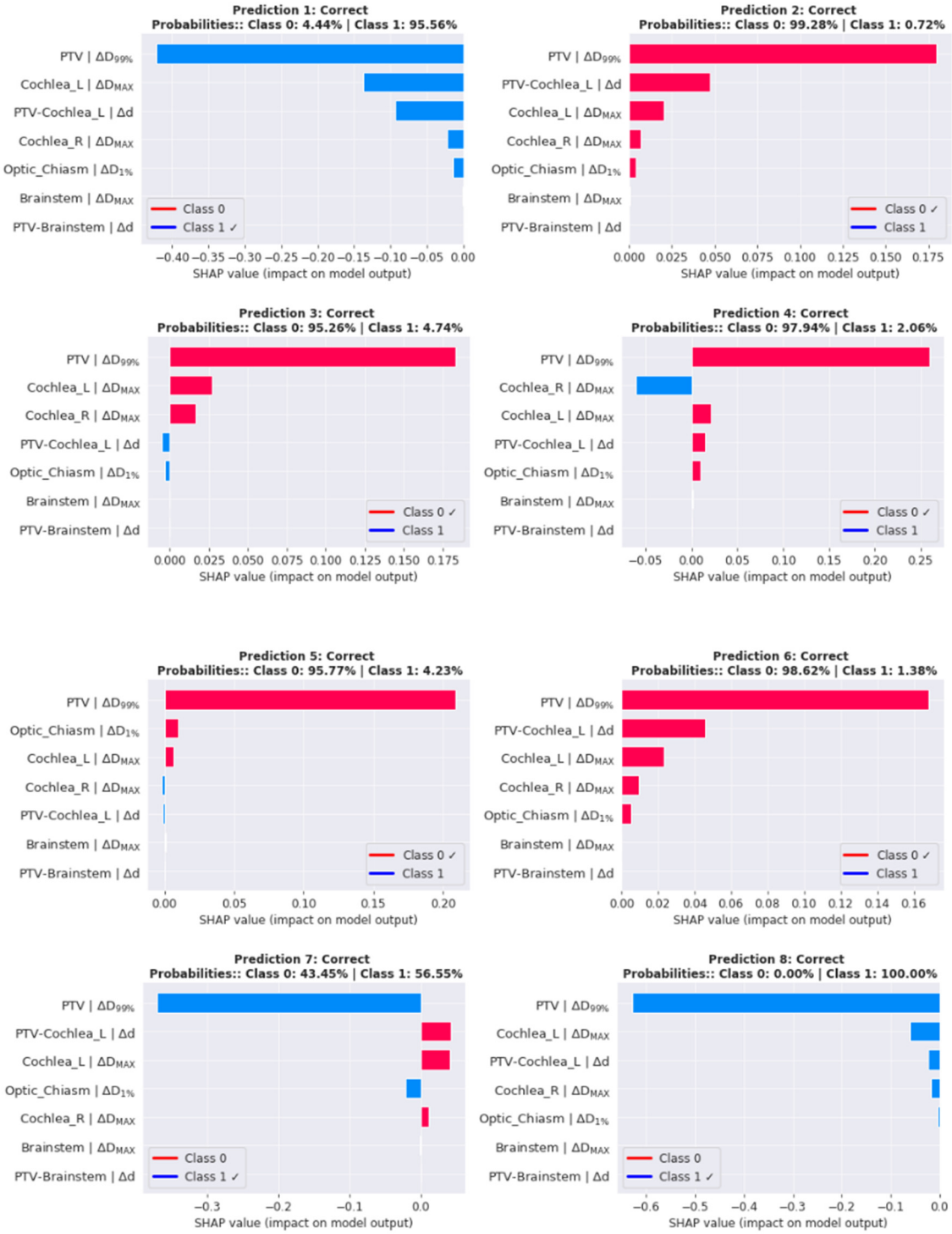



**Figure 8.90.** Cross-Validation Fold 3. Feature Importance for