Heterogeneous IMRT and VMAT Dose Calculation using Deep Learning with Beam Path Reconstruction and PTV Information

Loes Vandenbroucke[1,2], Emma Van Riet[3,4], Dylan Callens[3,4], Patrick Berkovic[3,4], Maarten Lambrecht[3,4], Sandra Nuyts[3,4], Wouter Crijns[3,4], and Frederik Maes[1,2]

[1] Department of ESAT/PSI, KU Leuven, Belgium
[2] Medical Image Research Center, UZ Leuven, Belgium
[3] Laboratory of Experimental Radiotherapy, KU Leuven, Belgium
[4] Department of Radiation Oncology, UZ Leuven, Belgium
loes.vandenbroucke@kuleuven.be

Abstract. Deep learning (DL) enables fast 3D dose calculation for radiotherapy, but existing approaches are restricted to a single treatment technique or tumor site. Furthermore, representing delivery parameters in the patient domain remains challenging. We propose a generalizable DL framework for full-plan dose calculation that reconstructs the full beam path by projecting the fluence maps into the patient domain and that incorporates PTV information to guide target dose prediction. A heterogeneous model trained jointly on lung IMRT and head-and-neck (H&N) VMAT was compared with modality-specific models to evaluate generalization across treatment techniques and tumor sites. Compared with homogeneous training, heterogeneous training increased the average gamma passing rate (98.42% vs. 98.26% for H&N VMAT; 98.60% vs. 98.54% for lung IMRT) and reduced mean dose difference for the planning target volume (2.07% vs. 3.34% for H&N VMAT; 2.16% vs. 2.82% for lung IMRT). Organs at risk mean dose differences remained below 3% for both approaches. Ablation experiments demonstrated that beam-path reconstruction and PTV information improve prediction accuracy.

Keywords: Radiotherapy · 3D Dose Calculation · Deep Learning

1 Introduction

Radiotherapy (RT) aims to deliver a high dose of radiation to the target (PTV), while minimizing exposure of organs at risk (OARs). Intensity-modulated radiotherapy (IMRT) [1] and volumetric modulated arc therapy (VMAT) [11] achieve conformal dose by modulating beam fluence (i.e., aperture and intensity) using multi-leaf collimators (MLC) [6]. IMRT and VMAT require careful treatment planning which relies on repeated dose calculations to optimize treatment delivery. Conventional dose calculation algorithms are a computational bottleneck [2, 3], especially for adaptive radiotherapy, where plans are re-optimized to

account for anatomical changes [4]. Deep learning (DL) has been investigated to accelerate treatment planning through both optimal dose prediction from patient anatomy [10, 13] and dose calculation from a treatment plan and its delivery parameters. The latter is particularly relevant for dose re-calculation during iterative plan optimization. DL-based dose calculation methods can be classified into single-beam [5, 7, 9] and full-plan approaches [15, 14, 17], with the latter estimating the plan dose in a single forward pass. These approaches require an effective representation for mapping treatment parameters from the machine domain in the patient domain. Existing methods address this challenge through computationally expensive approximations [5, 7] or by projecting 2D fluence maps (FMs) into the 3D patient volume. For single-beam dose calculation, Liang et al. [9] projected FMs along the beam path for H&N VMAT. In contrast, Yan et al. [17] projected the FMs onto a single plane at fixed distance from the isocenter for full-plan dose calculation for prostate VMAT. We extend beam-path fluence reconstruction to full-plan dose calculation and compare it with single-plane projection. We address the additional challenge of heterogeneous treatment techniques and tumor sites by training jointly on lung IMRT and head-and-neck (H&N) VMAT. Lung and H&N targets show large spatial variation within complex anatomical regions, increasing complexity for accurate dose calculation. Therefore, we explicitly incorporate PTV prescription information in our DL model.

