



Published in final edited form as:

Radiother Oncol. 2020 December ; 153: 228–235. doi:10.1016/j.radonc.2020.10.027.

Dose Prediction with Deep Learning for Prostate Cancer Radiation Therapy: Model adaptation to Different Treatment Planning Practices

Roya Norouzi Kandalan¹, Dan Nguyen², Nima Hassan Rezaeian³, Ana M Barragán-Montero⁴, Sebastiaan Breedveld⁵, Kamesh Namuduri⁶, Steve Jiang⁷, Mu-Han Lin⁸

¹Medical Artificial Intelligence and Automation (MAIA) Laboratory, Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, USA; Department of Electrical Engineering, University of North Texas, Denton, USA. ²Medical Artificial Intelligence and Automation (MAIA) Laboratory, Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, USA. ³Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, USA. ⁴Molecular Imaging, Radiotherapy and Oncology (MIRO), UCLouvain, Brussels, Belgium. ⁵Department of Radiation Oncology, Erasmus University Medical Center - Cancer Institute, Rotterdam, The Netherlands. ⁶Department of Electrical Engineering, University of North Texas, Denton, USA. ⁷Medical Artificial Intelligence and Automation (MAIA) Laboratory, Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, USA. ⁸Medical Artificial Intelligence and Automation (MAIA) Laboratory, Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, USA.

Abstract

Purpose: This work aims to study the generalizability of a pre-developed deep learning (DL) dose prediction model for volumetric modulated arc therapy (VMAT) for prostate cancer and to adapt the model, via transfer learning with minimal input data, to three different internal treatment planning styles and one external institution planning style.

Methods: We built the source model with planning data from 108 patients previously treated with VMAT for prostate cancer. For the transfer learning, we selected patient cases planned with three different styles, 14–29 cases per style, in the same institution and 20 cases treated in a different institution to adapt the source model to four target models in total. We compared the dose distributions predicted by the source model and the target models with the corresponding clinical plan dose used for patient treatments and quantified the improvement in the prediction quality for the target models over the source model using the Dice similarity coefficients (DSC) of 0% to 100% isodose volumes and the dose-volume-histogram (DVH) parameters of the planning target volume and the organs-at-risk.

Results: The source model accurately predicts dose distributions for plans generated in the same source style, but performs sub-optimally for the three different internal and one external target styles, with the mean DSC ranging between 0.81–0.94 and 0.82–0.91 for the internal and the

* Co-corresponding authors Steve.Jiang@UTSouthwestern.edu (Steve Jiang), Mu-Han.Lin@UTSouthwestern.edu (Mu-Han Lin).

external styles, respectively. With transfer learning, the target model predictions improved the mean DSC to 0.88–0.95 and 0.92–0.96 for the internal and the external styles, respectively. Target model predictions significantly improved the accuracy of the DVH parameter predictions to within 1.6%.

Conclusion: We demonstrated the problem of model generalizability for DL-based dose prediction and the feasibility of using transfer learning to solve this problem. With 14–29 cases per style, we successfully adapted the source model into several different practice styles. This indicates a realistic way forward to widespread clinical implementation of DL-based dose prediction.

1. Introduction

Radiation therapy treatment planning is a complex process, as the target dose prescription and normal tissue sparing are conflicting objectives. The lowest achievable dose for each individual organ-at-risk (OAR) is unknown a priori. Multiple iterations between the planner and the physician may be required to reach the optimal balance between target coverage and OAR sparing. Recent planning tool developments have focused on improving the efficiency of the planning iteration process. Interactive multi-criteria optimization (MCO) allows the planner or the physician to explore the trade offs among the possible solutions and eliminates the communication lag in the plan iteration process [1–5]. Knowledge-based planning (KBP) takes a data-driven approach to learn from the past high-quality clinical plans and predicts a patient’s specific dose-volume histogram (DVH) and OAR dose constraints for new patients by using the relationship between geometric and dosimetric information derived from historical patient plans in the library [6–13].

Advancements in the field of artificial intelligence (AI), especially in data-driven machine learning (ML) and deep learning (DL) algorithms for many challenging computer vision problems, have inspired many researchers in radiation oncology. Deep learning and, more specifically, convolutional neural network (CNN) architectures have significantly improved imaging and vision tasks. In particular, UNet [14], a complex architecture initially designed for biomedical image segmentation, has notably improved the performance of predicting the radiation dose distribution in the body without going through a real planning process [15–18]. The predicted dose distribution provides both visual input and DVH metrics to assist physician’s trade-off decision-making up front to provide more achievable planning objectives prior to the planning process.

While DL models holds promise as accurate predictors of the expected dose distribution prior to planning, the heterogeneity of data is a common challenge for AI modeling. Despite nationally-accepted practice guidelines that set the baseline of patient care, actual clinical practice is rarely clearly defined in black and white. Patient-specific medical reasons such as hip prosthesis and previous radiation can lead to deviation from the guideline. Strategies for solving a specific dosimetric trade-off problem will vary among different physicians and planners [19]. Treatment planning systems and optimization algorithms also introduce variations into clinical practice. Plans generated by different practice styles that meet the national guidelines, in terms of plan quality, can end up with different spatial dose

distributions. A DL dose prediction model built based on a particular dataset from one institution may not work well for different treatment planning styles even within the same institution. Therefore, the ability to adapt a pre-built DL dose prediction model to a given planning/practice style is desirable.