Single Predictions (1-8).

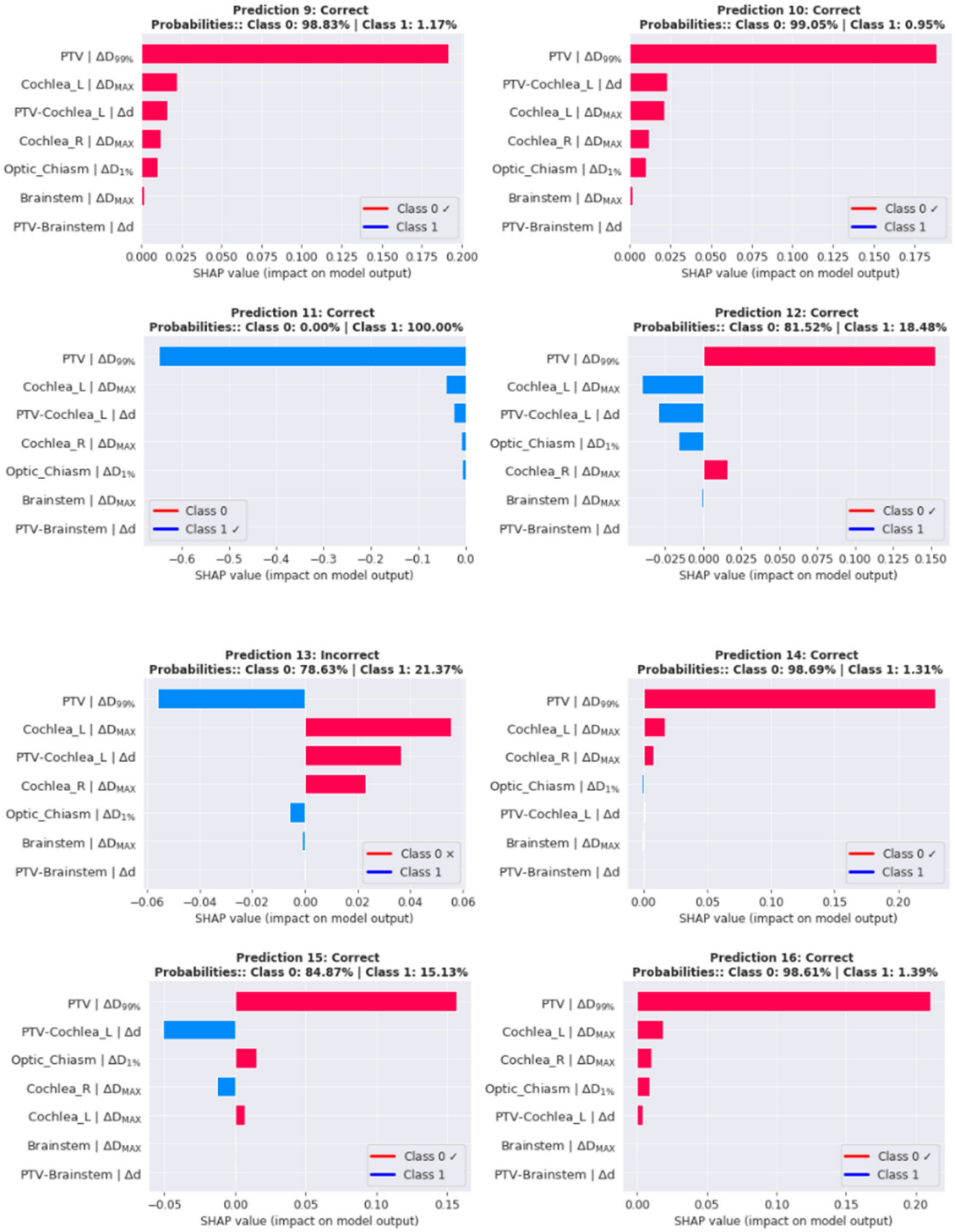
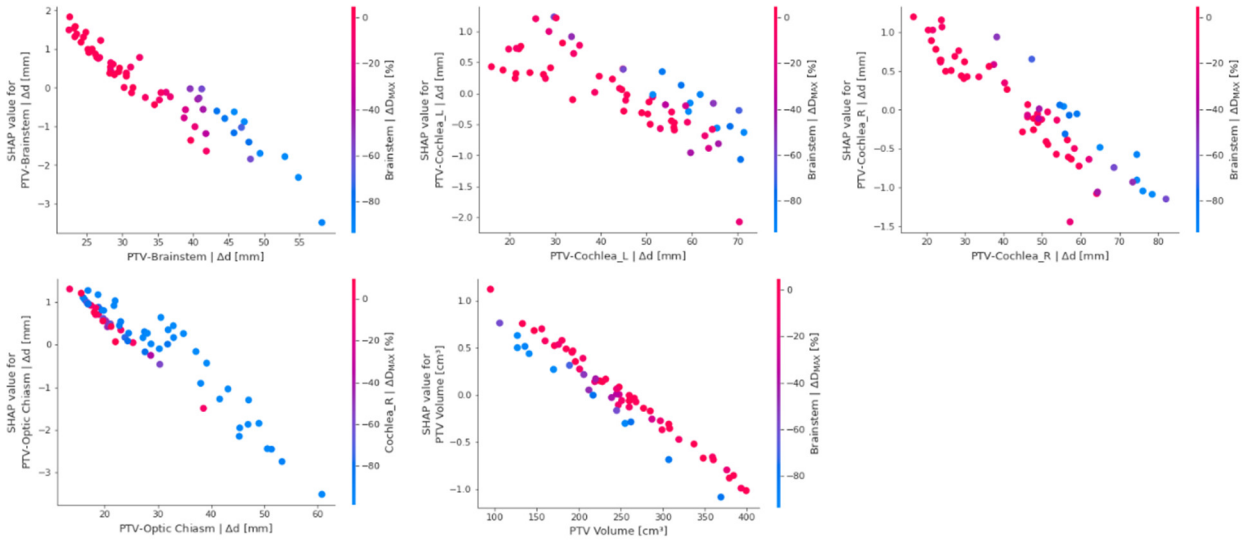



