HHS Public Access

Author manuscript

Med Phys. Author manuscript; available in PMC 2024 September 01.

Published in final edited form as:

Med Phys. 2023 September ; 50(9): 5354–5363. doi:10.1002/mp.16616.

Intentional deep overfit learning for patient-specific dose predictions in adaptive radiotherapy

Austen Maniscalco,

Xiao Liang,

Mu-Han Lin,

Steve Jiang,

Dan Nguyen

Medical Artificial Intelligence and Automation Laboratory, Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX, 75390, USA

Abstract

Background: The framework of adaptive radiation therapy (ART) was crafted to address the underlying sources of intra-patient variation that were observed throughout numerous patients' radiation sessions. ART seeks to minimize the consequential dosimetric uncertainty resulting from this daily variation, commonly through treatment planning re-optimization. Re-optimization typically consists of manual evaluation and modification of previously utilized optimization criteria. Ideally, frequent treatment plan adaptation through re-optimization on each day's computed tomography (CT) scan may improve dosimetric accuracy and minimize dose delivered to organs at risk (OARs) as the planning target volume (PTV) changes throughout the course of treatment.

Purpose: Re-optimization in its current form is time-consuming and inefficient. In response to this ART bottleneck, we propose a deep learning based adaptive dose prediction model that utilizes a head and neck (H&N) patient's initial planning data to fine-tune a previously trained population model towards a patient-specific model. Our fine-tuned, patient-specific (FT-PS) model, which is trained using the intentional deep overfit learning (IDOL) method, may enable clinicians and treatment planners to rapidly evaluate relevant dosimetric changes daily and re-optimize accordingly.

Methods: An adaptive population (AP) model was trained using adaptive data from 33 patients. Separately, 10 patients were selected for training FT-PS models. The previously trained AP model was utilized as the base model weights prior to re-initializing model training for each FT-PS model. 10 FT-PS models were separately trained by fine-tuning the previous model weights based on each respective patient's initial treatment plan. From these 10 patients, 26 ART treatment plans were withheld from training as the test dataset for retrospective evaluation of dose prediction performance between the AP and FT-PS models. Each AP and FT-PS dose prediction was

Dan.Nguyen@UTSouthwestern.edu .

CONFLICTS OF INTEREST

There are no conflicts of interest to disclose.

compared against the ground truth dose distribution as originally generated during the patient's course of treatment. Mean absolute percent error (MAPE) evaluated the dose differences between a model's prediction and the ground truth.

Results: MAPE was calculated within the 10% isodose volume region of interest for each of the AP and FT-PS models dose predictions and averaged across all test adaptive sessions, yielding 5.759% and 3.747% respectively. MAPE differences were compared between AP and FT-PS models across each test session in a test of statistical significance. The differences were statistically significant in a paired t-test with two-tailed p-value equal to 3.851×10^{-9} and 95% confidence interval equal to [−2.483, −1.542]. Furthermore, MAPE was calculated using each individually segmented structure as an ROI. 19 of 24 structures demonstrated statistically significant differences between the AP and FT-PS models.

Conclusion: We utilized the IDOL method to fine-tune a population-based dose prediction model into an adaptive, patient-specific model. The averaged MAPE across the test dataset was 5.759% for the population-based model vs 3.747% for the fine-tuned, patient-specific model, and the difference in MAPE between models was found to be statistically significant. Our work demonstrates the feasibility of patient-specific models in adaptive radiotherapy, and offers unique clinical benefit by utilizing initial planning data that contains the physician's treatment intent.

Keywords

radiation therapy; deep learning; artificial intelligence; adaptive; overfit; fine-tuning; dose prediction; head and neck cancer

I. INTRODUCTION

In radiation therapy (RT), personalized treatment plans are created for each patient based on their initial anatomical information as displayed by an acquired computed tomography (CT) image. An attending radiation oncologist may delineate segmented structures on the CT image such as gross tumor volumes (GTVs), clinical target volumes (CTVs), planning target volumes (PTVs), and surround organs at risk (OARs). A treatment planner can then select an arrangement of treatment fields. These fields determine the entrance angles from which particles will travel towards the patient. Next, the treatment planner utilizes an optimization algorithm specific to the type of particle to calculate a distribution of absorbed dose throughout the patient's body. This resultant dose distribution is tailored on a patient-specific basis to deliver high dose the patient's GTV(s), CTV(s) and PTV(s) while simultaneously minimizing dose delivered to surrounding OARs. The treatment planner and radiation oncologist collaborate over multiple days for each treatment plan in a laborious endeavor to generate the optimal, personalized dose distribution for each patient. However, this initial dose distribution is idealized because it is impractical to reproduce the exact same conditions as acquired by the initial CT image.

The accuracy and precision of initially planned dose distributions in radiation therapy can be limited by intra-patient anatomical variation caused by factors such as patient weight loss, changes in tumor size or geometric shape, and respiratory motion. Barker et al. conducted a study on head and neck (H&N) cancer patients and observed measurable changes in tumor

volumes and parotid glands throughout their course of radiation therapy, which could impact the delivered dose[1]. To improve treatment accuracy and precision, advanced image-guided radiation therapy (IGRT) techniques such as cone beam CT (CBCT) have been developed for imaging the patient at the treatment machine prior to treatment. While IGRT helps ensure consistent patient alignment, an IGRT review by Gupta et al. states that it alone is insufficient to address dosimetric impact resulting from anatomical changes[2]. In response to this shortcoming, adaptive radiation therapy (ART) techniques were developed to improve reproducibility of an initially planned dose distribution.

ART techniques can be described as either offline adaptation or online ART. Offline adaptation generally involves acquiring a new CT image of the patient and generating a new, corresponding treatment plan. More complex methods of offline adaptation have been utilized; Foroudi et al. proposed a process that utilizes information from numerous acquired CBCTs over a patient's treatment to update their delineated organs on the initial CT image[3]. These techniques may be used routinely in clinical workflows to address systematic anatomical discrepancies, but they are insufficient to address random daily variance[4]. On the other hand, online ART intends to rapidly adapt the treatment plan on the same day of delivery, therefore offering an avenue to minimize random, daily intra-patient variation. However, widespread utilization of online ART requires a streamlined workflow to minimize the treatment planning timeline and corresponding human intervention[5]. Vendors have recently introduced specialized online ART machines such as Varian's Ethos, Elekta's Unity, and Viewray's MRIdian, which generally offer streamlined workflows and advanced tools to assist clinical staff.