2 Methodology

Patient data Following institutional ethics approval (S65068, S69848), two datasets were retrospectively collected of lung and H&N cancer patients treated at our institution with RT between 2018 and 2025. The lung cohort comprises 217 patients with locally advanced non-small cell lung cancer (67 left, 150 right, prescription doses 30-66 Gy) treated with nine-beam IMRT according to our clinical class solution. The H&N cohort includes 106 patients treated for locally advanced squamous cell carcinoma with bilateral 3-arc VMAT (prescription doses 60-70 Gy). All treatments were delivered on a 6 MV FFF Halcyon™ linear accelerator (Varian, a Siemens Healthineers Company), and treatment plans were created in Eclipse™. Original dose distributions were recalculated using ACUROS (v18.1.0) to ensure a consistent dose calculation algorithm. Dose distributions were normalized such that the mean PTV dose equaled the prescription, using both the primary and nodal PTV for lung IMRT and high-dose primary PTV for H&N VMAT, reflecting clinical practice. CT Hounsfield units were converted to electron density using the TPS calibration curve, and CT volumes were resampled to a 2.5x2.5x3 mm grid centered around the treatment isocenter.

Patient-domain beam-path reconstruction Treatment delivery parameters were transformed from machine domain at treatment isocenter to patient domain, enabling the DL model to combine information on beam delivery and anatomy. For IMRT, treatment delivery parameters comprise FMs. For VMAT,

FMs are created from the MLC aperture and MUs at each control point, sampled every 2 degrees of gantry rotation. Previous work for full-plan dose calculation projects the FMs to a single plane at a fixed distance from the isocenter [17], resulting in a sparse representation of the treatment geometry, as illustrated in Figure 1a. Instead, we reconstruct the full beam path by projecting the FMs through the patient volume [8], accounting for gantry angle and beam divergence, as illustrated in Figure 1b, yielding a richer representation of treatment geometry. The beam-path reconstruction consists of geometric scaling and interpolation operations. The reconstructed beam paths from all gantry angles are superimposed into a single volume and masked by the patient body contour.

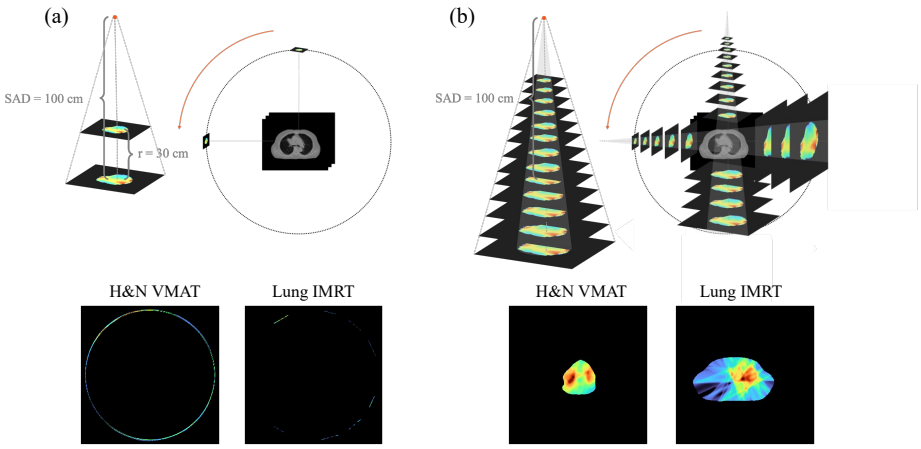

Fig. 1: Illustration of patient-domain encoding of FMs using (a) a sparse single-plane projection at a fixed distance r of the isocenter [16, 17] and (b) the proposed dense full beam path reconstruction, for a source-to-axis distance (SAD) of 100 cm.

PTV prescription encoding Extending prior work on prostate VMAT for full-plan dose calculation [17], we adapt to anatomically more complex lung and H&N cases by adding a PTV mask as an extra input channel to account for irregular target geometries, including primary and nodal targets. For H&N cases with multiple prescription levels, individual PTV masks were weighted by their prescribed dose and normalized by the total plan prescription.

Model architecture and training The proposed framework is based on a 3D U-Net architecture [12] with four encoder and decoder stages. The model input consists of the patient CT in electron density, FMs in patient domain and the PTV mask. Each encoder block consists of two 3x3x3 convolutions with instance normalization, ReLU activation and residual connections. Downsampling

combines average and max pooling [17], after which outputs are concatenated to capture both global and local dose features. The decoder follows the encoder structure, using trilinear and nearest-neighbor upsampling. The number of feature maps starts at 24 and is doubled every downsampling step. Training was performed on a single NVIDIA A100 GPU with 80 GB RAM using the Adam optimizer with L1 loss function for 500 epochs. The initial learning rate was set at 1e-4 and adjusted via cosine annealing.

