

Dosimetry Automation via Dose Mimicking

Contents

1	DVH Guided Deep Dose	2
1.1	Introduction	2
1.2	Material and Methods	2
	Data size	2
	Dataset and Patient Cohort	2
	Base Architecture	2
	Incorporation of DVHs	3
	Three models comparison	3
1.3	Results	3
	Quantitative Performance	3
	Prescription Adaptation	3
	Clinical Relevance	3
1.4	Conclusions	3
2	Attention Mechanism on Dose-Volume Histograms for Deep Dose Predictions (SFRO 2024)	5
2.1	Introduction	5
2.2	Material and Methods	6
	Patient Data and Preprocessing	6
	Model Architecture and Training	6
	U-net 0 (Baseline)	6
	U-net 1 (DVH with DAFT)	6
	U-net 2 (DVH with Cross-Attention)	6
	Loss Function and Optimization	6
2.3	Results	8
	Evaluation Metrics	8
	Models Performances	8
	Prescription Adaptability	8
2.4	Conclusions	8

Abstract

The previous chapter discussed automation techniques that enhance classical treatment plan optimization methods. This chapter focuses on novel automation techniques for radiotherapy treatment planning that bypass traditional optimization frameworks. With the advent of deep learning, dose prediction models have emerged, offering an alternative approach to generating treatment plans. These models predict synthetic 3D dose distributions, which can be used in dose mimicking to create clinically deliverable treatment plans. In this chapter, we explore advancements in dose prediction techniques, explicitly focusing on methods that improve the accuracy and adaptability of these predictions. By incorporating additional information, such as Dose-Volume Histograms, we propose methods to make dose prediction models more clinically relevant and capable of generating treatment plans that align with varying prescription protocols and dosimetrist preferences.

1 Dose-Volume Histograms Guided Deep Dose Prediction for Radiotherapy Treatment Planning (SFPM 2024)

1.1 Introduction

Traditionally, the creation of radiotherapy treatment plans has been a semi-manual process, where dosimetrists finetune importance factors assigned to structures and constraints. A cost function is then used through a classical optimization algorithm to calculate the optimal plan.

In recent years, deep learning in treatment planning has gained attention. Deep learning models can predict the three-dimensional dose distribution based on patient-specific anatomical data derived from medical imaging (CT scans). While the predicted dose distribution may not directly represent a deliverable treatment plan, it serves as the basis for determining a clinically viable plan through dose mimicking. Dose mimicking is an optimization technique that eliminates the need for manual adjustment of importance factors by dosimetrists. Therefore, the ability to predict a clinically acceptable and near-deliverable 3D dose distribution for any patient presents significant potential for fully automating the radiotherapy planning process. It is important to note that the successful application of dose mimicking [MWM⁺17, SXF⁺22] requires a target dose distribution that is nearly deliverable; thus, arbitrarily setting the target dose to zero for OARs is not feasible.

However, this approach requires further adaptation to accommodate specific clinical guidelines. A potential solution involves training individualized models for each treatment center, allowing institution-specific practices and guidelines to be incorporated. However, deep learning dose prediction models are computationally large, and implementing separate models for each center is resource-intensive. Furthermore, such models require substantial datasets for effective training. Consequently, smaller treatment centers may lack the necessary data volume to train a comprehensive model adequately. Additionally, a separate model may be required for each prescription type due to the variability in prescription doses, making manual treatment planning necessary for non-standard cases. Finally, clinicians and dosimetrists may prefer manually adjusting treatment plans in some cases. Such adjustments are not feasible within the current model framework.

We propose a novel approach that incorporates target DVHs directly into the input of the deep learning-based dose prediction model. Incorporating DVHs introduces interactivity into the model, allowing adjustments to the target DVH to yield corresponding changes in dose predictions. This methodology enables a workflow where dosimetrists can refine the predicted dose distribution according to specific clinical objectives. Furthermore, by establishing a template target DVH tailored to each clinic, the same model can be deployed across multiple centers while generating 3D dose predictions that align with the specific practices of each institution.