Varian's Ethos therapy, as an example, offers users visualizations of the recently acquired CBCT, automatically segmented OARs and PTVs, and a corresponding dose distribution. A new synthetic CT image is generated daily for dose calculation purposes through deformable image registration (DIR), a process in which the original CT image is deformed into the same geometry as that day's acquired CBCT[6]. While certain tools may assist in streamlining the online ART process, the automatically calculated dose distribution and clinical goals are not always sufficient to achieve an idealized dose distribution. For example, a publication by Sibolt et al. examined their clinical implementation of the Ethos therapy process and reported that complex cases with multiple dose levels required multiple revisions[7]. Additionally, Yoon et al. found that dose distributions calculated by the Ethos therapy process require a median time of 3 minutes and 10 seconds and as long as 7 minutes and 5 seconds[8]. Minimizing the patient's time on the table is vital to maximizing treatment plan reproducibility and optimizing overall clinical efficiency, but the time consumed by automatic dose calculation may be wasted if the resultant dose distribution is inadequate.

It is ideal to rapidly generate patient-specific dose distributions that simultaneously maintain the physician's original planning intent and yield an idealized dose distribution based on the patient's daily anatomy. A number of prior works have investigated deep learning (DL) based methods to generate H&N dose distribution predictions. Fan et al. utilized a now classical method to train a H&N dose prediction model with the PTV, OARs, and CT as input data and the ground truth dose distribution as a label in model training[9]. Yue et al. recently proposed an improved model training technique that includes distance information

in the input data of structures, yielding a H&N dose prediction model that demonstrated statistically superior performance over a model that withheld distance information[10]. Research in this field is actively ongoing with regards to clinical application of dose prediction models. Dose prediction as a quality assurance tool was recently suggested by Gronberg et al., in which they propose to flag OARs that receive excess dose as determined by an evaluation of the difference between an existing plan's dose and the model's dose prediction[11]. Despite recent activity in the topic of H&N dose prediction, we could not identify any prior works that trained patient-specific dose prediction models for ART.

In a publication by Chun et al., they identified an opportunity to integrate patient-specific information into DL model training for ART-involved tasks[12]. They posited that a pre-trained, generalized population model could be fine-tuned toward a patient-specific model for ART-specific tasks through a technique similar to transfer learning, which they termed an "intentional deep overfit learning" approach (IDOL). The IDOL terminology indicates that the resultant fine-tuned model is "overfit" for patient-specific tasks relative to a generalized population model. In other words, it is anticipated that an ART-specific model trained with the IDOL approach would perform well for patient-specific tasks at the cost of its generalizability in a broader population.

Inspired by the IDOL approach, we propose to fine-tune a previously trained population-based deep learning model using patient-specific data, resulting in a patient-specific ART dose prediction model. We anticipate that the fine-tuned, patient-specific (FT-PS) adaptive dose prediction model will offer immense clinical value in identification of systematic error, minimization of treatment planning burden and repetition, and standardization of plan quality across a patient's adaptive treatment sessions. The performance of the resulting FT-PS adaptive dose prediction model can then be compared to a more traditional population-based model that was trained on a dataset of numerous patients' plans from their ART course of treatment. We refer to the latter population-based model as an adaptive population (AP) model. We hypothesize that the fine-tuned, patient-specific adaptive dose prediction model will yield superior dose predictions for a respective patient as compared to a traditional, adaptive population-based DL dose prediction model.

II. METHODS

II.1. H&N Patient Data

For this study, we selected 43 H&N patients that were treated with ART intent at our institution on the Varian Ethos machine. Each patient contains an initial RT plan and a variable number of adaptive RT plans, dependent upon how many adaptive sessions were generated during the course of their radiation therapy. In the context of this study, only segmented structures, such as the relevant PTV and surrounding OARs, and a corresponding radiation dose distribution were required from each RT plan for model training.

All relevant patient files were extracted from the treatment planning system and saved in the Digital Imaging and Communications in Medicine (DICOM) format. For our data preparation process, all DICOM files were converted to NumPy arrays, resized the arrays to 5mm × 5mm × 5mm voxel spacing, and saved in the corresponding NumPy file format.

The resulting segmented structure sets were pruned of extraneous structures that were not representative of OARs. All OARs were stored as binary arrays such that an element containing 0 indicates the OAR is not present in its corresponding voxel, and an element containing 1 indicates the presence of the OAR in its corresponding voxel. All PTV arrays were summed into a single PTV array such that each element contained (in units of Gray) the largest prescription dose value for a PTV located in its corresponding voxel, or 0 if no PTV is present. The PTV and dose arrays were left unscaled at their original plan normalization values for model training.

33 out of the 43 total patients were dedicated to building the population-based dose prediction model, otherwise referred to as the AP model (Figure 1). In training the AP model, 30/33 patients were dedicated to the training dataset and 3/33 patients were dedicated to the validation dataset. Therefore, the training dataset was comprised of 30 initial RT plans and 55 ART plans and the validation dataset was comprised of 3 initial RT plans and 7 ART plans.

The remaining 10 out of 43 total patients were further partitioned: Their 10 initial RT plans were dedicated to training FT-PS models, and their remaining 26 ART plans were withheld from all model training as the test dataset. As a consequence of the FT-PS model training technique, and in the spirit of maintaining this study's clinical relevance, there was no viable validation dataset to use during model training in this study. In current clinical practice, it is unlikely to have more than one unique RT plan for an individual patient prior to their first day of treatment – see discussion section for more. The 10 unique FT-PS models were trained by initializing model training with the AP model's base weights and fine-tuning from there with only a single sample – the respective patient's initial RT plan.