Furthermore, expertise in AI modeling is not widely accessible in various clinical settings in the field of radiation oncology. Although some AI solutions are commercially available and the vendor provides AI expertise, building an initial model still requires a clean and large dataset of treated patient cases. This requires a tremendous effort from the user to collect and curate enough data for modeling. Sharing models between institutions could alleviate the challenges associated with building an initial AI model, but the heterogeneity in practice mentioned above would still yield unsatisfactory performance. Thus, the ability to easily adapt a pre-built dose prediction model to a different practice style could lead to practical clinical implementation of DL dose prediction models in real-world clinical settings.

The goals of this work are to investigate the problem of generalizing DL-based dose prediction models and to utilize transfer learning to adapt a DL prostate VMAT dose prediction model to various planning/practice styles with minimal data from each individual style. We built a source model based on the 108 patients treated with VMAT for prostate cancer in a large institution. This work tests two types of practice heterogeneity: first, different planning styles in the same institution where we built the source model; and second, planning practices at a different institution. The source model was adapted to three internal planning styles and one external planning style with 14–29 training cases for each model. To our knowledge, this is the first work to study the DL-based dose prediction model generalizability problem and to utilize transfer learning to provide a practical solution.

2. Methods

2.1. Data Groups

For this study, we selected a total of 248 cases of prostate cancer treated with the VMAT technique. 188 cases from one institution, planned with the Eclipse treatment planning system (TPS) utilizing the progressive resolution optimizer for VMAT optimization and delivered with both Varian TrueBeam and Elekta Versa HD linear accelerators; and the 60 cases from an external institution, planned with the Erasmus-iCycle TPS [20] a system for automated multi-criteria treatment planning, and delivered with Elekta Versa HD linear accelerators. Table 1 summarizes the datasets and shows the dose distributions of individual styles. Supplement document 1 illustrates the inherent plan quality variations among different training datasets.

The source model was trained with the "Source" dataset, which consists of 108 plans in the 'conformal' dose style, which is the most representative style of planning in our institution. In the same institution, we found three additional planning styles for different clinical indications, represented by three target datasets: "Internal-A," "Internal-B," and "Internal-C". Internal-A style is slightly more aggressive approach than the Source style featuring lower femur head doses while keeping similar bladder and rectum sparing as the Source style. Internal-B style trades the femur head doses and bladder dose for lower rectum

sparing. The Internal-C style is a more extreme approach that allows a higher dose to femur heads to spare both the bladder and rectum. We use these three target datasets to demonstrate the problem of model generalizability and to test the feasibility of using a small dataset for transfer learning by using 14–29 cases from each target dataset. We use 2–5 test cases to evaluate the target model's performance.

The "External" style represents plans from a different institution. As one can see the dose distribution of the External style has many intermediate dose spikes in both the lateral and the anterior directions, which resulted from different TPS and planning approach. Given the large differences between the External and the Source styles, we trained the target model with 20 cases to examine the concept of small training dataset for the transfer learning and thoroughly evaluated the model performance with 40 cases.

2.2. Network

In this work, we implemented a 3D UNet for the model architecture, [14, 21], as UNet has been extensively used in radiation therapy for dose prediction [15–18]. The inputs to this architecture are the contours of the planning target volume (PTV) and the OARs, including bladder, rectum, left and right femoral heads, and body. We use a rule-based method to automatically rename the structure to AAPM TG-263 nomenclature and follow by manual quality check process to pick up and fix error of matching. After the cleaning, the structure regions of interest (ROIs) are segmented into binary structure masks to indicate what structures each voxel belongs to. The data are resampled to have voxel spacing of $[5\text{mm} \times 5\text{mm} \times 5\text{mm}]$ across all patients, and then a $[120 \times 120 \times 64]$ volume was cropped to generate the base training data. Due to the memory limits of a 16 GB GPU (NVIDIA K80), we implemented patch-based training, with a patch size of $[96 \times 96 \times 48]$. At each training iteration, a random patient and plan is chosen, and then a $[96 \times 96 \times 48]$ volumetric patch is randomly selected from the data volume based on a Gaussian sampling scheme proposed by [22]. This random patch selection process is done on-the-fly during training, so resampled $[120 \times 120 \times 64]$ data is stored on the hard drive. The location of the patch is the same for input (anatomy) and output (dose) data for that given iteration, in order to correctly correlate the dose to the anatomy. We also implemented group normalization [23] after each set of convolution and rectified linear unit (ReLU) operations, which helps accelerate the convergence rate in the network. This model is trained to learn the mapping between the binary masks of the OARs and the clinical radiation dose distribution in the body. Mean Square Error (MSE) was the loss function, and the Adam optimizer with a learning rate of 10^{-4} was used to minimize the MSE. The final network consists of 85 layers with 7,870,177 trainable parameters.

2.3. Transfer learning

For both the internal Source dataset and the external institution dataset, the transfer learning method implemented for style transfer involves 3 components: 1) freezing the weights of the first half of the model, 2) training only the second half of the model, and 3) randomly reinitializing the weights of the very last layer of the model. Assuming that the major variations in dose distributions are owing to different planning styles across institutions, we can effectively freeze the first half of the UNet, which finds features from just the anatomy.