1.2 Material and Methods

Data size We defined a bounding box of dimensions $120 \times 180 \times 180 \text{ mm}^3$ centered on the PTV, with isotropic voxels of $5 \times 5 \times 5 \text{ mm}^3$. This box size was chosen to accommodate the PTV and the relevant OARs, namely the rectum and bladder, while maintaining a balance between computational feasibility and model accuracy. A larger bounding box would have increased model complexity and computational time without substantial benefit.

Dataset and Patient Cohort The dataset used for model training comprised 168 patients from the Institut régional du Cancer de Montpellier radiotherapy department. These patients were selected based on their anatomical conformity to the $120 \times 180 \times 180 \text{ mm}^3$ bounding box. These patients received either 62 Gy or 78 Gy prescribed doses to the PTV, with OARs including the bladder and rectum. The dataset was split into training, validation, and test subsets, with 80% used for training, 10% for validation, and 10% for testing.

Base Architecture The model architecture is based on a 3D U-net, a well-established neural network architecture for volumetric data. The input to the network consisted of four elements: the patient’s CT scan, the contour of the PTV, the rectum contour, and the bladder contour. The model output was the predicted three-dimensional dose distribution. The encoder part of the U-net consisted of four convolutional layers with residual connections to improve gradient flow during training, while the decoder section included five convolutional layers. Skip connections were implemented between the corresponding encoder and decoder layers to preserve spatial information across the model.

Incorporation of DVHs Dose-volume histograms represent one-dimensional curves, whereas the CT images, anatomical contours, and predicted dose distributions are inherently three-dimensional data. We employed the Direct Affine Feature Transforms (DAFT) technique to integrate these disparate data types within the neural network [PWW21]. DAFT dynamically scales latent feature maps within the network, enabling the combination of imaging data with DVH information.

For this study, we incorporated the DVHs for the primary structures of interest: the PTV, rectum, and bladder. Not all points along a DVH curve hold equal clinical significance. Dosimetrists typically focus on regions at the beginning and end of the curve, where the volume approaches 0% or 100%. To better capture these critical areas, we employed a non-uniform sampling strategy based on the Chebyshev distribution, which provides a higher density of points near the curve’s extremities. The Chebyshev points, defined in the range $[-1, 1]$, were remapped to the interval $[0, 1]$ for this purpose. Given their critical importance in clinical decision-making, this sampling technique allows us to prioritize accurate sampling at the curve’s extremities. The sampled points were subsequently processed by a two-layer perceptron, responsible for predicting the scaling parameters, α and β , used by the DAFT mechanism to modulate the feature maps.

Three models comparison We evaluated three model configurations with varying levels of DVH incorporation, the architectures of which are described in figure 1.

The first model referred to as "C" or the *classic* model (figure 1a), consists of a standard 3D U-net architecture without any incorporation of DVH information. The second model denoted as "B" or the *bottleneck* model (figure 1b), integrates target DVH data using the DAFT technique, as described previously, at the bottleneck layer of the U-net. In the third model, termed "A" or the *all connections* model (figure 1c), DVH information is incorporated via the DAFT technique both at the bottleneck layer and across all skip connections between the encoder and decoder of the U-net. During the model training process, the clinical DVHs corresponding to the real delivered doses were used as target DVHs for the optimization.

1.3 Results

Our results indicate that incorporating DVH data improves dose prediction’s quantitative and qualitative aspects.