II.2. Model Training

To train our DL-based dose prediction models, we set up the Hierarchically Densely Connected U-Net (HD U-Net) convolutional neural network (CNN) architecture as proposed and described in a paper by Nguyen et al[13] with four pooling layers, ReLU activations, and Spatial 3D Dropout at a rate of 0.1. The three-dimensional arrays of PTV and OARs were stacked into four-dimensional arrays such that the first element of the fourth dimension represented the PTV and sequential elements represented each unique OAR. In this study, we identified a PTV and 44 relevant OARs for each RT plan, so the fourth dimension contained 45 elements. If any particular OAR was not present in an RT plan, its three-dimensional array for that RT plan was entirely zeros. Finally, the three-dimensional dose distribution array was also expanded to four-dimensions, with only a single element in the fourth dimension. The PTV, OARs and dose distribution from a single RT plan encompass one sample. Each sample in this study was subject to randomized patch selection along the three spatial dimensions. The patch selection process only retained 96 sequential elements from the x- and y-axes, and 64 sequential elements from the z-axis. Furthermore, each selected patch was subject to random rotations and flips as an additional method of data augmentation. A number of these sample patches were then stacked in a newly created fifth dimension. The resulting five-dimension arrays were both converted to rank-5 tensors. A single batch in this study is composed of both rank-5 tensors. The tensor of PTV and OARs

was utilized as input and the dose distribution tensor served as the label for supervised learning during DL model training. With this technique, we were able to maintain the same shape for each type of rank-5 tensor across all RT plans.

Across all model training in this study, an NVIDIA V100 graphics processing unit (GPU) with 32 gigabytes (GB) of video random access memory (VRAM) was utilized. As a minor consequence of performing DL model training with the computationally intense HD U-Net, mini-batches were limited by VRAM to 5 samples patches at a time. All model training was performed with the Adam optimizer, a constant learning rate of 1e-4, and mean squared error as the loss function. The mean squared error (MSE) as a loss function is calculated by summing over the squared value of voxel-wise differences between the model's dose prediction and the dose distribution label, and then dividing the sum by the total number of voxels.

In the context of AP model training, 17 steps per epoch were required to iterate through all 85 RT plans one time. AP model training started with randomly initialized weights, and training was terminated when validation loss converged. AP model weights were saved at the epoch corresponding to the lowest validation loss.

For FT-PS model training, as there is only one RT plan in each training dataset, 5 sample patches were selected from the same RT plan for each mini-batch. All FT-PS models loaded the previously trained AP model weights prior to training. Limited by a lack of available validation data, fine-tuning of each FT-PS model was terminated after an arbitrarily large number of iterations (5000 iterations in this study) at a rate of 10 steps per epoch, and each FT-PS model's weights were saved at the end of its training.

II.3. Model Evaluation

To compare the performance of the AP and FT-PS models, the AP model and the corresponding patient's FT-PS model were used to predict dose distributions for each test patient's ART sessions. The resultant dose distributions were normalized based on PTV coverage, ensuring that 95% of the highest dose PTV receives at least 100% of its corresponding prescribed dose (D95%=100%). Mean Absolute Percent Error (APE) was then evaluated as:

$$MAPE(\%) = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{D_{GT_i} - D_{PRED_i}}{D_{RX}}\right| \times 100 \quad (1)$$

Here, D_{GT_i} represents the ground truth dose of the i^{th} voxel in units of Gy, D_{PRED_i} represents a model's predicted dose at the corresponding i^{th} voxel in units of Gy, and D_{RX} represents the patient's highest prescription dose in units of Gy. In overall model prediction performance analysis, we evaluated MAPE within the 10% isodose volume region of interest (ROI).

III. RESULTS

The average MAPE within the 10% isodose volume ROI across the entire test dataset was 5.759% for the AP model, as compared to 3.747% for the FT-PS model. For evaluation of statistical significance, we used a paired t-test to compare MAPE results from each model type across each individual test sample. The paired t-test yielded a two-tailed p-value of 3.851×10^{-9} along with a 95% confidence interval (CI) equal to [−2.483, −1.542]. The p-value indicates that the MAPE differences between the AP model and the FT-PS model are statistically significant across the test dataset. The 95% CI indicates a 95% probability that the population value lies within that range, or in other words, there is a 95% probability that the MAPE difference falls in that range across the population.

In our analysis of individually segmented structures, we used MAPE to evaluate the differences between a model's prediction for dose delivered to a structure as compared to the ground truth mean dose for that structure (Figure 2). Statistically significant differences in MAPE between the FT-PS and AP models were identified for 19 out of 24 structures. Additionally, the max dose a structure was predicted to receive (DMAX) was evaluated as compared to the ground truth dose for each individual structure (Figure 3). We found statistically significant differences in terms of DMAX error between the AP and FT-PS models for 13 of the 24 structures. For individual PTV statistical analysis, we evaluated the dose prediction for the highest dose PTV in each treatment plan as compared to the ground truth in terms of D98%, D99%, homogeneity ($\frac{D2\% - D98\%}{D50\%}$)and conformity index (Table 1).

To calculate conformity index, the volume of PTV receiving prescription dose was squared and then divided by the product of patient body volume receiving 100% prescription dose and the PTV volume itself. PTV p-values for D98%, D99%, homogeneity and conformity index exceeded 0.05, indicating that there was no statistically significant difference in these values between models.

We plotted the trend in MAPE over time for the AP and FT-PS models across all test patients, offering visualization of performance changes between adaptive fractions (Figure 4). Dose displays were generated for an individual patient's final adaptive session to visually demonstrate each model's performance (Figure 5). Dose difference displays were also generated, representing the difference between each predicted dose and the actual session dose (Figure 6). The Dose Volume Histogram (DVH) for this patient was generated for each of the ground truth, AP model predicted dose, and FTPS model predicted dose (Figure 7).

IV. DISCUSSION

One of the most important considerations around the integration of this work in a clinical workflow is duration of model training and how long it takes to generate an ART dose prediction with the corresponding model. For a single patient, FT-PS model training for 1000 iterations on our NVIDIA V100 GPU could be completed approximately 1 hour. It is typical for a treatment plan to be prepared multiple days in advance of a patient's start date, leaving sufficient time for FT-PS model training. However, if a shorter timeframe is required, it is still achievable to schedule multiple model trainings such that they complete overnight. In training multiple FT-PS models, it would be ideal to schedule all model

training immediately and sequentially based on patient start date. Once the FT-PS model is trained, generation of an ART dose prediction can be completed in approximately one minute or less.

The strong performance of the FT-PS adaptive dose prediction models indicates that they may integrate well in clinical workflows for the personalization of adaptive radiotherapy. One proposed integration of this work in clinical practice is to generate treatment planning optimization objectives based on the FT-PS dose prediction. These automatically generated optimization objectives would retain the physician's original dose distribution intent for the patient, while simultaneously utilizing trained knowledge from the relevant population and the current patient's anatomical information. Traditional, manual treatment planning requires a large degree of manual intervention by a treatment planner to achieve a physician's idealized dose distribution for each patient. We believe that the offering of patient-catered optimization objectives by FT-PS models is immensely valuable to widespread integration of ART in clinical workflows, as it may streamline the ART workflow, standardize plan quality across treatment sessions, and minimize manual intervention.