We can then train just the second half of the network to learn the dose distributions of a different planning style. Finally, to assist the network in converging and to prevent it from falling into a local minimum, we reinitialize the weights in the final layer of the network. The hyperparameters remain unchanged, except for the learning rate, which was reduced to one-tenth of the original value.

2.4. Model Evaluation

We first applied the Source model to the Source test cases to evaluate the model's performance on its own planning style. We then applied the Source model to the test cases of the three internal and one external target styles to investigate the model's performance on other planning styles and to set the baseline prediction quality before transfer learning. Individual target models trained via transfer learning were each applied to the test cases of their own planning style to investigate the prediction quality after adapting the Source model to the given target style. To further evaluate the improvement achieved by the transfer learning, we cross-compared each of the doses predicted by the Source and target models with the respective clinical plan dose.

The model was trained to predict the doses with their correct dose prescription values directly. We believe any kind of additional normalization would be hiding the true performance of the models, and potentially compromise any learned dose tradeoffs in the OARs. Comparing the direct predictions is the fairest assessment and comparison of the model performance. As far as clinical implementation, the clinical user can normalize the results to better suit their purpose.

We compared the absolute doses predicted by the DL models to the clinical plan doses without further normalization. We evaluated the spatial dose distributions and the dose-volume histograms (DVH). We calculated DVH metrics including structure mean and maximum doses (i.e. D2), PTV D98, and PTV D95. The clinical plan dose served as the baseline for the dose comparison. We calculated the differences in the DVH metrics between the clinical plan dose and the predicted doses to quantify the DVH agreement. Gamma evaluation with 3%/3mm criteria was applied to the pixels with higher than 10% prescribed dose to quantify the agreement between the clinical plan dose and the predicted dose for both source and target models. In addition, we calculated the dice similarity coefficients (DSC), $\frac{2(A \cap B)}{(A + B)}$, from 0% to 100% of the prescribed dose to quantify the agreement of the spatial dose distribution between the clinical plan dose and the predicted doses. The Dice score examines the overlap as well as the volume similarity between the clinical plan dose and the predicted dose. We used a paired t-test to calculate the statistical significance of the results for the External style, which had enough test cases for a meaningful test.

3. Results

Figure 1 shows a typical example of the dose agreement between the clinical plan and the Source model prediction for the Source style test cases. Upon visually inspecting the dose distribution, percent dose difference, and the DVH, one can see that the Source model predicts the PTV dose correctly with minimal differences in the OAR doses. The violin plots

present the differences in the PTV and the OAR DVH metrics between the clinical plan and the Source model prediction. The differences in the PTV DVH metrics are mostly within 3%, as reflected by the width of the individual data clusters. The median of the PTV dose differences are all within 2%. Similarly, the median of the mean and maximum OAR dose differences are within 1% and 2%, respectively.

Figure 2 illustrates the Source model and target model predictions on the test cases of individual target styles. The improvement in the model performance after transfer learning can be seen in the dose comparisons among the clinical plan, the Source model prediction and the target model prediction. The Source model fails to predict the style-specific dose distribution features. However, the target models have learned those dose distribution features via transfer learning, so they predict a distribution similar to the clinical plan and exhibit a better agreement in the DVH, especially for the Internal-C and External styles. The gamma evaluation distribution is shown in the supplement document 2. The source model on the source data achieves a gamma passing rate (mean \pm s.d.) of 0.71 ± 0.08 . On the internal target data, the source model achieves a gamma passing rate of 0.52 ± 0.10 , while the target model achieves an improved rate of 0.61 ± 0.10 . On the external data we have also found that the source and target model had achieved passing rates of 0.47 ± 0.05 and 0.84 ± 0.09 , respectively. Using a two-sided t-test, the improved gamma passing rates for the target model over the source model for both the internal data and external data were found to be statistically significant ($p < 0.01$).

We compared each of the Source and target model predictions to the respective clinical plan dose and calculated the DSC of individual isodose volumes accordingly. The DSC values increased with the dose agreement with a coefficient of 1 considered to be a perfect match. As presented in Figure 3, the target models improved upon the Source model in term of DSC regardless of the planning style. However, the Internal and the External styles exhibit different trends. For the Internal styles, both the Source and the target models predict the high dose bath (70% to 100%) fairly well. The main improvements appear in the intermediate dose bath (60% to 30%). The mean DSC ranges between 0.81–0.94 for the three Source models and between 0.82–0.91 for the corresponding target models. The largest improvement may be seen in the Internal-C style, as the target model achieves a 7% improvement over the Source model (mean DSC is 0.80 for the Source model and 0.87 for the Internal-C target model) transfer learning. For the External style, the target model improves significantly upon the Source model throughout all isodose volumes ($p < 0.05$), with the improvement increasing as the isodose volume decreases. An average of 5% and 8% mean DSC improvements were achieved in the high and the intermediate dose volumes, respectively. A systematic 10% improvement was seen in the low dose volume.

Figure 4 illustrates the Source and the External target model performances on predicting the DVH metric of the External style test cases. Overall, as indicated by the length of the data clusters, the Source model demonstrates a large variation in the prediction quality, while the External target model demonstrates more consistent and accurate predictions. The Source model overestimates the PTV doses, especially for the Dmean and the D2. The External target model significantly improved the predictions and resulted in mean differences within 1.6% (Table 2) comparing to the treated plan approved by the medical doctor. In terms of the

OAR dose predictions, the target model achieved an agreement within 1.5%. The target model also results in satisfactory agreement in the clinical relevant DVH metric, V75 and V65 of bladder and rectum. In summary, the External target model improved upon the Source model by up to 6.4% and 6.0% in the PTV and the OAR DVH predictions, respectively ($p < 0.05$).