**Figure 8.91.** Cross-Validation Fold 3. Feature Importance for Single Predictions (9-16).

### Geometric Features




### Dosimetric Features

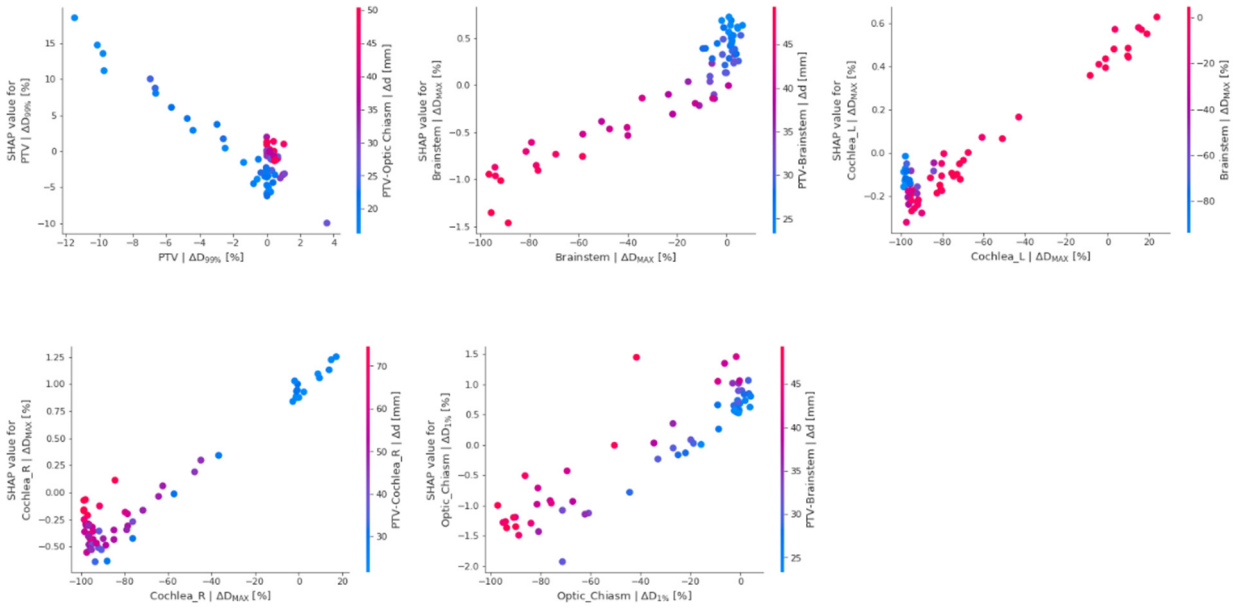