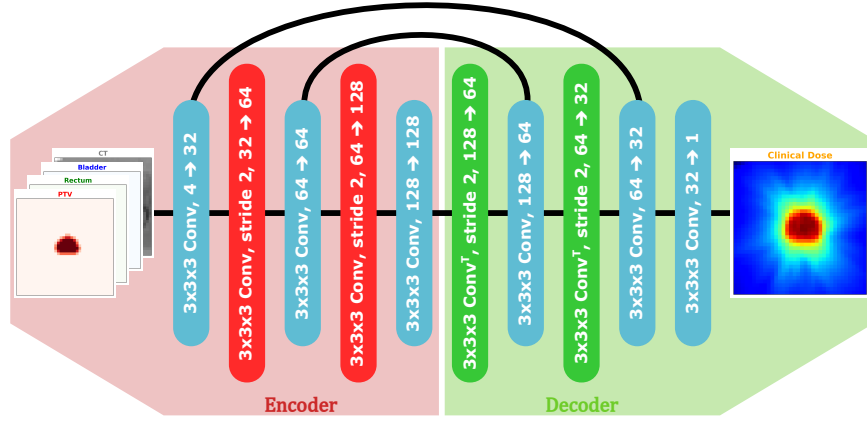
Quantitative Performance We used the Mean Absolute Error (MAE) between the predicted and ground-truth dose distributions to evaluate model performance. The incorporation of dose-volume histogram data into the networks resulted in improved quantitative performance. The MAE measured on the test dataset was 2.42 Gy for model A, 2.58 Gy for model B, and 3.18 Gy for model C.

Prescription Adaptation In addition to the quantitative improvements, a qualitative analysis of the DVHs associated with the predicted dose distributions confirmed the benefit of including DVH information. A key finding from our study was that models A and B could adapt their deep dose predictions based on the prescribed dose, with model A showing a slight advantage over model B regarding accuracy (see figure 2). The dataset comprised patients with two distinct prescription doses: 62 Gy and 78 Gy to the PTV. Model C consistently predicted dose distributions resembling a 65 Gy prescription, demonstrating a lack of adaptability to the varying prescription levels (see figure 2). In contrast, models A and B displayed greater flexibility, successfully adjusting their dose predictions following the prescribed doses for each patient. This adaptive behavior highlights the effectiveness of incorporating DVH information, allowing the models to tailor dose predictions to specific prescription requirements.

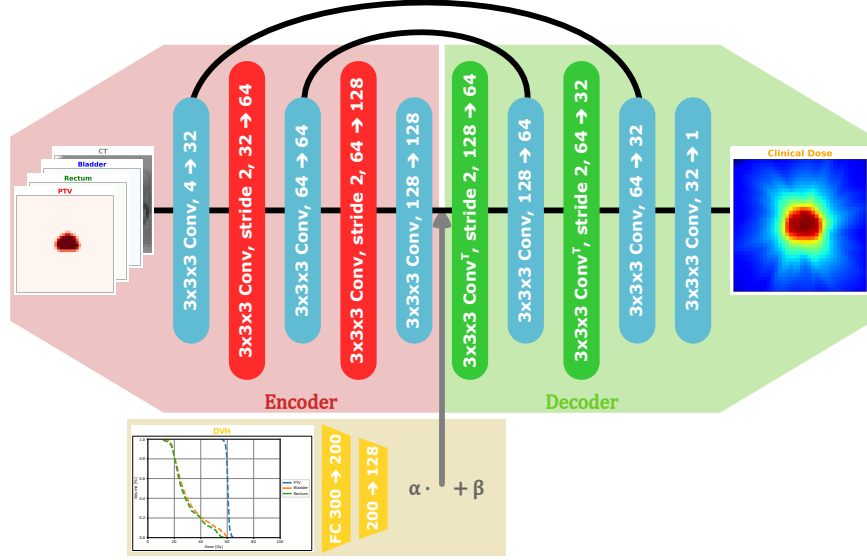
Clinical Relevance The ability to adjust dose predictions based on DVH data demonstrates a significant advantage of our approach. The prescription dose can vary in clinical practice depending on the tumor type, stage, and patient characteristics. By incorporating DVH data, our models provide a more flexible and personalized approach to dose prediction, allowing the TPS to generate a plan that aligns with clinical objectives and dosimetrist input.