Experiments The datasets were randomly divided into training, validation, and test subsets, consisting of respectively 131/43/43 cases for lung IMRT and 64/21/21 cases for H&N VMAT, ensuring balanced representation of tumor laterality and subregion across splits. We first compare modality-specific homogeneous models (IMRT-only, VMAT-only) with a heterogeneous model trained on both modalities jointly (IMRT+VMAT). We then perform several ablation experiments, comparing our beam path reconstruction (BP) to the single plane representation (SP) of [17] and evaluating the impact of adding PTV information in both approaches (BP+PTV, SP+PTV). Performance is evaluated on the test set using gamma passing rate (GPR) (2%/2mm, 10% threshold) and clinically relevant dose-volume histogram (DVH) metrics, including PTV coverage (D_{95}, D_{98}, D_{99}), near-maximum dose (D_2), mean dose (D_{mean}) and conformity index (CI), and OAR mean dose (D_{mean}). Statistical significance is assessed using the Wilcoxon signed-rank test ($\alpha = 0.05$). To provide a preliminary assessment of robustness to an unseen delivery paradigm, the heterogeneous and homogeneous models were evaluated on 5 additional left lung RT cases treated with RapidArc Dynamic (RAD). RAD combines continuous rotation of VMAT with static IMRT beams. RAD treatment was delivered on a TrueBeam™ linear accelerator (Varian, a Siemens Healthineers Company) with prescription doses of 45-60 Gy.

3 Results

Heterogeneous vs homogeneous models The heterogeneous model achieves accurate dose estimation with a GPR of 98.42% for the lung IMRT cases and 98.60% for the H&N VMAT cases, slightly outperforming the homogeneous models (Table 1a). Table 1b, summarizes the difference for key metrics for the primary PTV between the heterogeneous and homogeneous models. Overall, dose differences were within 3% for the heterogeneous model, except for D_{98} in the lung IMRT cases. For the H&N VMAT cohort, the heterogeneous model achieves lower errors in PTV coverage (D_{95}, D_{98} and D_{99}), as well as in the PTV mean dose and near-maximum dose, compared to the homogeneous VMAT-only model. The improvements in D_{98}, D_2, and D_{mean} are statistically significant. Although the homogeneous model shows better conformity, the difference in conformity is not significant. For the lung IMRT cohort, the heterogeneous model achieves lower errors in D_{95}, D_{98}, D_2, D_{mean} and CI, with the improvement in CI reaching statistical significance. The homogeneous IMRT-only model achieves slightly lower errors for D_{99}. However, this difference is not statistically significant. Table

1c summarizes the differences in OAR D_{mean} between the heterogeneous and homogeneous models. Overall, both approaches achieve comparable OAR sparing, with neither model consistently outperforming the other across the evaluated OARs. No statistically significant differences were observed. Figure 2 shows representative test cases for the heterogeneous and homogeneous models. Consistent with the quantitative analysis, the heterogeneous model results in better target coverage resulting in closer agreement with the ground truth DVHs.

Ablation experiments Table 2a shows significant improvement in GPR for the proposed BP+PTV model compared to SP, SP+PTV and BP, for both H&N VMAT and lung IMRT. Table 2b summarizes primary PTV dose errors for the H&N VMAT cohort. Overall, both beam path reconstruction and inclusion of PTV information improve target dose coverage. The proposed BP+PTV model achieves lowest errors for D_{95}, D_{98}, D_{99} and PTV mean and maximum dose. Compared to SP and BP without PTV information, these improvements are statistically significant. The only exception is the CI, for which the SP+PTV model is significantly closer to the ground truth. This may indicate that the model relies more strongly on the PTV input, producing a sharper target boundary that improves CI despite a less accurate overall dose distribution, as reflected by the lower GPR. Beam path reconstruction (BP) consistently outperforms the single-plane representation (SP) across all investigated PTV metrics. The OAR results in Table 2c further support these findings. The proposed BP+PTV model achieves the lowest D_{mean} errors for three out of the four investigated OARs and both beam path models (BP, BP+PTV) outperform the single-plane models (SP, SP+PTV), demonstrating that the proposed beam path reconstruction substantially contributes to a more accurate dose estimation. For the lung IMRT cohort, PTV results are summarized in Table 2d. Similar to H&N VMAT, adding PTV information reduces target dose errors and beam path reconstruction consistently outperforms the single-plane representation. The proposed BP+PTV model significantly outperforms SP and BP, but SP+PTV achieves significantly lower PTV dose errors. This trend does not extend to OAR dose prediction (Table 2e), where BP+PTV yields the lowest OAR D_{mean} errors and both BP and BP+PTV outperform SP and SP+PTV. This suggests that although single-plane representation with PTV information improves target dose prediction for lung IMRT, reconstructing the full beam path provides a better overall representation of the treatment geometry, resulting in more accurate OAR dose prediction. This is further confirmed by a significantly higher GPR for BP+PTV (Table 2a). These findings suggest that SP+PTV relies predominantly on PTV information while capturing limited treatment geometry. This is illustrated in Figure 2b, where the SP+PTV model fails to accurately reconstruct the beam trajectories for the IMRT cases, while the BP+PTV model successfully captures them. However, BP comes with a cost of a longer preprocessing time (60.7ms per FM) than SP (30.7ms). Consequently, full-plan inference times of BP versus SP increased from 0.9s to 1.1s for IMRT and from 17.2s to 33.4s for VMAT.