Another proposed integration of this work is an alert-based system that analyzes the changes in FT-PS dose predictions over each adaptive session and reports when significant deviance is identified from the original plan intent. As an example, the PTV may initially encroach on an OAR during the first few fractions of radiation therapy. Over time, if the PTV recedes from the OAR, it may be possible to perform a significant plan alteration that allows for the remainder of treatment to spare that OAR. To identify such a case, a notification could be send to relevant parties requesting manual intervention when a pre-specified dose threshold for an OAR is exceeded.

Our work encountered many limitations, partially a consequence of the relative novelty of the Varian Ethos treatment machine and ART software. We identified 43 eligible patients who met all necessary study criteria, notably limiting our sample size for population model training. However, these 43 patients were comprised of 43 initial RT plans and 88 ART plans. From this, the population model training dataset was comprised of 30 patients and their 85 unique RT plans. Furthermore, randomized patch selection and data augmentation may alleviate sample size concerns. While a larger sample size could potentially improve AP model performance, we anticipate that this may also improve the FT-PS model performance. As another limitation, our input data only utilized a portion of all available RT data. We suspect that addition of more patient-specific data, such as the CT image or field geometry, could improve personalization of dose prediction models and offer a corresponding improvement in performance.

Lastly, our study was limited by the lack of a validation dataset in training of FT-PS models. Without a validation dataset, an arbitrary cutoff point for FT-PS model training was required. As a consequence, the current training technique has no method for evaluation of whether the model is "best fit" or potentially "too overfit". We suspect that FT-PS model performance can be better optimized if a suitable validation dataset could be utilized during model training. One potential solution for a validation dataset could be generation of

randomly deformed, computer-generated patient data based on the initial RT data. Relatedly, it could be hypothesized that this training technique yields a prediction model that is initial plan specific rather than patient-specific, though the distinction may not be essential so long as the resulting model is adequate for the course of the respective patient's ART.

For our future work, we anticipate integration of additional patient-specific information as input for model training to improve model prediction performance. Figure 4 suggests that MAPE may increase for a patient as time from initial CT simulation increases, but a future study would need to evaluate this for statistical significance. We suspect that the consistency of an ART dose prediction model's performance over a patient's course of treatment can be improved through additional fine-tuning the FT-PS model as new data becomes available. With regards to clinical integration of this work, we plan to utilize FT-PS models in our ART workflow to generate adaptive dose predictions. The resulting adaptive dose predictions would then be used as previously described to minimize manual intervention and standardize plan quality across sessions in each patient's course of adaptive radiotherapy.

V. CONCLUSION

We fine-tuned a previously trained, population-based deep learning model with a single sample – a single patient's initial radiation therapy plan – based on the intentional deep overfit learning (IDOL) principle. The resulting patient-specific dose prediction model can generate dose distribution predictions on that respective patient's future radiation therapy data, which is particularly valuable in an adaptive radiotherapy clinical workflow. The mean absolute percent error (MAPE) was calculated within the 10% isodose volume region of interest for each testing session, and averaged for each model. The population-based dose prediction model yielded an average value of 5.759% as compared to the fine-tuned, patient-specific dose prediction models value of 3.747%. We reported a two-tailed p-value of 3.851×10^{-9} with a confidence interval of [−2.483, −1.542], indicating that the fine-tuned, patient-specific models yielded statistically significant MAPE differences across the test dataset as compared to the population-based model. The strong performance of the fine-tuned, patient-specific dose prediction models on the test dataset indicates that it may be well-suited to highlighting dose distribution changes over time throughout organs at risk as the patient anatomy changes. Our proposed fine-tuned, patient-specific model offers significant clinical benefit in adaptive radiation therapy due to its retention of the physician's original planning intent for that patient. In future work, we are considering integration of additional patient-specific information to improve model prediction performance and further personalize our models. Additionally, we anticipate integration of this work into a clinical workflow to save time during the adaptive radiation therapy treatment planning process.

ACKNOWLEDGEMENTS

This study is supported by NIH grants R01CA237269, R01CA254377, and R01CA258987.

Note regarding data availability:

Authors will share data upon request to the corresponding author.

REFERENCES

1. Barker JL Jr, Garden AS, Ang KK, et al. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. International Journal of Radiation Oncology* Biology* Physics. 2004;59(4):960–970. [PubMed: 15234029]
2. Gupta T, Narayan CA. Image-guided radiation therapy: Physician's perspectives. Journal of medical physics/Association of Medical Physicists of India. 2012;37(4):174.
3. Foroudi F, Wong J, Haworth A, et al. Offline adaptive radiotherapy for bladder cancer using cone beam computed tomography. Journal of medical imaging and radiation oncology. 2009;53(2):226–233. [PubMed: 19527372]
4. Sonke J-J, Aznar M, Rasch C. Adaptive radiotherapy for anatomical changes. Paper presented at: Seminars in radiation oncology 2019.
5. 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;35:43–52. [PubMed: 27909957]
6. Byrne M, Archibald-Heeren B, Hu Y, et al. Varian ethos online adaptive radiotherapy for prostate cancer: Early results of contouring accuracy, treatment plan quality, and treatment time. Journal of Applied Clinical Medical Physics. 2022;23(1):e13479. [PubMed: 34846098]
7. Sibolt P, Andersson LM, Calmels L, et al. Clinical implementation of artificial intelligence-driven cone-beam computed tomography-guided online adaptive radiotherapy in the pelvic region. Physics and imaging in radiation oncology. 2021;17:1–7. [PubMed: 33898770]
8. Yoon SW, Lin H, Alonso-Basanta M, et al. Initial evaluation of a novel cone-beam CT-based semi-automated online adaptive radiotherapy system for head and neck cancer treatment–a timing and automation quality study. Cureus. 2020;12(8).
9. Fan J, Wang J, Chen Z, Hu C, Zhang Z, Hu W. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. Medical physics. 2019;46(1):370–381. [PubMed: 30383300]
10. Yue M, Xue X, Wang Z, et al. Dose prediction via distance-guided deep learning: Initial development for nasopharyngeal carcinoma radiotherapy. Radiotherapy and Oncology. 2022;170:198–204. [PubMed: 35351537]
11. Gronberg MP, Beadle BM, Garden AS, et al. Deep Learning–Based Dose Prediction for Automated, Individualized Quality Assurance of Head and Neck Radiation Therapy Plans. Practical radiation oncology. 2023.
12. Chun J, Park JC, Olberg S, et al. Intentional deep overfit learning (IDOL): A novel deep learning strategy for adaptive radiation therapy. Medical physics. 2022;49(1):488–496. [PubMed: 34791672]
13. Nguyen D, Jia X, Sher D, et al. 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture. Physics in medicine & Biology. 2019;64(6):065020. [PubMed: 30703760]