1.4 Conclusions

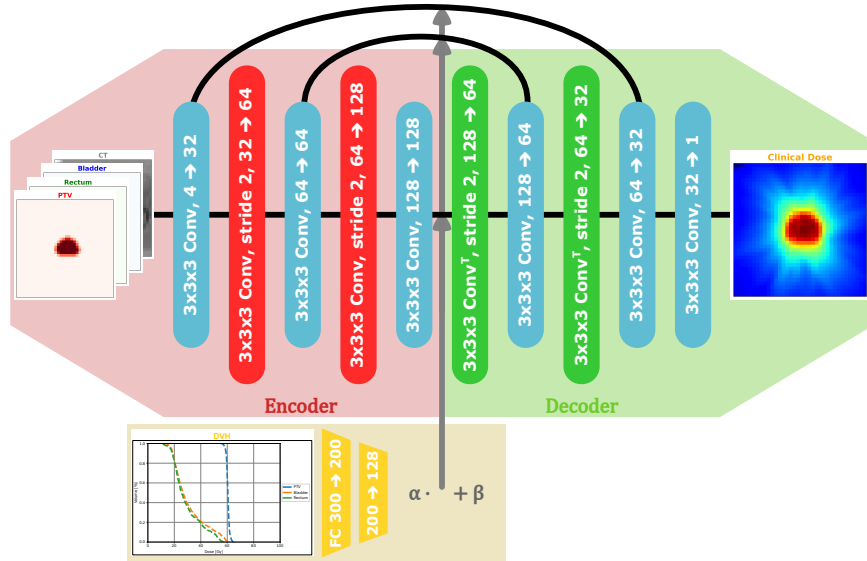
Our study shows the importance of incorporating DVH information into deep learning-based dose prediction models. In this study, the target DVH links the clinical objectives defined by the dosimetrist and the predicted dose distribution. We create a system allowing center-dependent adjustments by embedding this information into the model. This system also allows interactive adjustment of the dose distribution when the proposed treatment plan



(a) Model C: No DVH data, **classical** U-net.



(b) Model B: DVH data using DAFT on the **bottleneck** of the U-net.



(c) Model A: DVH data using DAFT on **all connections** between the encoder and the decoder part of the U-net.

Figure 1: Architecture diagram of models A, B and C.