4. Discussion and Conclusion

There has been growing attention on leveraging AI-based decision support tools (DST) to improve the treatment planning process. Incorporating AI DST as a part of the clinical pathway [1] could standardize evidence-based practices to ensure high quality and cost-effective medical care. AI algorithms, including the DL dose prediction model in this work, require large repositories of high quality data. This becomes one of the barriers to widespread deployment of AI-based solutions in the radiation oncology field. In this study, we utilized transfer learning to solve the data size problem and demonstrated the ability to adapt a source model to three different internal planning styles and one external planning style with minimal data input.

Heterogeneity in clinical practices, whether intra- or inter-institutional, is commonly seen. The clinical protocols and guidelines set the floor of treatment plan acceptance, but users set the ceiling for their own practice. Table 1 clearly shows that, even within the same institution, the individual trade-off preferences of different physicians and planners in combination result in a variety of dose distribution styles. On top of these differences in practice, the different treatment planning system employed in the External institution in this study yielded dramatically different dose distributions from the Source style. The treatment plans in the Source style were optimized with a specific objective function (normal tissue objective, NTO) to shape the dose fall-off conformally, whereas plans from the External institution were optimized with a unique 23-beam starting condition of the VMAT optimization that led to distinct spikes in the low dose bath. Regardless of the dose distribution style, they are all clinically accepted plans that comply with the clinical guideline. The AI model's adaptation to the user's practice style can achieve standardization with respect to the user's prior experience. This would be practical and beneficial for implementing AI-guided planning in practice, because the user can precisely interpret the predicted dose for meaningful clinical use.

Directly deploying a model built in a different practice style may lead to unsatisfactory predictions. As shown in the Source model predictions on the cases planned with target styles in Figure 2, the predicted doses inherit the conformal dose distribution of the Source style. They meet the planning objectives, but they fail to represent the dose distribution features of the target styles. This reflects a common frustration of institutions that are trying to clinically implement a model provided by a vendor or developed by a different institution. In this work, we demonstrated the transfer learning can adapt a source model into various practice styles.

Transfer learning with an additional 14–29 cases in the target style, allowed the target models to learn the features of the new planning style quickly. Taking the External style as

an example, the distinct low dose spikes were precisely predicted by the External target model. We saw this improvement globally in all target styles, and it increased as the variation between the Source and target styles increased (Figure 2). For example, the Internal-A and Internal-B styles pull the dose from the rectum only slightly more aggressively than the Source style, so the inherent differences are small. Therefore, the Source model still achieved more than a 0.78 mean DSC, which the target models only slightly outperformed. In contrast, the Internal-C style turns off the NTO and intentionally trades femur doses for lower rectum and bladder doses, so it differs more substantially from the Source style. Accordingly, the Source model predictions on the Internal-C style cases demonstrated a significant dip in DSC in the intermediate dose levels, but the Internal-C target model improved the DSC to a satisfactory level. For the External style, which is fundamentally different from the Source style, the External target model improved upon the Source model's DSC by up to 10% DSC and had a DSC higher than 0.87 among all test cases and isodose volumes.

The intermediate dose bath of the VMAT plan is more unpredictable because of its complex nature. The variations in trade-offs and planning approaches can tremendously polarize the spatial dose distribution in this dose range. Moreover, with the large numbers of beam angles in VMAT optimization, the optimizer has more options to pull the dose off a specific area, which makes the dose more uncertain to predict. Traditionally, the DL dose predictor has a hard time handling these style-specific features. This is reflected in figure 3, which shows that the DSC of the Source model prediction ranges between 0.80–0.94. However, with transfer learning, the target models can represent these dose features, and the DSC of target model prediction increased to 0.87–0.99, which is satisfactory.

The input data to the model are the structure contours of the PTV, bladder, rectum, femoral heads and body. The dose prediction is dependent on these structures, and the dose distribution may change with changes in the segmentation. However, assuming that the same set of contours that are used for planning are also used as input to the model (even if they are “poor” quality segmentation), we do not expect any larger prediction errors against a ground truth clinical dose, than what was reported in the study.

It is worth to mention that it is not guaranteed the model will produce the “best” possible dose distribution since the clinical training data is not guaranteed to be “optimal” either. However, since the clinical doses that were used for training the models have been approved by the physicians, the trained models can produce dose distributions that are meeting or close to meeting the acceptability criteria outlined by the physicians. This can provide a lot of benefits as well, which include improving planning efficiency and reducing plan quality variation. The doses predicted in the paper do not guarantee machine deliverability. However, we should note that the three internal styles and the external data that we have are all machine deliverable plans. Therefore, the model is expected to be able to predict overall realistic dose tradeoffs. The model is not striving to create a “perfect plan” but instead, create a “physician acceptable” plan.