**Figure 8.92.** Cross-Validation Fold 3. Partial Dependency Charts.

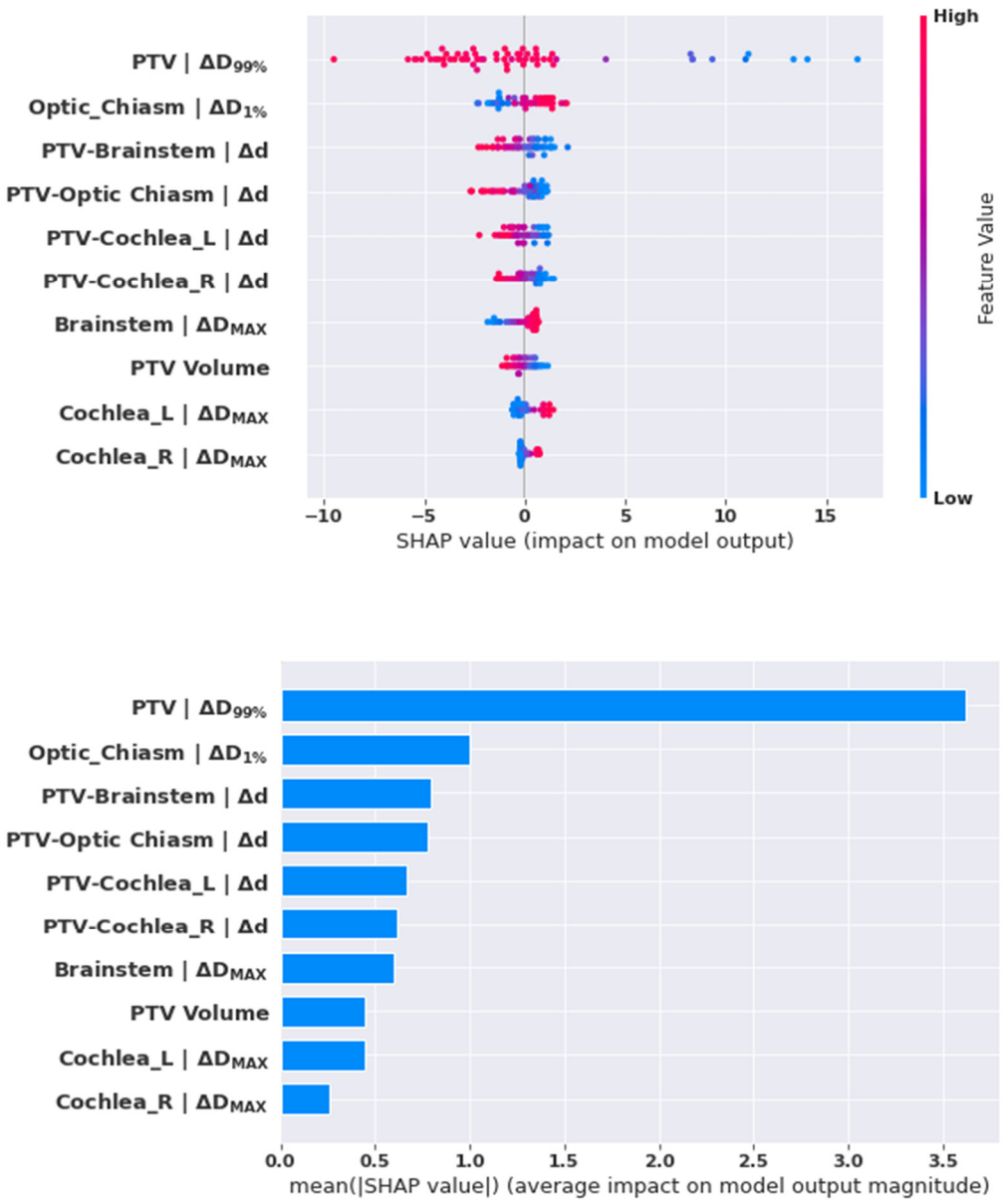



**Figure 8.93.** Cross-Validation Fold 4. Mean Feature Importance.

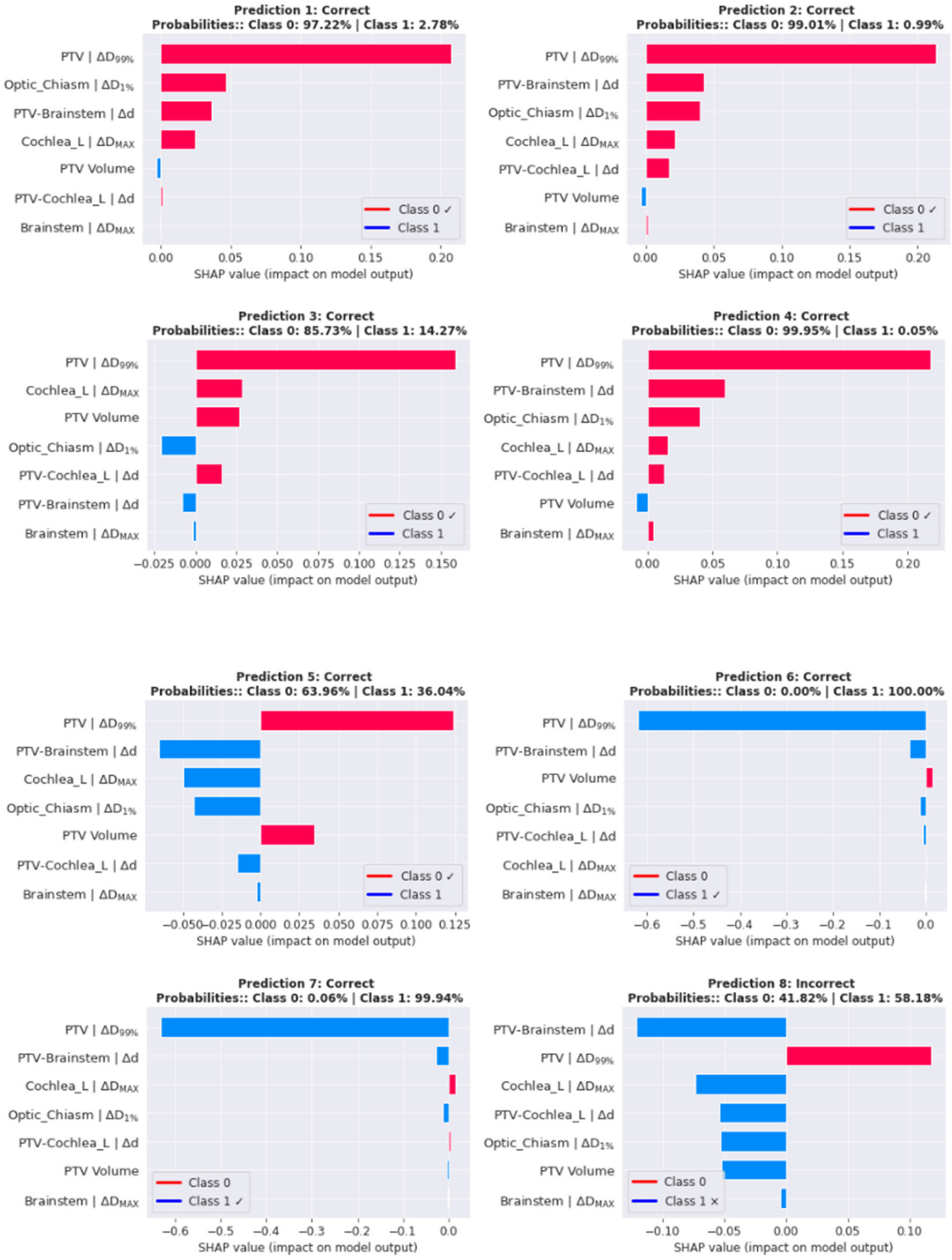