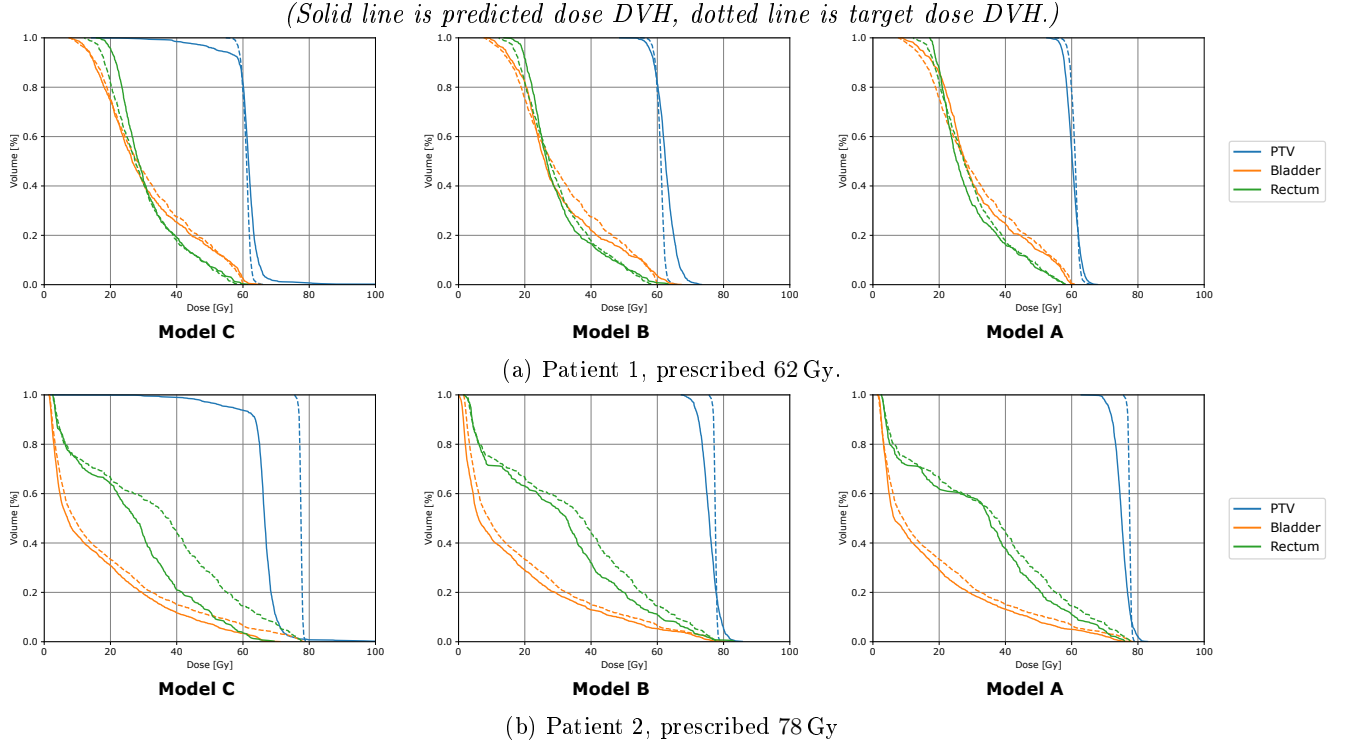


Figure 2: DVHs of the dose predicted by each model on two test set patients.

is close to clinically acceptable. The comparison of models A, B, and C highlights the advantages of integrating DVH data in the network. The comparison also shows that integrating the information once at the bottleneck is sufficient.

An additional advantage of the proposed method is the capacity to generate standardized target DVH templates for treatment planning. While averaging 3D dose distributions across multiple patients is not feasible, it is possible to compute average DVHs. These average DVHs can be stratified by anatomical site, prescription dose, and clinical practice, providing a target for dose prediction with no effort. Dosimetrists and doctors can further modify these templates to meet specific clinical requirements in case of non-standard patients. Once an optimal set of DVHs is established for a given center’s protocols, it can be reused for future patients with only minor adjustments. This framework could enhance the efficiency and consistency of the treatment planning process.

Our proposed approach demonstrates the feasibility and benefits of incorporating DVH data into deep learning-based dose prediction models for radiotherapy treatment planning. By embedding DVH information, we improve dose prediction accuracy and allow for interactive fine-tuning based on clinical objectives. This technique opens the door to a new workflow where dosimetrists can design target DVHs, and the TPS generates the deliverable treatment plan that best matches these targets. Further studies will explore the generalizability of our model across different cancer types and radiotherapy modalities. Additionally, clinical validation studies will be crucial to assess the real-world impact of our proposed method on treatment outcomes and workflow efficiency.

2 Attention Mechanism on Dose-Volume Histograms for Deep Dose Predictions (SFRO 2024)

2.1 Introduction

This study builds on work from the previous section We explore using attention mechanisms to improve the incorporation of target Dose-Volume Histogram information into deep-learning models for radiotherapy dose prediction. In traditional radiotherapy planning, DVHs are essential for assessing and optimizing dose distributions for the Principal Target Volume and Organs at Risk. However, existing deep learning models for dose prediction, while capable of generating accurate 3D dose distributions, often fail to integrate DVH constraints fully [PXR⁺23]. This

limitation reduces their utility in real-time treatment planning.

In previous work, deep learning approaches primarily focused on predicting dose distributions directly from patient imaging data, such as CT scans and anatomical contours of the PTV and OARs. These methods typically rely on convolutional neural networks (CNNs), like the Unet, trained with voxel-wise loss functions [RFB15]. While successful in producing accurate 3D dose maps, they generally treat DVH-related information as secondary, with little attention given to adapting predictions based on these critical clinical metrics.

In the last section, we proposed integrating target DVH into the deep dose model to address this gap. In this section, we extend prior approaches by proposing attention mechanisms to integrate target DVH information more effectively into the dose prediction process. Attention mechanisms, widely adopted in domains such as natural language processing and image recognition, enable models to focus on relevant features within the input data selectively [NZY21]. By leveraging attention, we aim to enhance the model’s capacity to dynamically adapt its dose predictions following specified DVH constraints, providing a more precise and clinically responsive framework for radiotherapy treatment planning.