There are several potential applications. The current product can be used as a decision support tool for the MDs or a driving guidance tool for the treatment planners, who then

create the deliverable plan. Second application is MD decision support between different treatment modalities, such as 3D versus IMRT. Third application is efficient plan quality check for peer-review. In addition, the dose prediction can be used for education and training for residents and students. Our institution implemented the dose prediction tool as the physician's planning directive assist. Physician triggers the prediction upon the time of contour completion. Physician will be presented the predicted dose distribution, the predicted dose-volume histogram, and a predicted objective metrics. Physician will review the predicted objective and make adjustment as needed (ex. prior radiation) before submitting the directive to the medical dosimetrist.

In light of the barriers to accessing the expertise of AI modeling, collecting sufficient data to train a model, and sharing data among institutions, a broad AI tool that can easily adapt a maturely built model to different practice styles could allow more practices to access and clinically implement an AI model for dose prediction. We used prostate VMAT as an example to demonstrate the feasibility of leveraging transfer learning for model adaptation. However, this methodology can be employed for other AI tools in the field of radiation oncology and allow for widespread application of such tools. As future work, we plan to develop an approximate Bayesian method in order to have the deep learning model produce uncertainties within its voxel-wise prediction. This uncertainty can be converted to confidence intervals in other dose-related metrics, such as the DVH. This can be used to flag the users on the reliability or achievability of the prediction. In addition, we plan to investigate the addition of this uncertainty measure to the uncertainty of anatomy and tradeoff options. For example, the model can create prediction based on base model and various tradeoff styles, to provide the options with uncertainties for the MDs to make their best clinical judgment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We would like to thank the National Institutes of Health (NIH) for supporting this study through a research grant (R01CA237269) and Dr. Jonathan Feinberg for editing the manuscript.

References

- [1]. Craft David L., Hong Theodore S, Shih Helen A, and Bortfeld Thomas R. Improved planning time and plan quality through multicriteria optimization for intensity-modulated radiotherapy. *International Journal of Radiation Oncology * Biology * Physics*, 82(1):e83–e90, 1 2012.
- [2]. Monz M, Küfer Karl-Heinz, Bortfeld T, and Thieke Christian. Pareto navigation - algorithmic foundation of interactive multi-criteria imrt planning. *Physics in medicine and biology*, 53:985–98, 03 2008. [PubMed: 18263953]
- [3]. Craft David L, Halabi Tarek F, Shih Helen A, and Bortfeld Thomas R. Approximating convex pareto surfaces in multiobjective radiotherapy planning. *Medical Physics*, 33(9), 9 2006.
- [4]. Haveren Rens, Heijmen Ben J. M., and Breedveld Sebastiaan. Automatic configuration of the reference point method for fully automated multi-objective treatment planning applied to oropharyngeal cancer. *Medical Physics*, 2 2020.

- [5]. Breedveld Sebastiaan, Craft David, van Haveren Rens, and Heijmen Ben. Multi-criteria optimization and decision-making in radiotherapy. *European Journal of Operational Research*, 277(1):1–19, 2019.
- [6]. Wu Binbin, Ricchetti Francesco, Sanguineti Giuseppe, Kazhdan Misha, Simari Patricio, Chuang Ming, Taylor Russell, Jacques Robert, and McNutt Todd. Patient geometry-driven information retrieval for imrt treatment plan quality control. *Medical Physics*, 36(12):5497–5505, 2009. [PubMed: 20095262]
- [7]. Wu Binbin, Ricchetti Francesco, Sanguineti Giuseppe, Kazhdan Michael, Simari Patricio, Jacques Robert, Taylor Russell, and McNutt Todd. Data-driven approach to generating achievable dose–volume histogram objectives in intensity-modulated radiotherapy planning. *International Journal of Radiation Oncology Biology Physics*, 79(4):1241–1247, 3 2011.
- [8]. Yuan Lulin, Ge Yaorong, Lee W. Robert, Yin Fang Fang, Kirkpatrick John P, and Wu Q. Jackie. Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in imrt plans. *Medical Physics*, 39(11):6868–6878, 2012. [PubMed: 23127079]
- [9]. Appenzoller Lindsey M, Michalski Jeff M, Thorstad Wade L, Mutic Sasa, and Moore Kevin L. Predicting dose-volume histograms for organs-at-risk in imrt planning. *Medical Physics*, 39(12):7446–7461, 2012. [PubMed: 23231294]
- [10]. Wu Binbin, Pang Dalong, Lei Siyuan, Gatti John, Tong Michael, McNutt Todd, Kole Thomas, Dritschilo Anatoly, and Collins Sean. Improved robotic stereotactic body radiation therapy plan quality and planning efficacy for organ-confined prostate cancer utilizing overlap-volume histogram-driven planning methodology. *Radiotherapy and oncology: journal of the European Society for Therapeutic Radiology and Oncology*, 112(2):221–6, 2014. [PubMed: 25108808]
- [11]. Moore Kevin L, Brame Scott R, Low Daniel A, and Mutic Sasa. Experience-based quality control of clinical intensity-modulated radiotherapy planning. *International Journal of Radiation Oncology * Biology * Physics*, 81(2):545–551, 10 2011.
- [12]. Shiraishi Satomi and Moore Kevin L.. Knowledge-based prediction of three-dimensional dose distributions for external beam radiotherapy. *Medical Physics*, 43(1):378–387, 2016. [PubMed: 26745931]
- [13]. Wu Binbin, Pang Dalong, Simari Patricio, Taylor Russell, Sanguineti Giuseppe, and McNutt Todd. Using overlap volume histogram and imrt plan data to guide and automate vmat planning: A head-and-neck case study. *Medical Physics*, 40(2):021714, 2013. [PubMed: 23387737]
- [14]. Ronneberger Olaf, Fischer Philipp, and Brox Thomas. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, page 234–241, 2015.
- [15]. Nguyen Dan, Jia Xun, Sher David, Lin Mu-Han, Iqbal Zohaib, Liu Hui, and Jiang Steve. 3d radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected u-net deep learning architecture. *Physics in Medicine & Biology*, 64(6):065020, 3 2019. [PubMed: 30703760]
- [16]. Willems Siri, Crijns Wouter, Sterpin Edmond, Haustermans Karin, and Maes Frederik. Feasibility of ct-only 3d dose prediction for vmat prostate plans using deep learning. In Nguyen Dan, Xing Lei, and Jiang Steve, editors, *Artificial Intelligence in Radiation Therapy*, pages 10–17, Cham, 2019. Springer International Publishing.
- [17]. Nguyen Dan, Long Troy, Jia Xun, Lu Weiguo, Gu Xuejun, Iqbal Zohaib, and Jiang Steve. A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning. *Scientific Reports*, 9(1):1076, 2019. [PubMed: 30705354]
- [18]. Barragán-Montero Ana María, Nguyen Dan, Lu Weiguo, Lin Mu-Han, Norouzi-Kandalan Roya, Geets Xavier, Sterpin Edmond, and Jiang Steve. Three-dimensional dose prediction for lung imrt patients with deep neural networks: robust learning from heterogeneous beam configurations. *Medical Physics*, 46(8):3679–3691, 2019. [PubMed: 31102554]
- [19]. Chan J, Chen J, Valdes G, Yom SS, Pattison A, Carpenter C, and Solberg TD. The impact of dataset size on an artificial intelligence-guided clinical decision support system for radiation therapy planning. *International Journal of Radiation Oncology • Biology • Physics*, 102(3):e535, 11 2018.