Generalization to lung RAD On the 5 lung RAD cases, the heterogeneous model achieved an average GPR of 98.15%, which is a slight improvement compared with 98.06% and 97.15% for the IMRT-only and VMAT-only homogeneous models, respectively.

Table 1: Quantitative comparison of GPR, PTV and OAR DVH metrics for heterogeneous (IMRT+VMAT) and homogeneous (VMAT-only, IMRT-only) models (mean ± standard deviation). Dose differences are reported as % of the prescription dose. The best-performing model for each metric is marked in **bold**. * indicates a statistically significant difference between both models ($p<0.05$, Wilcoxon signed-rank test).

(a) GPR (2%/2mm)

	H&N VMAT		Lung IMRT	
Model	IMRT+VMAT	VMAT-only	IMRT+VMAT	IMRT-only
GPR	**98.42 ± 0.53**	98.29 ± 0.76	**98.60 ± 0.83**	98.54 ± 0.92

(b) PTV metrics

	H&N VMAT		Lung IMRT	
Model	IMRT+VMAT	VMAT-only	IMRT+VMAT	IMRT-only
D_{95}	**1.86 ± 1.12**	3.45 ± 3.07	**2.95 ± 2.69**	3.20 ± 2.95
D_{98}	**2.03 ± 1.36** *	3.75 ± 3.03	**3.51 ± 3.77**	3.53 ± 4.32
D_{99}	**2.59 ± 1.44**	4.08 ± 3.19	3.85 ± 3.96	**3.59 ± 5.21**
D_2	**2.42 ± 1.58** *	3.50 ± 3.00	**2.27 ± 1.59**	2.73 ± 1.97
D_{mean}	**2.07 ± 1.17** *	3.34 ± 2.79	**2.16 ± 1.53**	2.82 ± 2.16
CI	0.21 ± 0.20	**0.18 ± 0.24**	**0.10 ± 0.10** *	0.17 ± 0.14

(c) OAR metrics

Model				
H&N VMAT	Brainstem	Parotid left	Parotid right	Spinal cord
IMRT+VMAT	1.02 ± 1.25	**2.87 ± 2.22**	**2.41 ± 2.43**	**1.41 ± 1.54**
VMAT-only	**0.91 ± 1.01**	2.92 ± 3.11	2.94 ± 2.29	2.05 ± 2.07
Lung IMRT	Esophagus	Heart	Lungs	Spinal canal
IMRT+VMAT	1.11 ± 1.11	**0.56 ± 0.61**	**0.55 ± 0.46**	1.05 ± 0.94
IMRT-only	**1.04 ± 1.02**	0.60 ± 0.71	0.57 ± 0.53	**0.99 ± 0.96**

Table 2: Quantitative comparison of GPR, PTV and OAR DVH metrics for the ablation experiments (mean ± standard deviation). Dose differences are reported as % of the prescription dose. The best-performing method for each metric is marked in **bold**. * indicates a statistically significant difference compared with the proposed BP+PTV model ($p<0.05$, Wilcoxon signed-rank test).