**Figure 8.94.** Cross-Validation Fold 4. Feature Importance for Single Predictions (1-8).

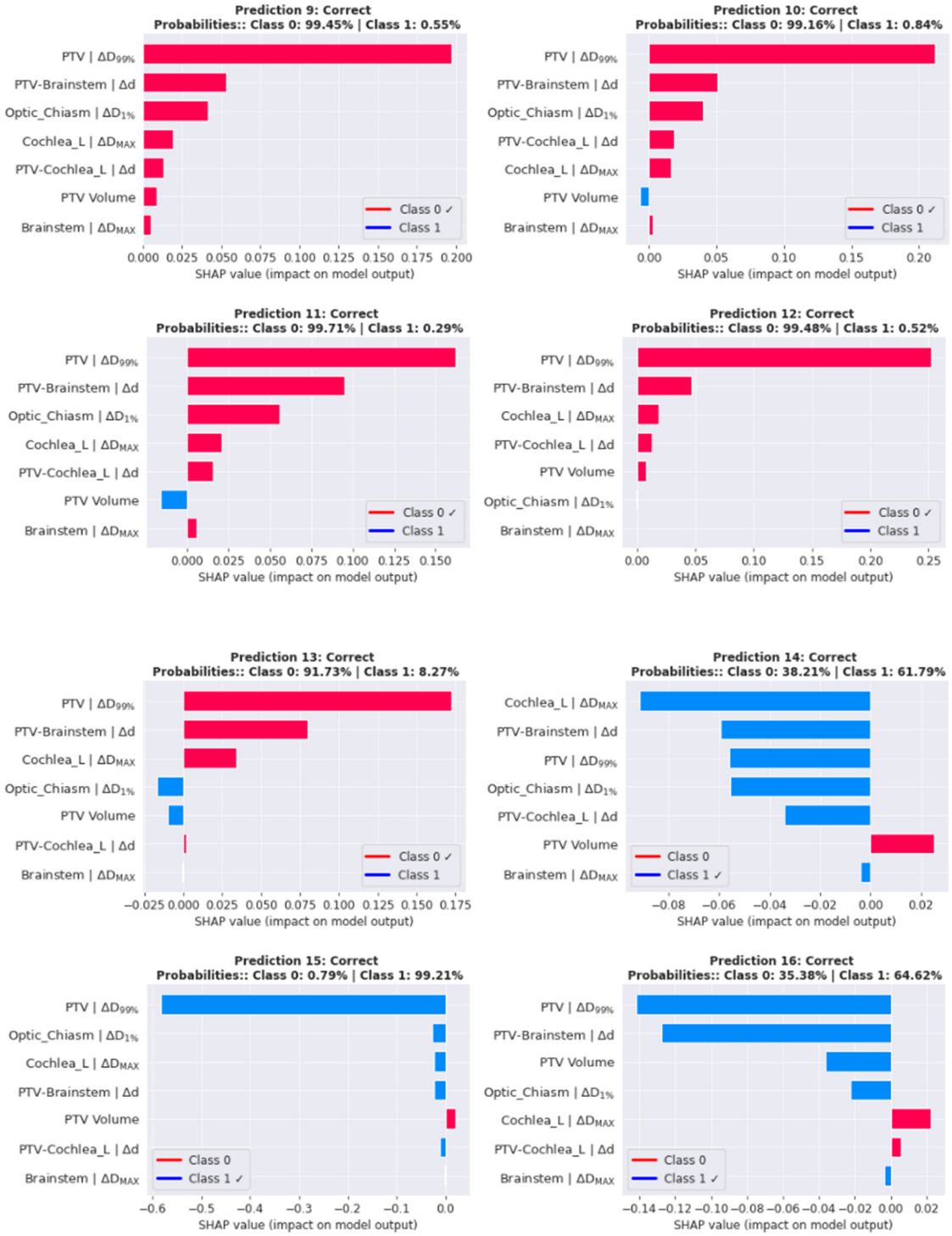
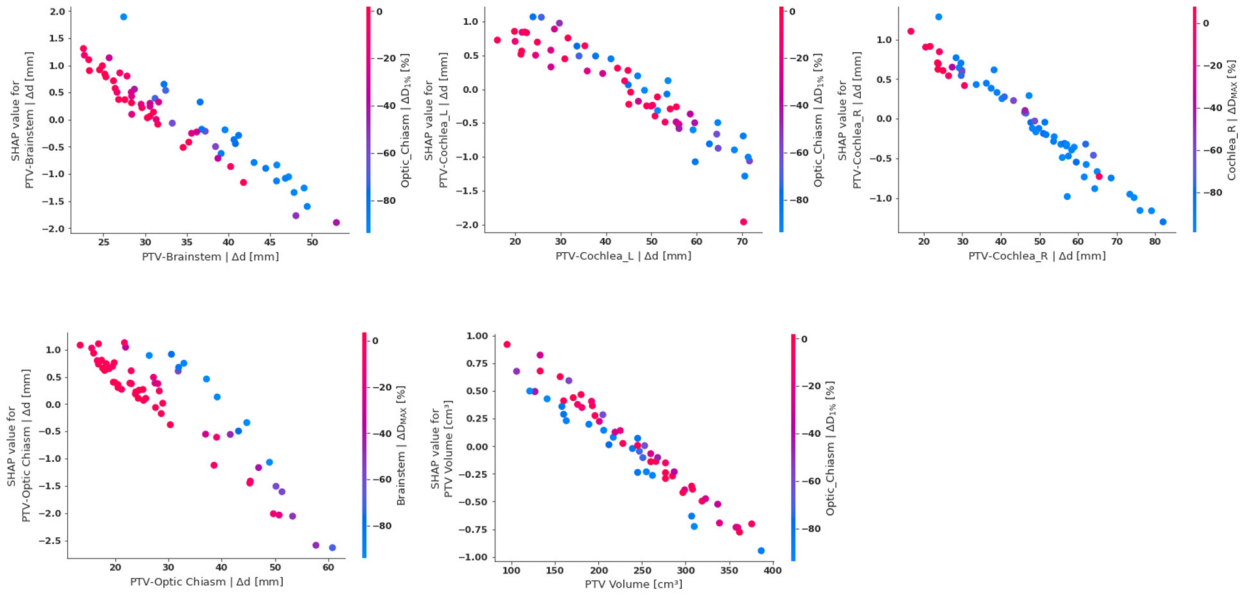