- [20]. Haveren Rens van, Heijmen Ben J M, and Breedveld Sebastiaan. Automatically configuring the reference point method for automated multi-objective treatment planning. *Physics in Medicine & Biology*, 64(3):035002, 1 2019. [PubMed: 30566906]
- [21]. Çiçek Özgün, Abdulkadir A, Lienkamp S, Brox T, and Ronneberger O. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In *MICCAI*, 2016.
- [22]. Nguyen Dan, O'Connor Daniel Dan Ruan, and Sheng Ke. Deterministic direct aperture optimization using multiphase piecewise constant segmentation. *Medical physics*, 44(11):5596–5609, 11 2017. PMC5689109[pmcid]. [PubMed: 28834556]
- [23]. Wu Yuxin and He Kaiming. Group normalization. In *The European Conference on Computer Vision (ECCV)*, 9 2018.

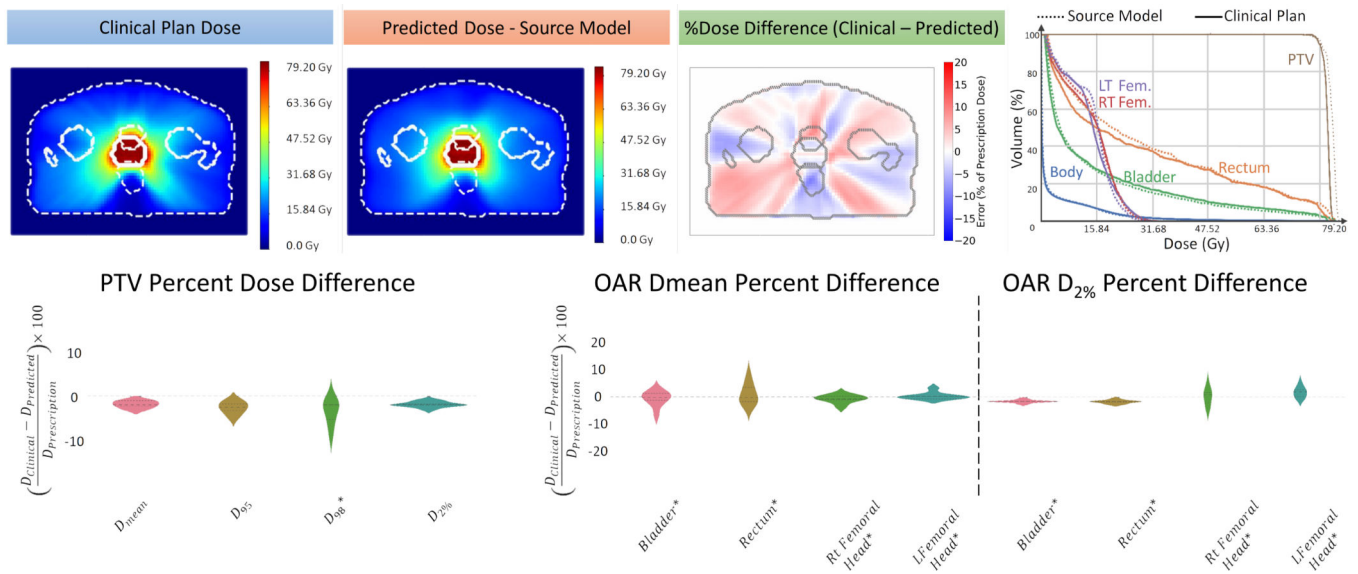


Figure 1: Performance analysis of the source model including the dose distributions with the structure contours outlined in white dotted lines; the percentage dose difference; the DVH of the clinical plan (solid lines) and the predicted dose (dotted lines) and the violin plots of the target and the OAR DVH metrics.

