Physics-Guided Radiotherapy Treatment Planning with Deep Learning

Anonymized Authors

Anonymized Affiliations
email@anonymized.com

Abstract. Radiotherapy (RT) is a critical cancer treatment, with volumetric modulated arc therapy (VMAT) being a commonly used technique that enhances dose conformity by dynamically adjusting multileaf collimator (MLC) positions and monitor units (MU) throughout gantry rotation. Adaptive radiotherapy requires frequent modifications to treatment plans to account for anatomical variations, necessitating time-efficient solutions. Deep learning offers a promising solution to automate this process. To this end, we propose a two-stage, physics-guided deep learning pipeline for radiotherapy planning. In the first stage, our network is trained with direct supervision on treatment plan parameters, consisting of MLC and MU values. In the second stage, we incorporate an additional supervision signal derived from the predicted 3D dose distribution, integrating physics-based guidance into the training process. We train and evaluate our approach on 133 prostate cancer patients treated with a uniform 2-arc VMAT protocol delivering a dose of 62 Gy to the planning target volume (PTV). Our results demonstrate that the proposed approach, implemented using both 3D U-Net and UNETR architectures, consistently produces treatment plans that closely match clinical ground truths. Our method achieves a mean difference of $D_{95\%} = 0.42 \pm 1.83$ Gy and $V_{95\%} = -0.22 \pm 1.87\%$ at the PTV while generating dose distributions that reduce radiation exposure to organs at risk. These findings highlight the potential of physics-guided deep learning in RT planning.

Keywords: Radiotherapy · Deep Learning · Physics-Guided

1 Introduction

Radiotherapy (RT) is a critical treatment modality for cancer, with approximately 50% of cancer patients undergoing external beam radiation therapy during their disease [8]. The aim of RT to deliver a sufficiently high radiation dose to the tumor while sparing surrounding healthy tissues as much as possible [5]. Over the past decades, various advanced delivery methods have been introduced to increase dose conformity and reduce toxicity. Notably, intensity-modulated radiation therapy (IMRT) [3] and volumetric-modulated arc therapy (VMAT) [23] have transformed clinical practice. VMAT delivers highly conformal doses by continuously adjusting the leaves and jaws of the multileaf collimator (MLC)

system while modulating the monitor units (MU) throughout the gantry rotation [2].

Online adaptive radiotherapy (ART) personalizes RT by updating plans based on anatomical changes. [18]. This continual replanning process requires solving a high-dimensional, non-convex optimization problem [27]. Although conventional algorithms based on Monte Carlo simulations [29] can be highly accurate, they are computationally expensive leading to inefficient workflows. To address these limitations, deep learning methods are increasingly being investigated as a way to accelerate treatment planning.

One prominent deep learning strategy is Deep Reinforcement Learning (DRL), where an agent learns to optimize treatment plans by maximizing a cumulative reward through repeated interactions with an environment [26]. DRL has shown promise in IMRT planning for prostate cancer [25] and has been extended to VMAT planning [16, 15]. In the latter case [15], 3D collapsed cone convolution algorithms [1] serve as the environment, while the Deep Deterministic Policy Gradient (DDPG) algorithm [19] is used for optimization. However, DRL-based approaches face several challenges, particularly their reliance on the accuracy of the environment and the design of reward functions used during training [32].

An alternative strategy uses supervised deep learning on pre-calculated, clinically approved plans generated by commercial treatment planning systems (e.g., Pinnacle from Philips [24] or Monaco from Elekta [10]). Early models used single-arc IMRT data and a four-layer 3D U-Net [34] to predict the MLC apertures from the patient's Computed Tomography (CT) images and the corresponding masks of the Planning Treatment Volume (PTV) and organs-at-risk (OAR) [22]. Later, an MU-decoder was added to predict both MLC configurations and MU values for three-arc breast cancer treatments as warm start for the optimisation [28]. A key limitation of such direct supervision lies in the non-uniqueness of optimal solutions: multiple different MLC and MU configurations can produce clinically equivalent dose distributions, making a single ground truth plan inherently ambiguous.

To address these issues, we propose a two-stage, physics-guided [11], [30] training pipeline for deep learning–based RT planning. In the first stage, our *Deep RT Planner* is trained with direct supervision on the treatment plan parameters. In the second stage, we introduce an additional supervision signal derived from the 3D dose distribution corresponding to the predicted treatment plan, thereby incorporating physics guidance and training in a clinically relevant domain. This dose is generated by the *RT Dose Predictor*, a fully differentiable gated recurrent unit (GRU) neural network [6], pretrained to predict 3D dose distributions from CT scans and treatment plans [33], and remains frozen during the second stage of training.

We evaluate our method on a dataset of 133 patients with prostate cancer, all treated under a uniform 2-arc VMAT protocol delivering 62 Gy to the PTV. We implement two variants of our *Deep RT Planner*: a 3D U-Net [34] with dual decoders—one for MLC masks and one for MU values, and a UNETR [13] architecture, which employs Vision Transformer (ViT) [9] encoders while retaining the

dual-decoder structure. For both architectures, the second-stage physics-guided training significantly improves plan quality and dose accuracy.