**Figure 8.95.** Cross-Validation Fold 4. Feature Importance for Single Predictions (9-16).

### Geometric Features




### Dosimetric Features

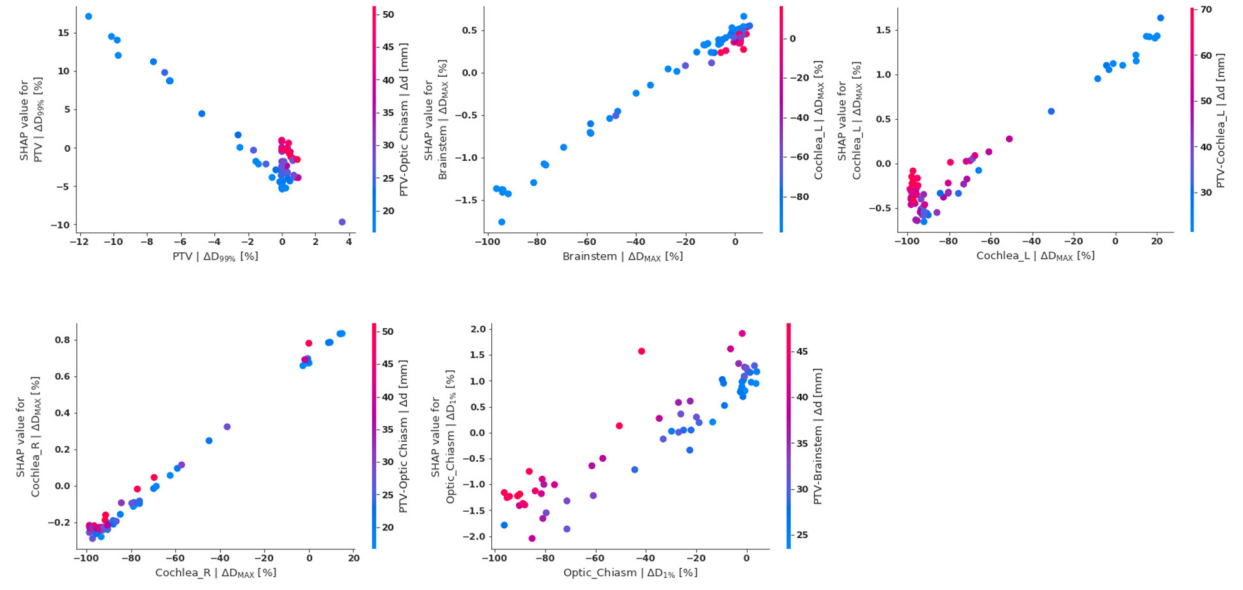



**Figure 8.96.** Cross-Validation Fold 4. Partial Dependency Charts.

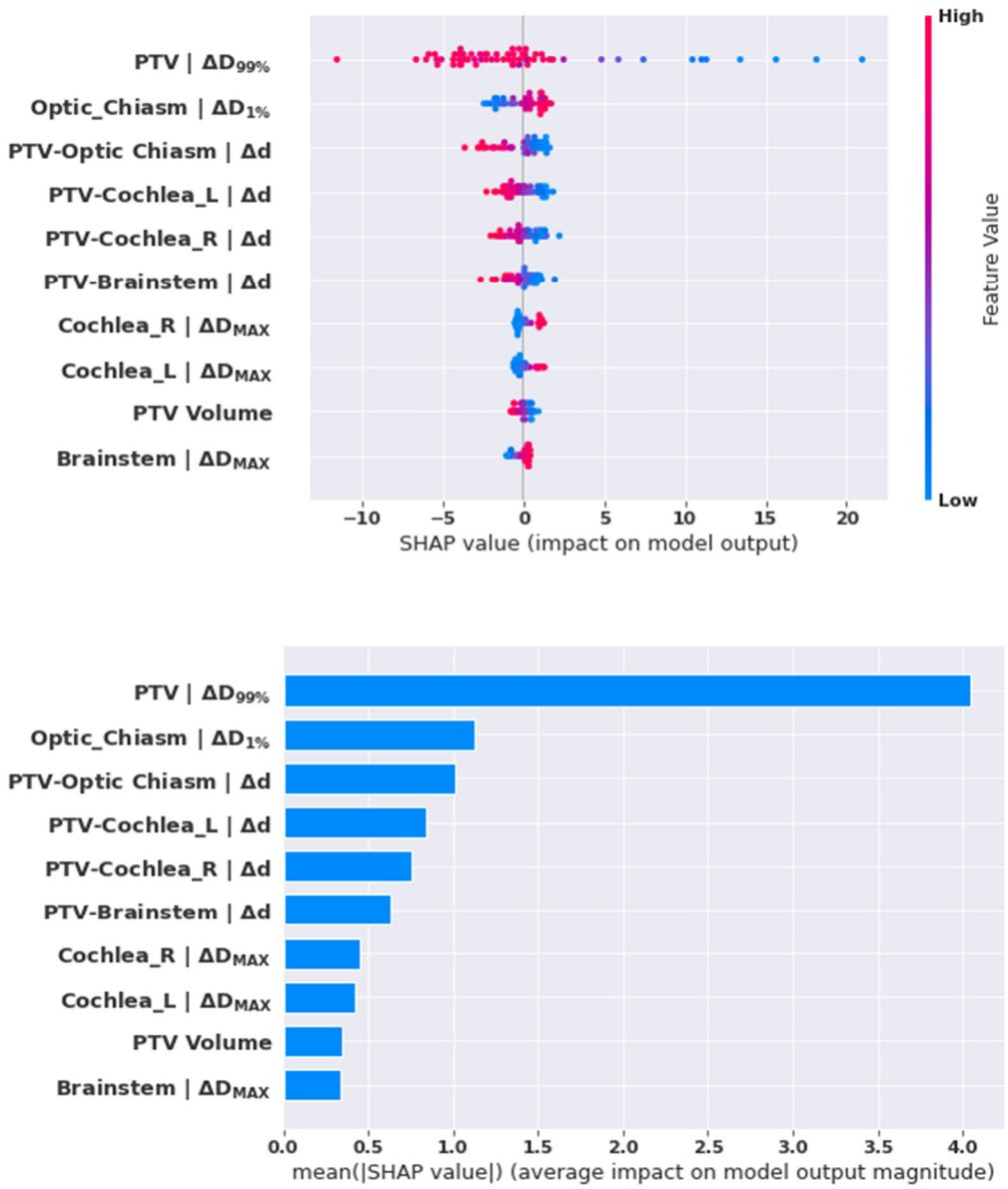