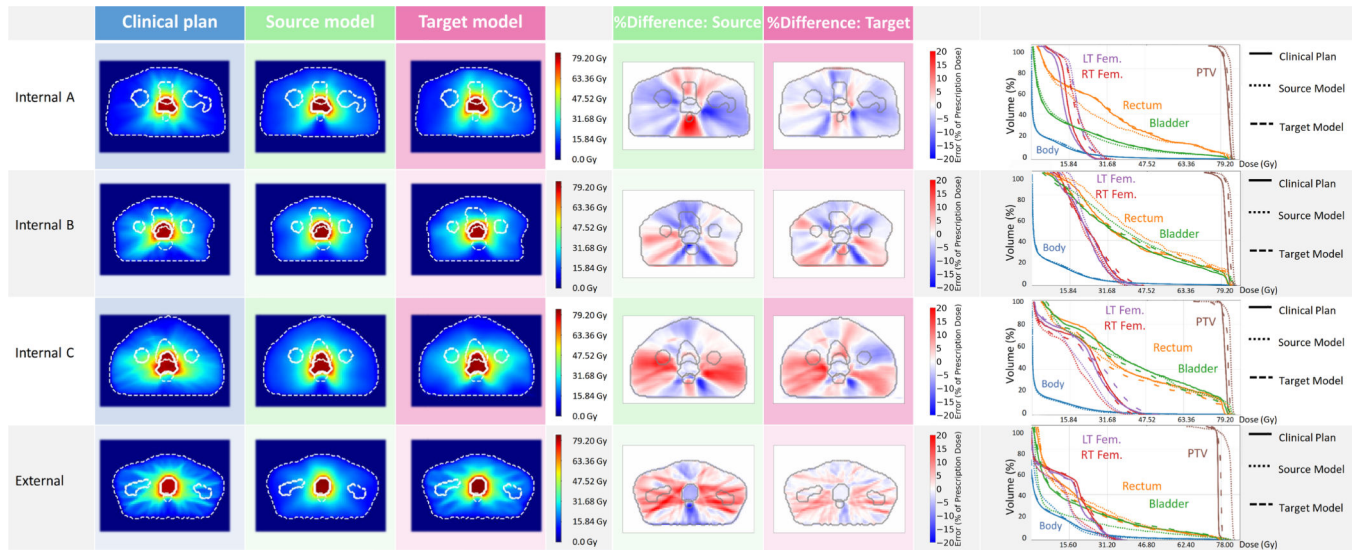


Figure 2:
Dose comparison between the clinical plan and the predicted plan using the source and the target models including the dose distributions, the percentage dose difference, and the DVH comparisons.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

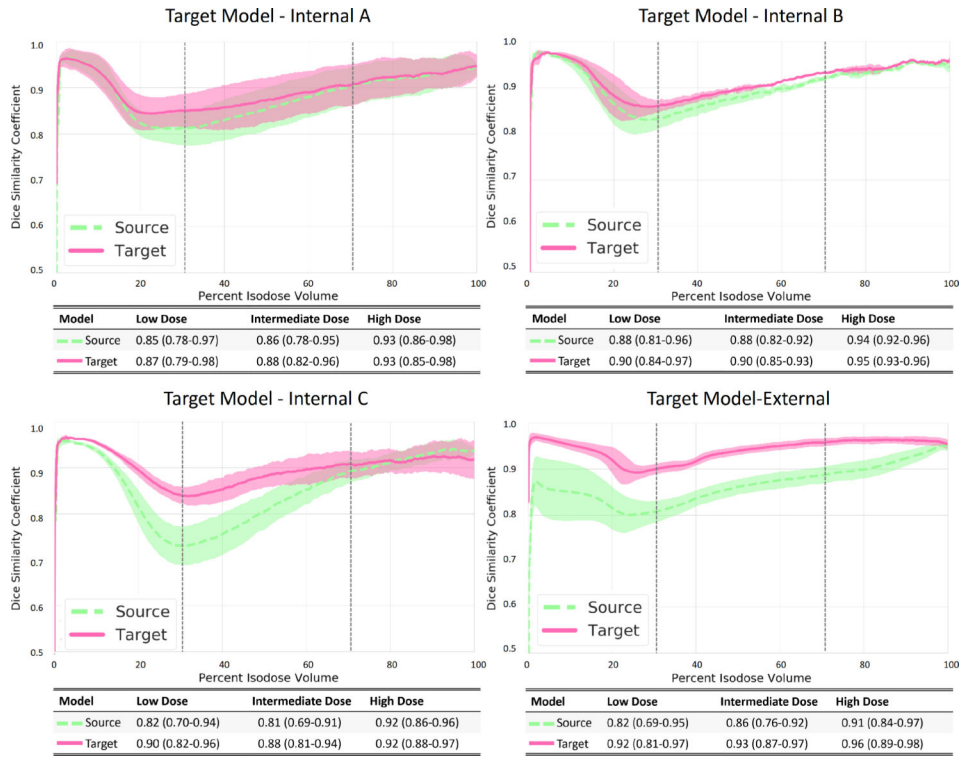


Figure 3: Dice similarity coefficients analysis, $\left(\frac{2(A \cap B)}{A + B}\right)$. Comparing the isodose volumes between the clinically planned dose and the predicted dose. One standard deviation represented the error in the graphs. The 30%-70% range is indicated in the figure by vertical dotted lines.