This work evaluates three U-net-based architectures, including a novel model using a cross-attention mechanism to incorporate DVH information into the dose prediction process directly. Our goal is to enhance the accuracy and adaptability of deep learning models for dose prediction, ultimately enabling dosimetrists to have greater control over the treatment planning process through DVH-guided model interactions.

2.2 Material and Methods

Patient Data and Preprocessing We again used the cohort of 168 patients from the previous section who had undergone radiotherapy treatment. The dataset included patients prescribed 62 Gy or 78 Gy on the PTV. The dataset was randomly split into 80% for training, 10% for validation, and 10% for testing. We collected the CT scans, PTV contours, and OAR contours for each patient. The CT data was resampled to a voxel size of $0.5 \times 0.5 \times 0.5 \text{ mm}^3$ to standardize inputs across patients. As in the last section, we used the same box $120 \times 180 \times 180 \text{ mm}^3$ centered on the PTV.

Model Architecture and Training The backbone architecture for all models in this study is a 3D-convolutional Unet, a commonly used CNN for image-to-image translation tasks, such as segmentation and dose prediction. Our U-net is of depth two, meaning it performs two levels of down-sampling and up-sampling. The U-net outputs a 3D dose prediction for each voxel within the patient’s anatomy. Details architecture diagrams can be found in figure 3.

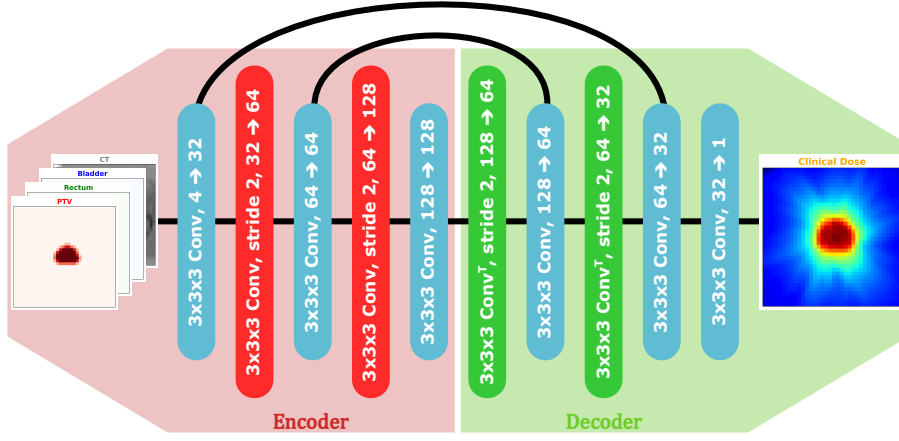
U-net 0 (Baseline) The first model, U-net 0 (architecture diagram can be found on figure 3a), serves as a baseline and does not incorporate target DVH information. It receives the CT scan, PTV, and OAR contours as input and predicts the 3D dose distribution purely from this spatial data.

U-net 1 (DVH with DAFT) The second model, U-net 1 (architecture diagram can be found on figure 3b), incorporates DVH data using the DAFT approach, like in the previous section. In this model, the target DVHs are provided as additional inputs, and the model is trained to predict 3D dose distributions that match these target DVHs.