**Figure 8.97.** Cross-Validation Fold 5. Mean Feature Importance.

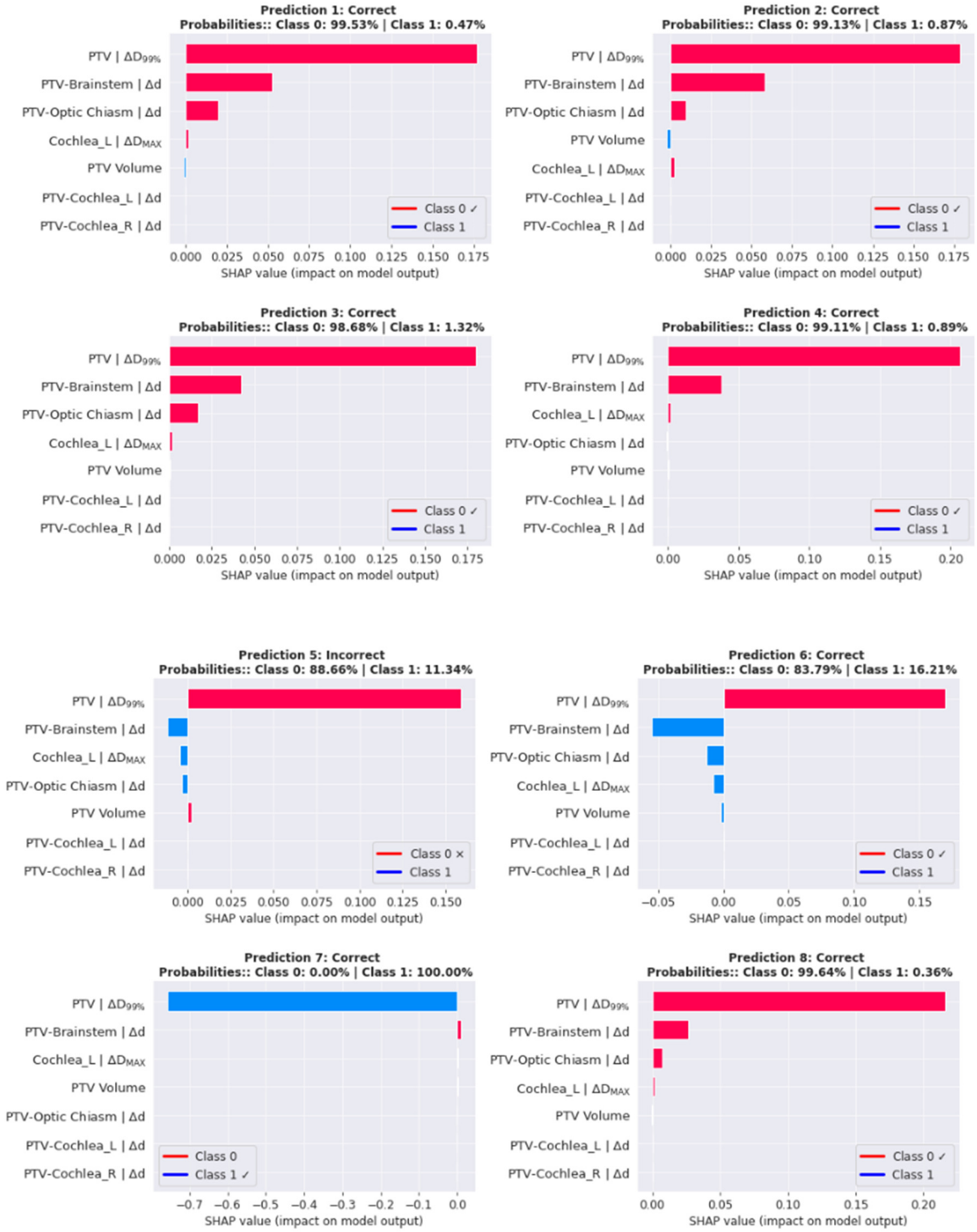



**Figure 8.98.** Cross-Validation Fold 5. Feature Importance for Single Predictions (1-8).

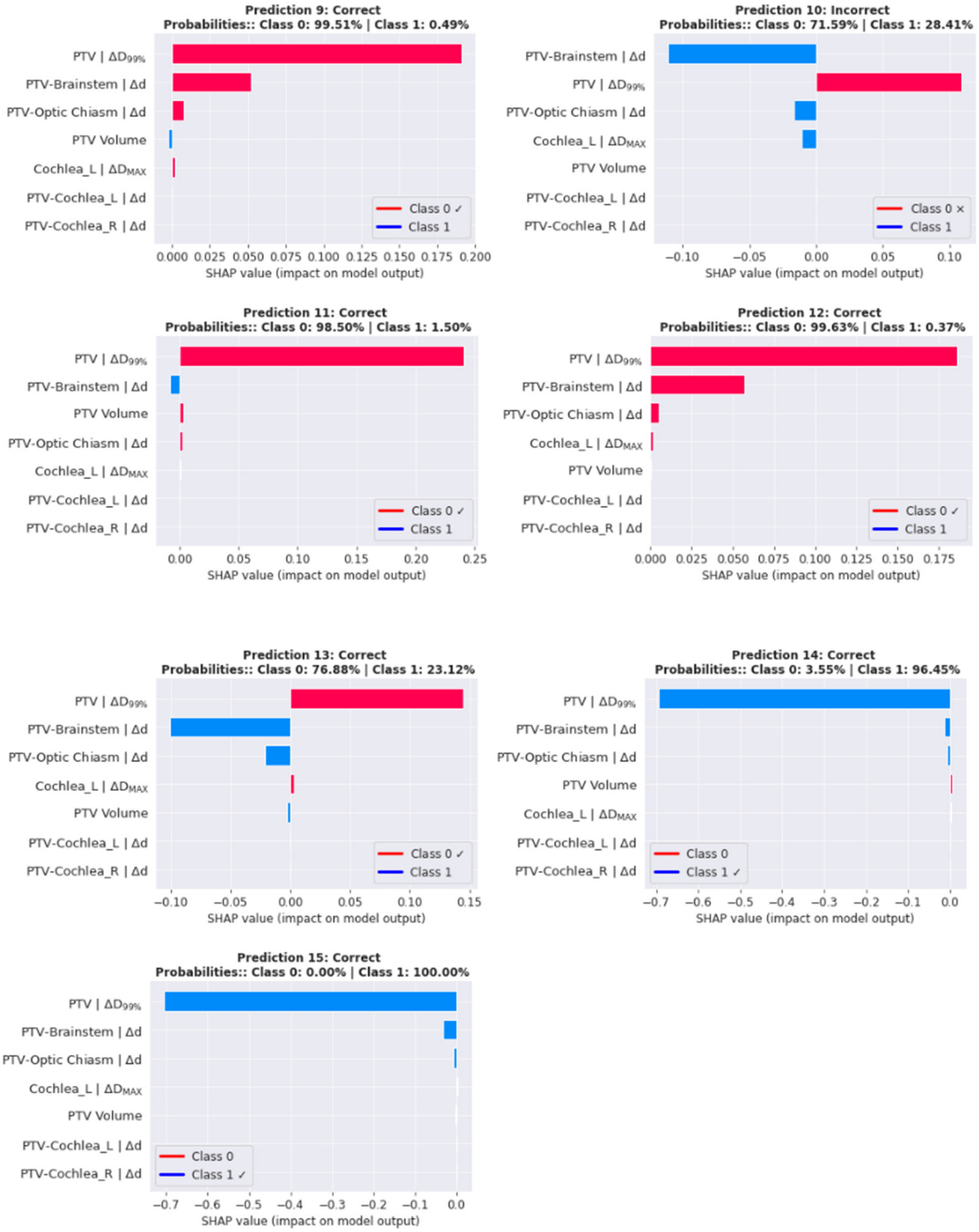