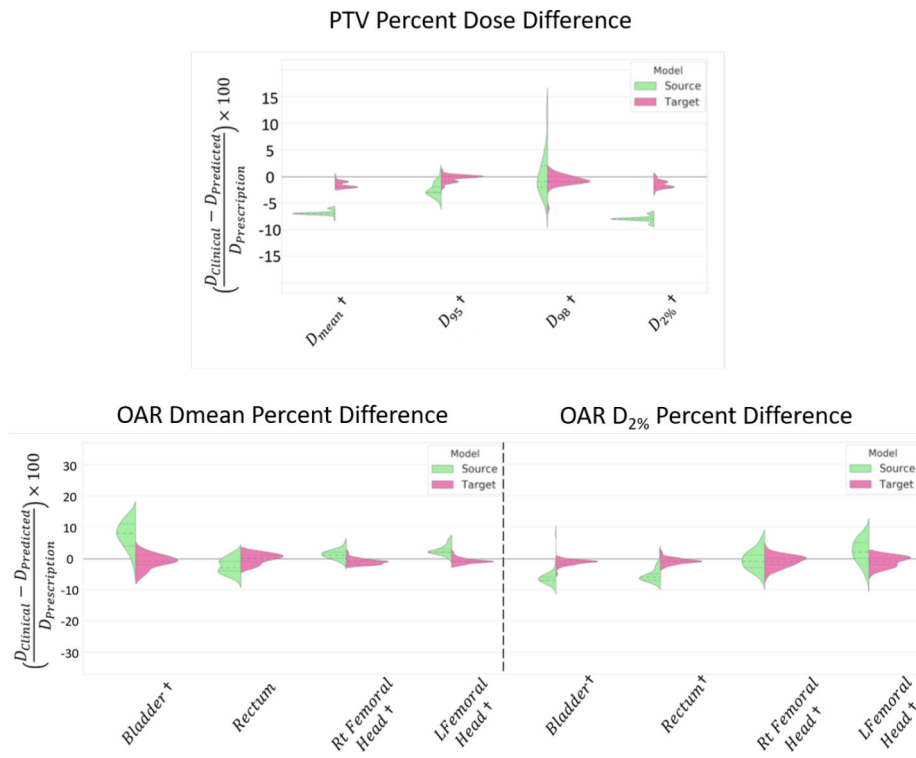


Figure 4: Quantitative performance evaluation of the error in the DVH metric in target model dose prediction vs the source model dose prediction in respect to the clinically planned dose for the external data.

† Represents $p < 0.05$

Table 1:

Patient datasets used for building the source model (Source) and for testing the model generalizability and then for adapting model (Internal-A, Internal-B, Internal-C, and External). First column shows a typical dose distribution of the planning style represented by each dataset.

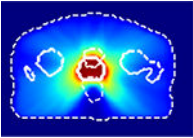
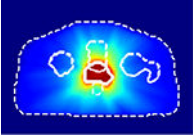
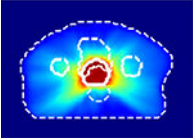
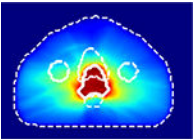
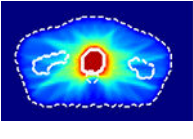
<u>Dose Style</u>	<u>Name</u>	<u>Dataset</u>	<u>Training</u>	<u>Testing</u>
	Source	118	108	10
	Internal-A	34	29	5
	Internal-B	16	14	2
	Internal-C	20	17	3
	External	60	20	40

Table 2:

Quantitative performance evaluation based on the error in target model dose prediction vs the source model dose prediction for the 'External' style. The comparison was made between the clinically approved plans and the predicted dose from the source and the target models, respectively, and normalized to the prescription dose.

ROI	DVH Metric	Source Model	Target Model
PTV	Dmean	%6.90±0.40	%1.60±0.55†
	D95	%2.60±1.15	%0.40±0.60†
	D98	%2.80±2.65	%0.90±1.00†
	Dmax	%8.00±0.50	%1.55±0.70†
Bladder	Dmean	%7.75±3.80	%1.80±1.60†
	Dmax	%6.55±1.25	%1.35±0.95†
	V75	%9.94±8.18	0.9±1.04†
	V65	%13.55±7.91	1.09±1.18†
Rectum	Dmean	%2.80±1.95	%1.25±0.95
	Dmax	%5.55±1.80	%1.10±0.80†
	V75	%5.14±3.23	0.48±0.51†
	V65	%5.7±3.37	0.78±0.71†
Rt Femur Head	Dmean	%1.55±1.20	%1.25±0.70†
	Dmax	%2.55±1.85	%1.55±1.45†
Lt Femur Head	Dmean	%2.60±1.45	%1.05±0.65†
	Dmax	%3.65±2.50	%1.50±1.35†