(a) GPR (2%/2mm)

Model	BP+PTV	SP	SP+PTV	BP
H&N VMAT	**98.42 ± 0.53**	97.41 ± 0.89 *	98.08 ± 0.67 *	98.40 ± 1.07 *
Lung IMRT	**98.60 ± 0.83**	97.15 ± 1.44 *	97.57 ± 1.21 *	98.40 ± 1.07 *

(b) H&N VMAT - PTV metrics

Model	BP+PTV	SP	SP+PTV	BP
D_{95}	**1.86 ± 1.12**	33.04 ± 14.79 *	2.85 ± 2.01 *	6.68 ± 4.65 *
D_{98}	**2.03 ± 1.36**	34.97 ± 16.11*	2.67 ± 2.60	6.46 ± 4.75 *
D_{99}	**2.59 ± 1.44**	36.37 ± 17.22 *	3.45 ± 3.09	7.13 ± 4.69 *
D_2	**2.42 ± 1.58**	5.21 ± 3.93 *	2.48 ± 1.67	3.52 ± 2.10
D_{mean}	**2.07 ± 1.17**	15.10 ± 8.68 *	2.16 ± 1.35	3.78 ± 3.09 *
CI	0.21 ± 0.20	0.37 ± 0.29	**0.17 ± 0.24** *	0.20 ± 0.17

(c) H&N VMAT - OAR metrics

Model	BP+PTV	SP	S+PTV	BP
Brainstem	1.02 ± 1.25	4.01 ± 8.78 *	1.82 ± 2.50	**0.94 ± 1.15**
Parotid left	**2.87 ± 2.22**	6.67 ± 3.83 *	6.10 ± 6.27	3.53 ± 2.57
Parotid right	**2.41 ± 2.43**	10.56 ± 8.15 *	5.16 ± 4.73	3.01 ± 2.29
Spinal cord	**1.41 ± 1.54**	2.73 ± 2.82	2.06 ± 2.25	1.89 ± 2.00

(d) Lung IMRT - PTV metrics

Model	BP+PTV	SP	SP+PTV	BP
D_{95}	2.95 ± 2.69	27.96 ± 18.80 *	**1.92 ± 2.85** *	5.13 ± 5.63
D_{98}	3.51 ± 3.77	34.32 ± 19.75 *	**2.95 ± 5.48** *	6.12 ± 6.25 *
D_{99}	3.85 ± 3.96	38.16 ± 20.36 *	**3.72 ± 6.89**	6.72 ± 6.40 *
D_2	2.27 ± 1.59	3.46 ± 2.34 *	**1.84 ± 1.48**	4.09 ± 3.24 *
D_{mean}	2.16 ± 1.53	7.99 ± 7.15 *	**1.42 ± 1.28** *	2.85 ± 3.84
CI	0.10 ± 0.10	0.21 ± 0.22 *	**0.07 ± 0.08**	0.13 ± 0.13

(e) Lung IMRT - OAR metrics

Model	BP+PTV	SP	SP+PTV	BP
Esophagus	**1.11 ± 1.11**	3.19 ± 3.32 *	1.99 ± 1.77 *	1.42 ± 1.62
Heart	**0.56 ± 0.61**	1.80 ± 2.22 *	1.74 ± 2.28 *	0.75 ± 1.04
Lungs	**0.55 ± 0.46**	1.76 ± 1.25 *	1.07 ± 0.98 *	0.60 ± 0.60
Spinal canal	**1.05 ± 0.94**	2.82 ± 3.07 *	1.75 ± 2.17 *	1.21 ± 1.42