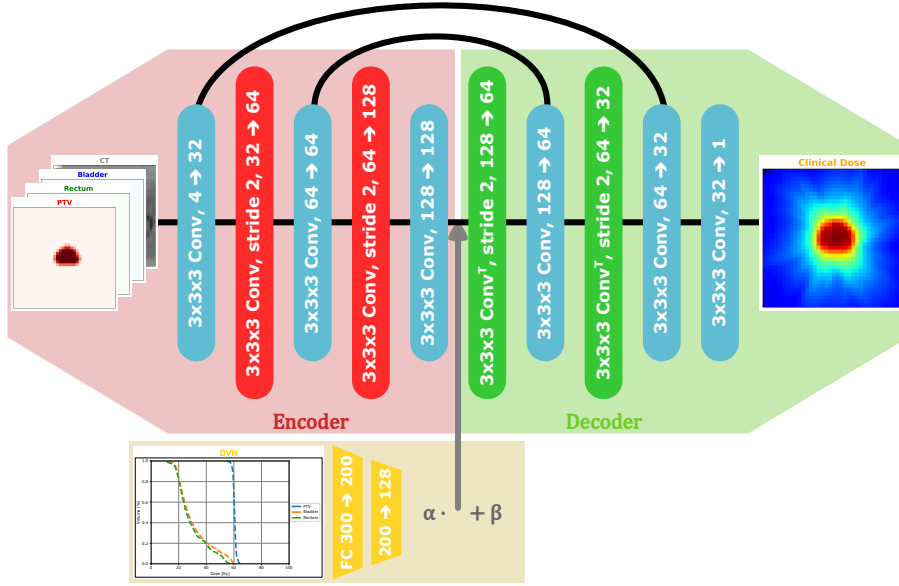
U-net 2 (DVH with Cross-Attention) Our novel contribution, U-net 2 (architecture diagram can be found on figure 3c), integrates DVH data using a cross-attention mechanism. In this architecture, the model takes queries from the 3D dose prediction and keys and values from the target DVHs. The result of the 4-head attention mechanism is the same shape as that of the 3D input. The cross-attention mechanism allows the model to adjust its dose predictions dynamically based on the desired DVH information. The output of the attention mechanism is re-injected into the U-net during the decoding process, modifying the predicted dose map to match the target DVHs better.

Loss Function and Optimization The models were trained using a loss function that combines voxel-wise Mean Absolute Error (MAE) with an L1 loss on the Dose-Volume Histogram (DVH), designed to penalize deviations between the predicted and target DVHs. Incorporating the L1 loss on DVH significantly improved the model’s convergence. The total loss function is expressed as:

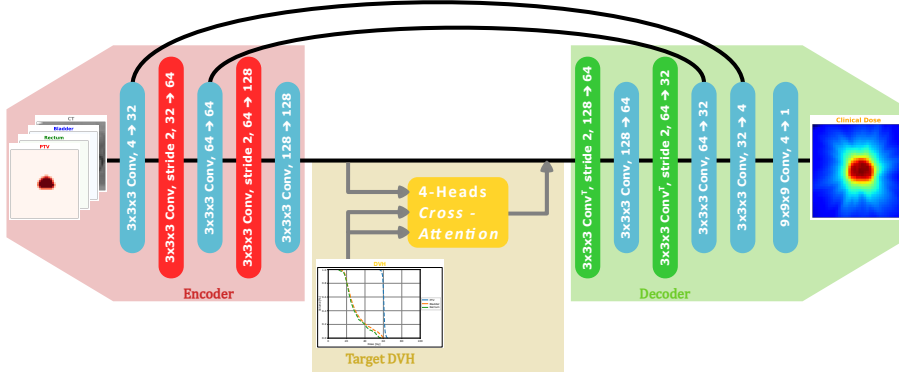
$$\text{Total Loss} = \alpha \cdot \text{MAE}_{\text{voxel}} + \beta \cdot \text{L1}_{\text{DVH}}$$



(a) U-net 0: no DVH information.



(b) U-net 1: DVH with DAFT.



(c) U-net 2: DVHs with cross-attention.

Figure 3: Architecture diagrams of U-nets 0, 1 and 2.

Metric	Unet-0	Unet-1	Unet-2
3D Dose MAE (Gy)	3.093	2.254	2.210
Mean DVH Deviation (Gy)	1.942	1.051	0.930

Table 1: Models performances comparison.

where α and β are hyper-parameters that control the relative contribution of the two terms. For this study, we set both α and β to 1, based on performance observed on the validation set. However, future work could further optimize these parameters to enhance model performance.

The models were trained using the Adam optimizer with a 10^{-4} learning rate. Using the validation loss criterion, early stopping was applied to prevent overfitting and ensure robust generalization.