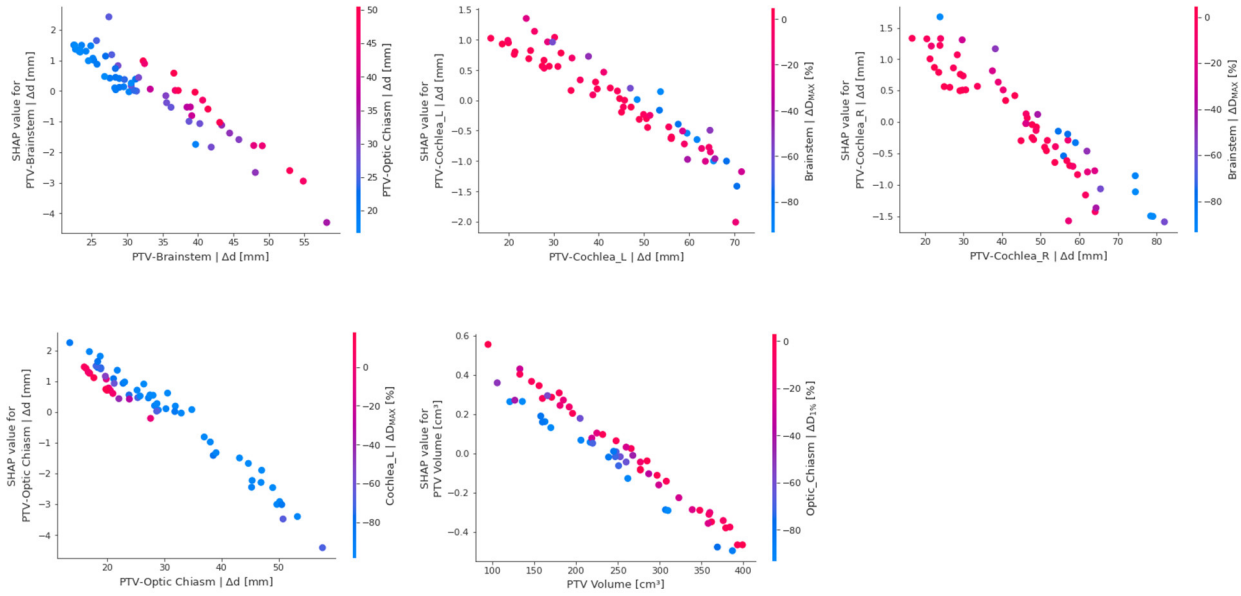



**Figure 8.99.** Cross-Validation Fold 5. Feature Importance for Single Predictions (9-15).

**Geometric Features**




**Dosimetric Features**

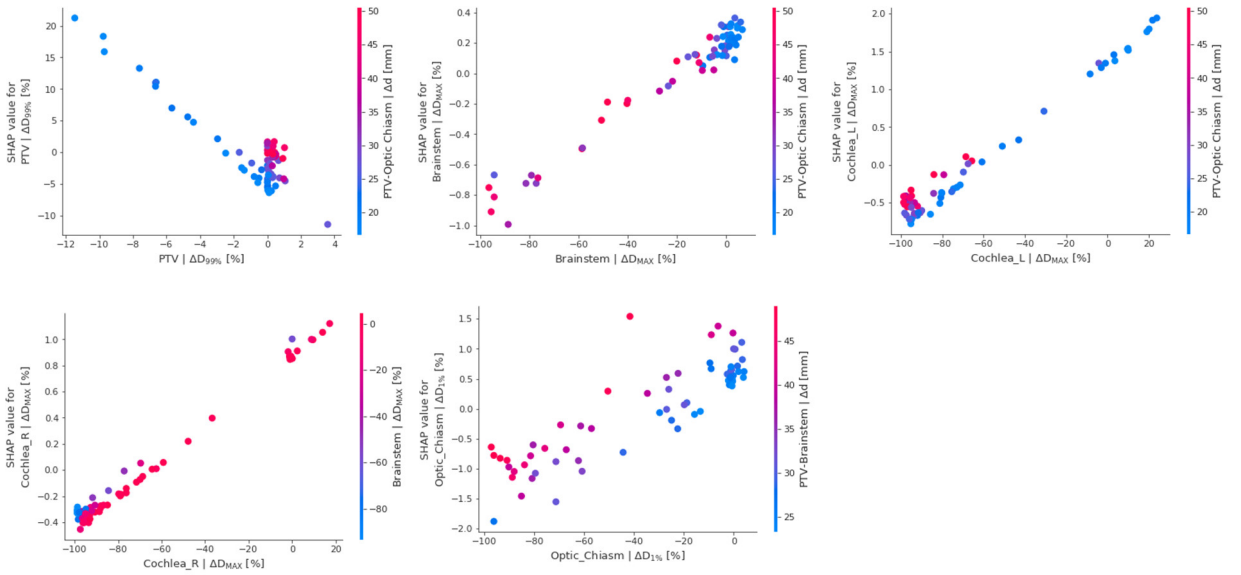



**Figure 8.100.** Cross-Validation Fold 5. Partial Dependency Charts.