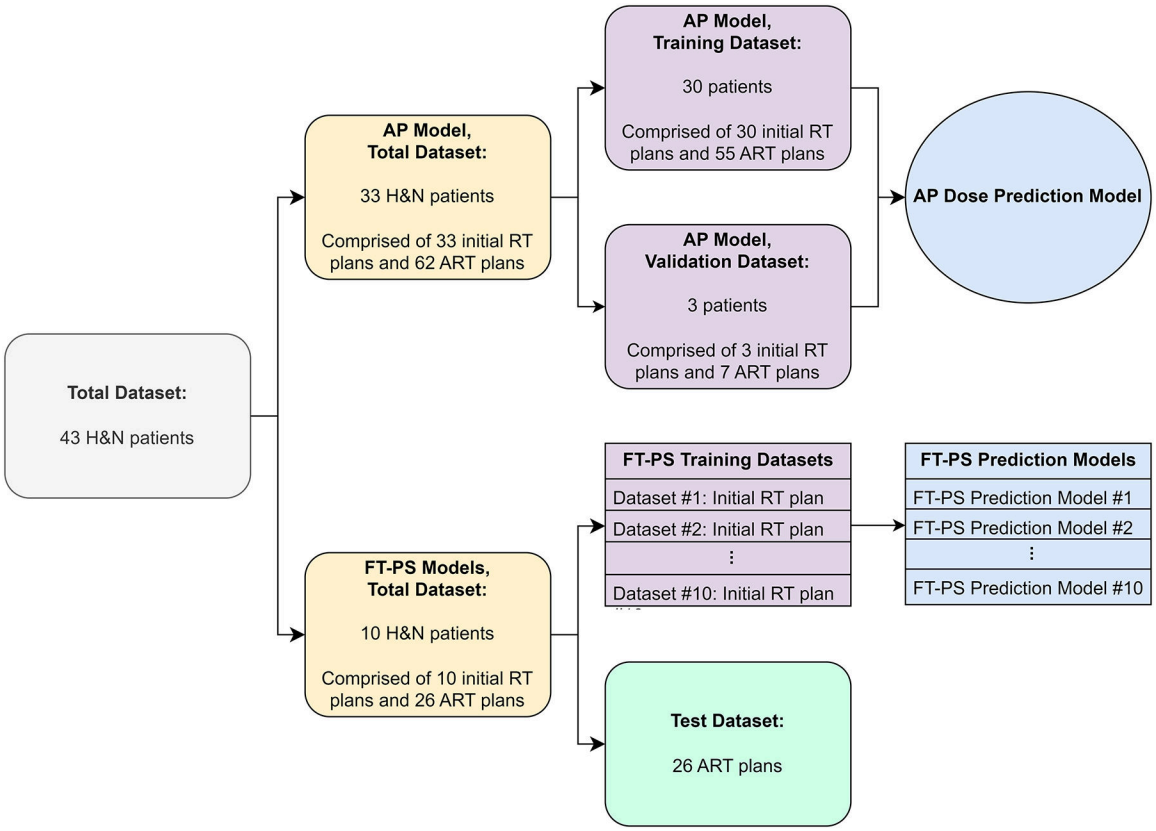

Figure #1:
H&N Patient Data Allocation

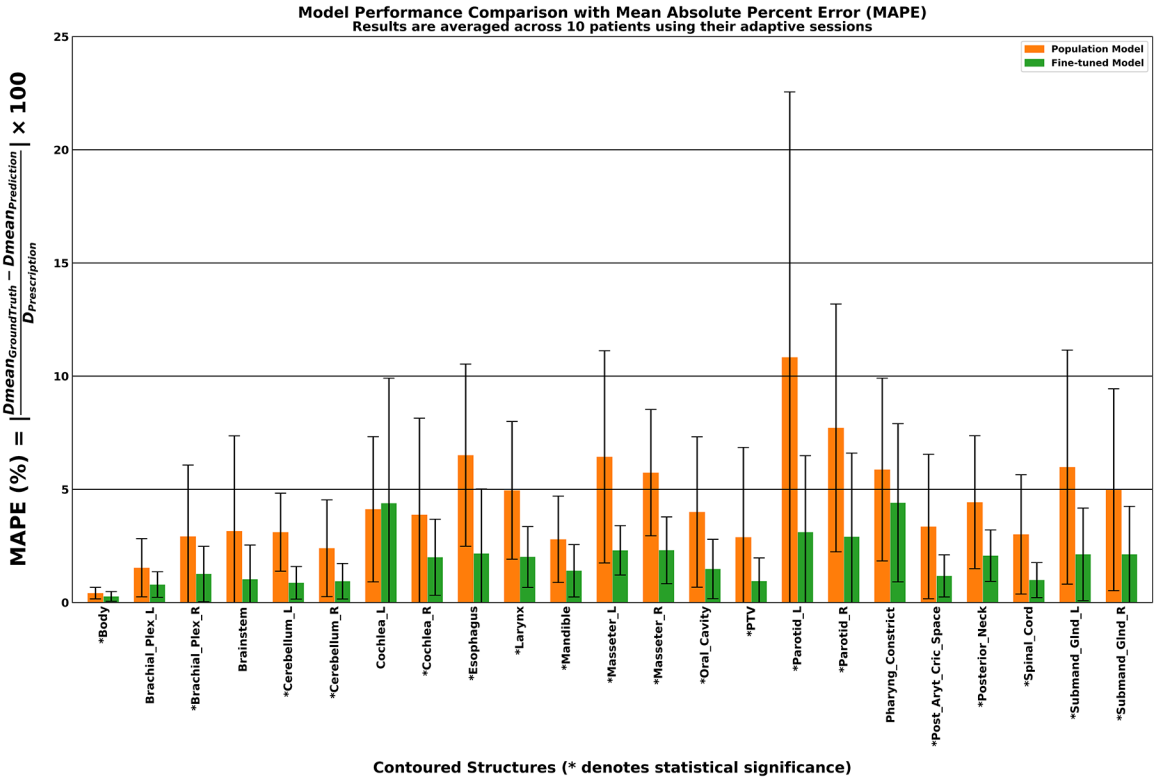

Figure #2:

Dose prediction model comparison in terms of MAPE between the adaptive population model and the fine-tuned patient-specific model. Results are averaged across all 26 adaptive treatment plans.