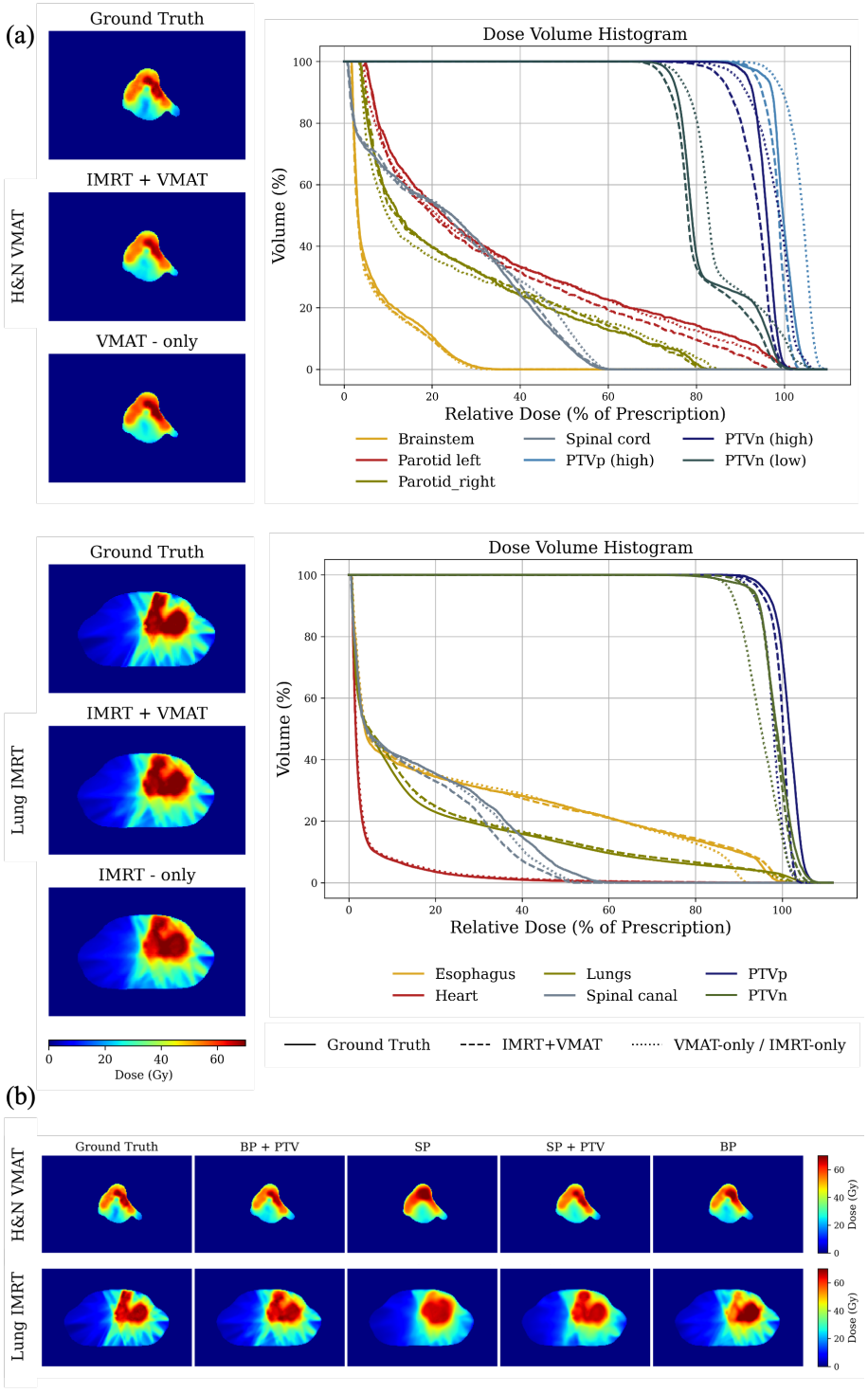

Fig. 2: **(a)** Ground truth and predicted dose and DVHs for a representative H&N VMAT (top) and lung IMRT (bottom) test case, comparing heterogeneous and homogeneous models. **(b)** Ground truth and predicted dose for the ablations.

4 Discussion

This study demonstrates that full-plan DL-based 3D dose calculation can be generalized across treatment techniques and anatomically distinct tumor sites using a heterogeneously trained model. Heterogeneous training improved prediction accuracy over modality-specific methods in terms of voxel-wise agreement and PTV coverage. Furthermore, reconstructing the beam path yielded a more accurate dose prediction compared to a single-plane representation [17], suggesting that a richer representation of treatment delivery helps the network to better infer the relationship between beam configuration, patient anatomy and resulting dose distribution. While single-plane representations have shown promising performance for prostate VMAT, our results indicate that more complex treatment scenarios such as lung IMRT and H&N VMAT benefit from explicitly modeling the beam path. Including PTV information also proved to be beneficial in these settings. To our knowledge, this is the first DL model capable of predicting modulation-specific, plan-conditioned dose across multiple treatment modalities. Multi-modality modeling is particularly relevant in the context of rapidly evolving RT technologies. Preliminary robustness assessment revealed that the proposed model generalized well to unseen RAD treatments, demonstrating adaptability to emerging delivery paradigms. Plan optimization involves intermediate optimization states that are not fully represented in the finalized clinical treatment plans used for current model training. Future work will include investigation of the performance of the proposed framework on intermediate and hypothetical plans, enabling practical integration in optimization workflows, including both conventional iterative or DL-based approaches. Another aspect to consider is the inclusion of the prescription-weighted PTV mask as an input channel. Providing explicit target and prescription information could potentially bias the model toward enforcing high dose conformity in the PTV region. In the current setup, we assume that FMs and MLC apertures have already undergone an initial optimization step, such that a conformal high-dose region around the target is expected.