2.3 Results

Evaluation Metrics The performance of the models was assessed using two primary metrics. The first metric was the 3D Dose Mean Absolute Error (MAE), which measures the voxel-wise MAE between the predicted dose distributions and the corresponding ground truth values. The second metric was the Dose-Volume Histogram (DVH) deviation, calculated as the mean L1 deviation between the predicted and target DVHs for both the Principal Target Volume (PTV) and Organs at Risk (OARs). These metrics provide a comprehensive evaluation of the models’ accuracy in predicting dose distributions and adherence to clinical DVH constraints.

Models Performances The performance of the three models is summarized in Table 1. Both U-net 1 and U-net 2, which incorporated DVH information, outperformed the baseline U-net 0 across all metrics.

As shown, the inclusion of DVH data significantly reduced the MAE in the predicted dose distributions. U-net-2, which used a cross-attention mechanism, showed the best overall performance, with a minor error reduction compared to U-net 1.

Prescription Adaptability The ability of the models to adapt to varying prescriptions was also evaluated. Similarly to the last section, U-net 0 struggled to predict the correct dose when the prescription varied from its training distribution. Regardless of the actual prescription, the model consistently predicted doses around 65 Gy. In contrast, U-net 1 and U-net 2 adapted their dose predictions to the provided DVHs, making them more versatile in clinical scenarios where the prescription may differ across patients. There were only a few differences between the performances of U-net 1 and U-net 2.

2.4 Conclusions

This study re-demonstrates the feasibility of incorporating DVH data into deep-learning models for radiotherapy dose prediction. We have marginally improved model performances and adaptability by using an attention mechanism to integrate target DVHs. The cross-attention mechanism, in particular, offers a novel way to align predicted dose distributions with clinical goals flexibly.

Our findings indicate that this approach could facilitate the development of a more efficient radiotherapy treatment planning workflow. In this proposed framework, dosimetrists would focus primarily on creating standardized target DVH templates rather than manually adjusting dose distributions for each patient. These templates would serve as a foundation, requiring minimal adjustments to accommodate individual patient-specific characteristics, thereby streamlining the treatment planning process.

Designing a template for 3D dose distributions is not feasible, as 3D doses are highly patient-specific and cannot be transferred between patients. In contrast, DVHs are more generalizable and can be applied across different patients. This generalizability makes them suitable for the proposed workflow, where template DVHs can be established and used with deep learning-based dose prediction models guided by target DVH. This workflow could reduce planning time and improve consistency across patients. Future work will explore incorporating more complex DVH structures and refining the attention mechanism to enhance model interpretability and clinical usability.

References

- [MWM⁺17] Chris McIntosh, Mattea Welch, Andrea McNiven, David A Jaffray, and Thomas G Purdie. Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method. *Phys. Med. Biol.*, 62(15):5926–5944, jul 2017.
- [NZY21] Zhaoyang Niu, Guoqiang Zhong, and Hui Yu. A review on the attention mechanism of deep learning. *Neurocomputing*, 452:48–62, 2021.
- [PWW21] Sebastian Pölsterl, Tom Nuno Wolf, and Christian Wachinger. *Combining 3D Image and Tabular Data via the Dynamic Affine Feature Map Transform*, page 688–698. Springer International Publishing, 2021.
- [PXR⁺23] Jinghui Pan, Jinsheng Xiao, Changli Ruan, Qibin Song, Lei Shi, Fengjiao Zhuo, Hao Jiang, and Xiangpan Li. Deep-learning-driven dose prediction and verification for stereotactic radiosurgical treatment of isolated brain metastases. *Front. Oncol.*, 13:1285555, nov 2023.
- [RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.
- [SXF⁺22] Zihan Sun, Xiang Xia, Jiawei Fan, Jun Zhao, Kang Zhang, Jiazhou Wang, and Weigang Hu. A hybrid optimization strategy for deliverable intensity-modulated radiotherapy plan generation using deep learning-based dose prediction. *Med. Phys.*, 49(3):1344–1356, mar 2022.