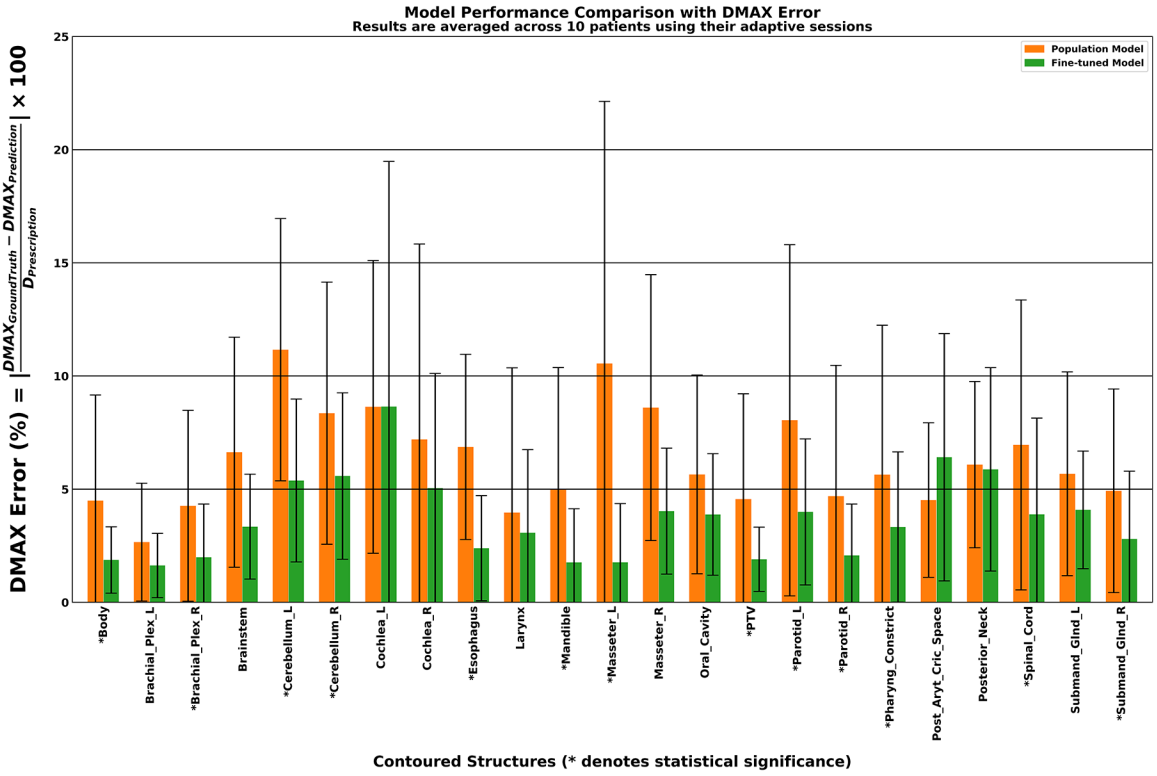

Figure #3:

Dose prediction model comparison in terms of each predicted structure's max dose as compared to the ground truth, referred to as DMAX error. Results are averaged across all 26 adaptive treatment plans.

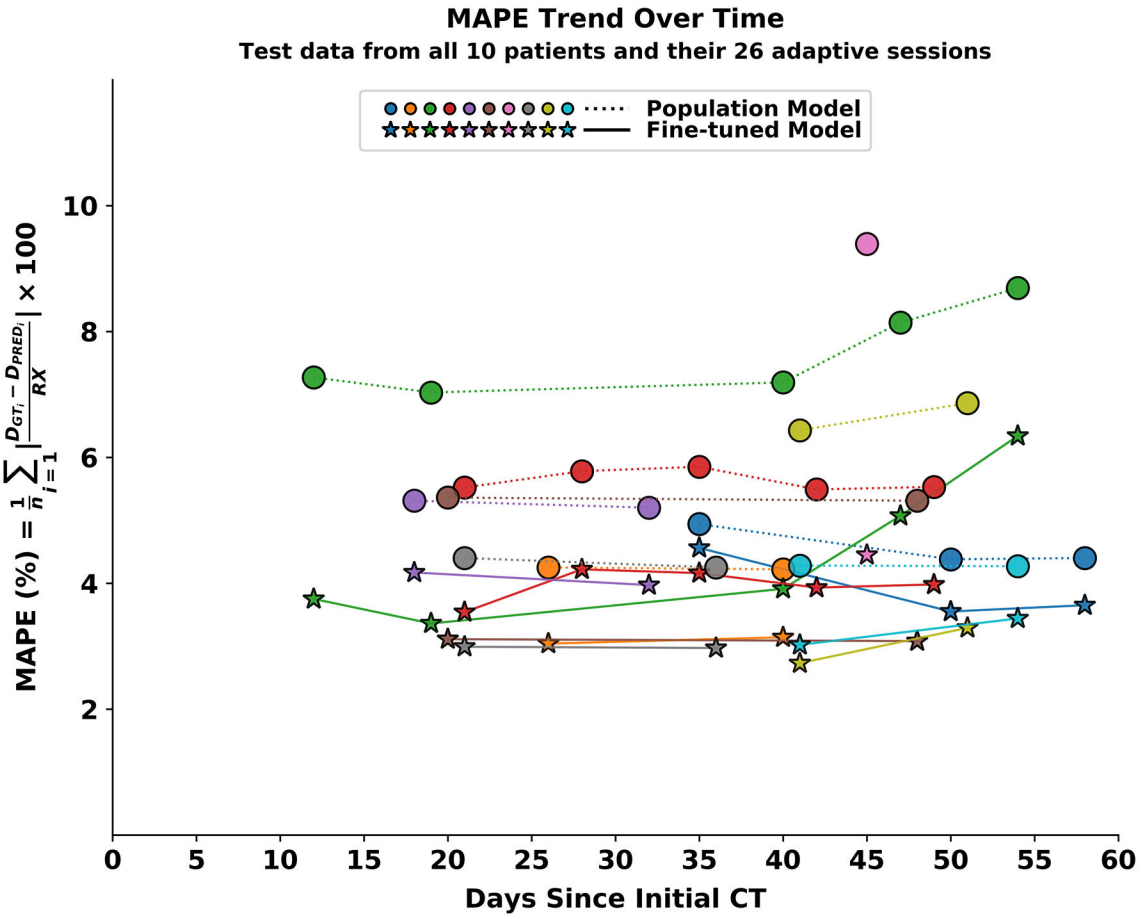

Figure #4:

MAPE was calculated using Equation (1) and plotted over time, with time in units of calendar days since the initial CT simulation. The color of each data point represents a unique patient, and data points are connected for patients with greater than 1 adaptive session. Data points shaped as a circle represent the performance of the population model for that specific adaptive session and may be connected by dotted lines. Data points shaped as a star represent the performance of the fine-tuned patient-specific model and may be connected by a solid line.