5 Conclusion

We developed and validated a generalizable DL framework for 3D dose calculation outside the TPS. By transforming the treatment parameters into the patient domain by reconstructing the beam path, the model learns the plan-dose relationship across multiple delivery techniques and tumor sites.

References

1. Bortfeld, T.: IMRT: a review and preview. Physics in Medicine & Biology **51**(13), R363 (2006)
2. Bragg, C.M., Conway, J.: Dosimetric verification of the anisotropic analytical algorithm for radiotherapy treatment planning. Radiotherapy and Oncology **81**(3), 315–323 (2006)

3. Cho, W., Suh, T.S., Park, J.H., Xing, L., Lee, J.W.: Practical implementation of a collapsed cone convolution algorithm for a radiation treatment planning system. Journal of the Korean Physical Society **61**(12), 2073–2083 (2012)
4. Dona Lemus, O.M., Cao, M., Cai, B., Cummings, M., Zheng, D.: Adaptive radiotherapy: next-generation radiotherapy. Cancers **16**(6), 1206 (2024)
5. Fan, J., Xing, L., Dong, P., Wang, J., Hu, W., Yang, Y.: Data-driven dose calculation algorithm based on deep U-Net. Physics in Medicine & Biology **65**(24), 245035 (2020)
6. Jeraj, M., Robar, V.: Multileaf collimator in radiotherapy. Radiology and Oncology **38**(3) (2004)
7. Kontaxis, C., Bol, G., Lagendijk, J., Raaymakers, B.: DeepDose: Towards a fast dose calculation engine for radiation therapy using deep learning. Physics in Medicine & Biology **65**(7), 075013 (2020)
8. Li, T., Scheuermann, R., Lin, A., Teo, B.K.K., Zou, W., Swisher-McClure, S., Alonso-Basanta, M., Lukens, J.N., Ghiam, A.F., Kennedy, C., et al.: Impact of multi-leaf collimator parameters on head and neck plan quality and delivery: a comparison between halcyon™ and truebeam® treatment delivery systems. Cureus **10**(11), e3648 (2018)
9. Liang, B., Xia, W., Wei, R., Xu, Y., Liu, Z., Dai, J.: A deep learning-based dose calculation method for volumetric modulated arc therapy. Radiation Oncology **19**(1), 141 (2024)
10. Nguyen, D., Jia, X., Sher, D., Lin, M.H., Iqbal, Z., Liu, H., Jiang, S.: 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 (2019)
11. Otto, K.: Volumetric modulated arc therapy: IMRT in a single gantry arc. Medical Physics **35**(1), 310–317 (2008)
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer (2015)
13. Song, Y., Hu, J., Liu, Y., Hu, H., Huang, Y., Bai, S., Yi, Z.: Dose prediction using a deep neural network for accelerated planning of rectal cancer radiotherapy. Radiotherapy and Oncology **149**, 111–116 (2020)
14. Tseng, W., Liu, H., Yang, Y., Liu, C., Lu, B.: An ultra-fast deep-learning-based dose engine for prostate VMAT via knowledge distillation framework with limited patient data. Physics in Medicine & Biology **68**(1), 015002 (2023)
15. Xing, Y., Nguyen, D., Lu, W., Yang, M., Jiang, S.: A feasibility study on deep learning-based radiotherapy dose calculation. Medical physics **47**(2), 753–758 (2020)
16. Yan, S., Maniscalco, A., Wang, B., Nguyen, D., Jiang, S., Shen, C.: Quality assurance for online adaptive radiotherapy: a secondary dose verification model with geometry-encoded U-Net. Machine Learning: Science and Technology **5**(4), 045013 (2024)
17. Yan, S., Maniscalco, A., Wang, B., Zhao, H., Nguyen, D., Jiang, S., Shen, C.: Geometry-encoded deep learning (GeoDL) framework for real-time 3D dose verification for online adaptive radiotherapy. Machine Learning: Health **1**(1), 015004 (2025)