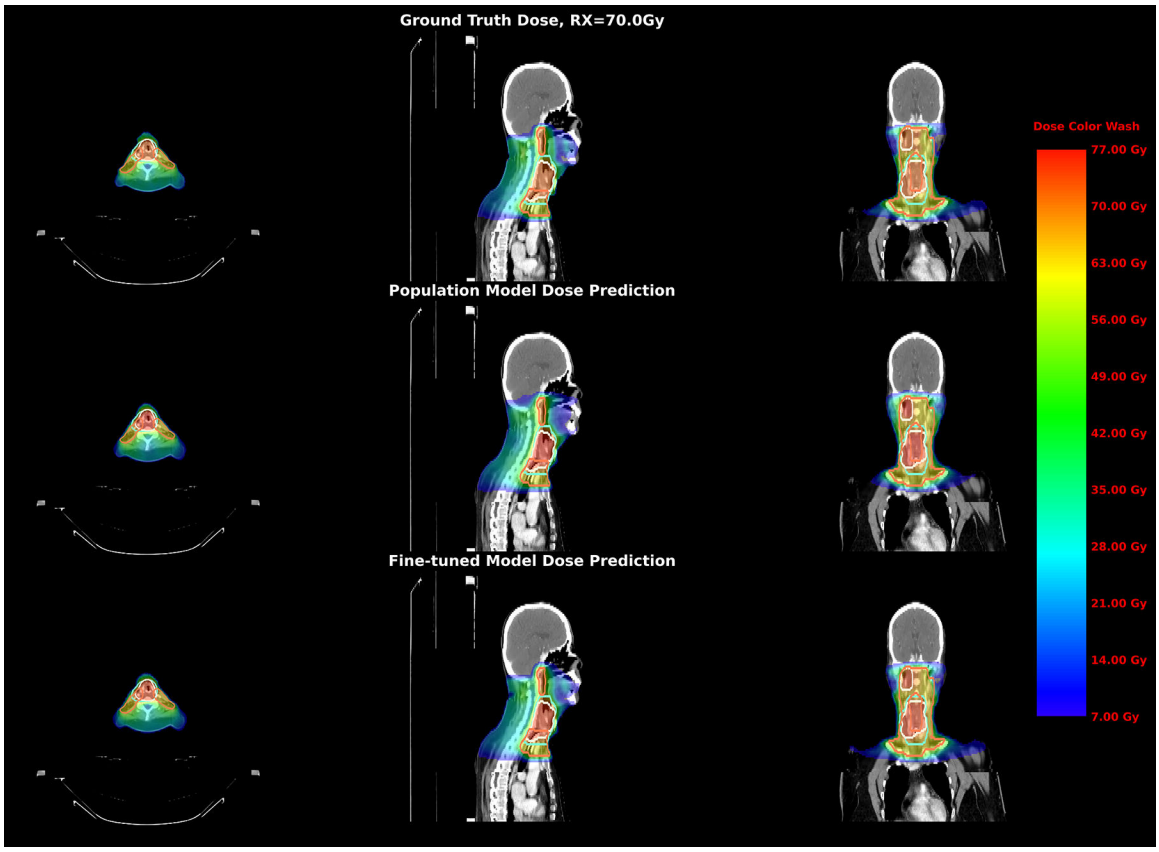

Figure #5:
Dose displays, shown for a session from an individual patient. The top panel displays the ground truth dose, while the middle and bottom panels display dose as predicted by the AP and FT-PS models respectively.

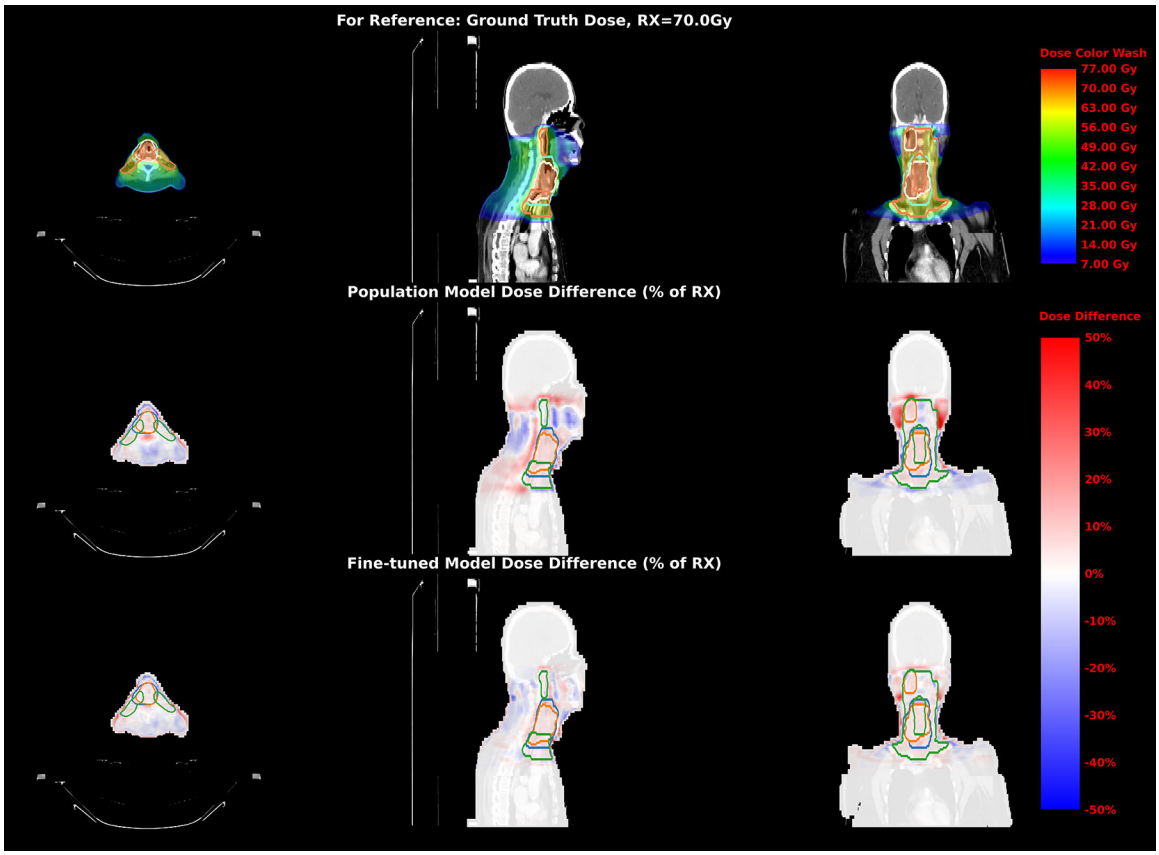

Figure #6:

Dose displays, shown for the same patient displayed in figure #5. The top panel displays the ground truth dose. The middle and bottom panel display dose differences in terms of percent error, where a positive value indicates over-prediction of dose and a negative value indicates under-prediction of dose.

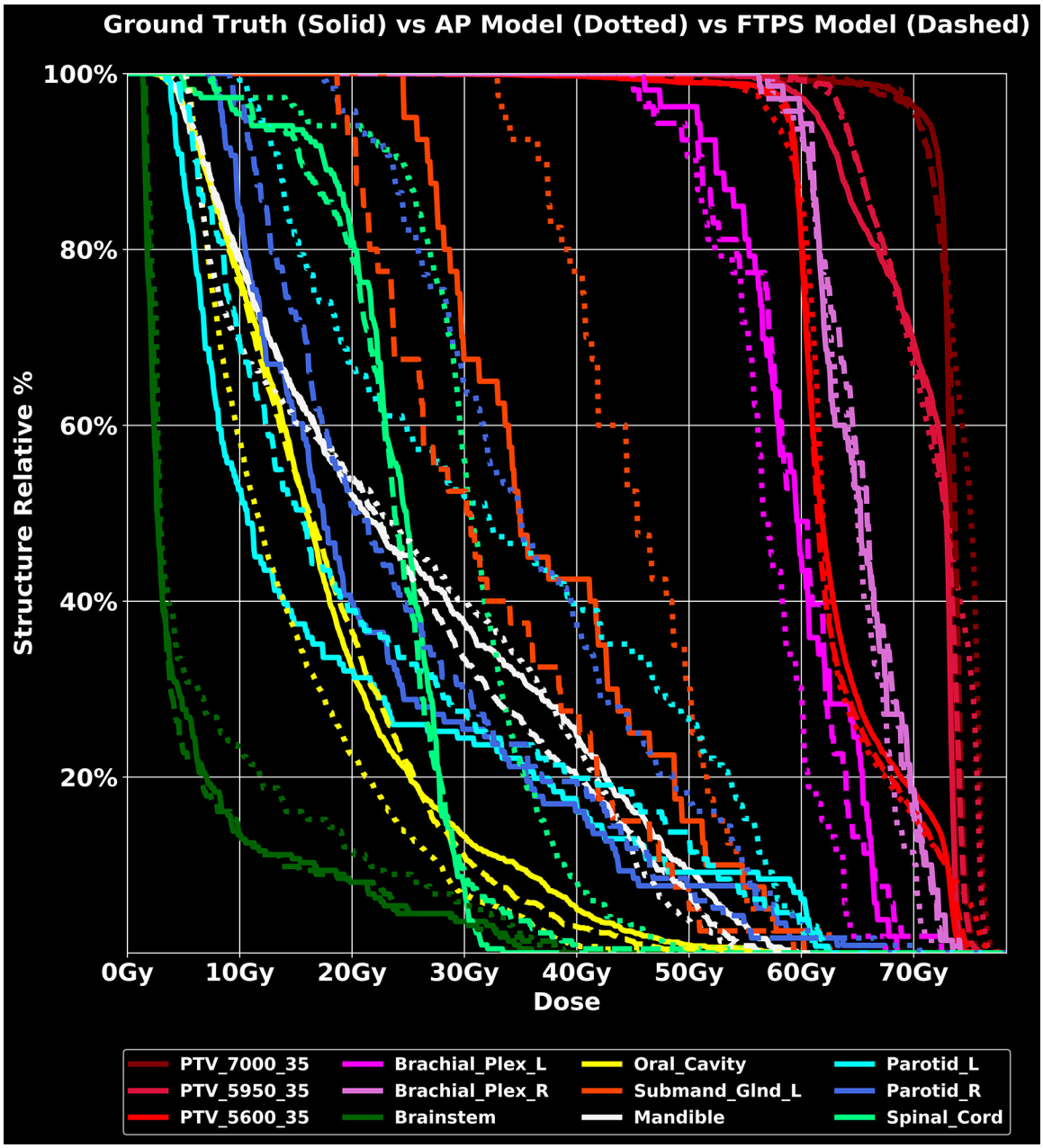

Figure #7:

The dose volume histogram for the patient displayed in figure #5 and figure #6. The solid lines display the ground truth dose delivered to each OAR, whereas dotted lines represent the AP model's prediction for each OAR and dashed lines represent the FTPS model's prediction for each OAR.

Table #1:

The dose delivered to the highest dose PTV in each treatment plan was compared in terms of percent deviance from the ground truth using the formula: $(\frac{Prediction}{Ground\ Truth}*100)$. A value of 100.0% indicates identical values for the dose prediction and the ground truth, whereas a lower value represents an under-prediction and a higher value represents an over-prediction.

Model Type	PTV D95%	PTV D98%	PTV D99%	PTV Homogeneity	PTV Conformity Index
AP Model	100.0 ± 0.0%	100.1 ± 3.2%	100.8 ± 4.4%	115.6 ± 25.5%	94.1 ± 22.9%
FTPS Model	100.0 ± 0.0%	100.6 ± 2.2%	101.4 ± 4.4%	109.6 ± 21.6%	102.7 ± 14.4%