2 Method

2.1 Dataset

We collected data from 133 prostate cancer patients treated at our institution between 2018 and 2022, following approval of the Institutional Review Board. For each patient, the planning CT scan and RT Structure Set (RTSS) were obtained, which includes the precise locations of the routinely considered CTV, PTV, rectum and femoral heads.

We recalculated all plans using Pinnacle [24] with standardized parameters: 7 MeV beam energy, 20 treatment sessions, flattening filter-free mode, and a linear accelerator with 160 MLC leaves. Identical dose objectives were applied across all plans to ensure consistency, and an experienced radiation oncologist reviewed each case for clinical validity.

The final dataset included 104 training, 16 validation, and 13 test cases, each comprising a CT scan, binary masks for the CTV, PTV, and OARs, and a corresponding treatment plan. The CT volumes, centered on the isocenter, were resampled to a $144 \times 144 \times 144$ grid (approximately $500 \times 500 \times 500$ mm^3) with an isotropic resolution of 3.5 mm^3. Hounsfield unit (HU) values were clipped to $[-1000, 3000]$ and normalized to $[-1, 1]$, with RTSS masks geometrically aligned to the CT grid.

Each VMAT plan consists of 144 control points (72 per arc). At each control point, a 144×144 binary mask encodes the MLC aperture, representing MLC leaf and jaw positions relative to the isocenter. A corresponding scalar value specifies the monitor units (MU), resulting in 144 MLC aperture masks and 144 MU values per plan.

2.2 Deep RT Planner and RT Dose Predictor

For the *Deep RT Planner*, we experimented with two architectures: a 3D U-Net [34] and a UNETR [13] model. Both architectures take as input a tensor of shape (B, C, D, H, W), where B is the batch size, C is the number of channels, and D, H, and W are the depth, height, and widths. In our setting, $B = 4$, $H = W = D = 144$, and $C = 5$. The first channel contains the CT scan, while the remaining four channels contain the rotation and projection at each control point for the CT, PTV, CTV, and OARs, respectively. This representation predefines the Beam's Eye View [5] at the input level, allowing the model to process spatially aware information. With a single forward pass, the models predict the complete RT plan, consisting of 144 MLC apertures and their corresponding monitor units.

3D U-Net Architecture. We employ a 3D U-Net with an encoder-decoder structure and skip connections across four resolution levels. The encoder applies repeated $3 \times 3 \times 3$ convolutions with batch normalization and ReLU activation,

followed by $2 \times 2 \times 2$ max pooling. The decoder mirrors this structure, using upsampling before applying convolutional layers. Skip connections concatenate encoder features with decoder stages to retain spatial information. A final $1 \times 1 \times 1$ convolution with sigmoid activation predicts the 144 binary MLC apertures.

A global average pooling layer extracts a latent representation from the deepest encoder features, which is processed by fully connected layers to predict the 144 MU values.

UNETR Architecture. We utilize a compact UNETR-based model with a lightweight ViT encoder and dual decoders for MLC mask and MU prediction. The 3D input volume is split into non-overlapping $16 \times 16 \times 16$ patches, which are embedded into a latent space with positional encoding. A streamlined ViT with four transformer blocks (instead of twelve) processes the sequence, generating multi-scale feature representations while retaining intermediate states.

For MLC mask prediction, deconvolution and upsampling restore the original $144 \times 144 \times 144$ resolution, with feature maps concatenated at multiple scales. A $1 \times 1 \times 1$ convolution with softmax activation outputs the 144 binary MLC apertures. Meanwhile, the MU decoder branches from the ViT bottleneck and processes the latent sequence through fully connected layers to predict the 144 MU values.

RT Dose Predictor. For predicting the 3D dose distribution given the RT plan and CT scan, we utilized a convolutional gated recurrent unit neural network, following [33]. We modified the architecture to be fully differentiable in our pipeline, enabling gradient-based optimization during training, while keeping it frozen during the physics-guided stage of training.

The network was trained on 350 cancer patients across multiple tumor sites, including prostate cancer, using treatment plans generated by Monaco [10]. Using a gamma pass rate criterion [31] of 2% and 2 mm for voxels receiving at least 10% of the maximum dose, the model achieves a 99.6% pass rate.

2.3 Two-Stage Physics-Guided Training

First Training Stage. During the first stage of training, the *Deep RT Planner* takes as input the CT scan and the binary masks of the RT Structure Set. The network encodes these inputs, with the MLC aperture decoder predicting the binary MLC apertures and the MU decoder predicting the MU values for the complete 2-arc plan. The network is supervised using the ground truth plan and minimizes the following loss function:

$$\mathcal{L} = \mathcal{L}_{\text{BCE}}(M_{\text{pred}}, M_{\text{true}}) + \lambda \cdot \|\text{MU}_{\text{pred}} - \text{MU}_{\text{true}}\|_1$$

where the first term represents the Binary Cross-Entropy (BCE) loss [12] for the predicted MLC aperture M_{pred} and its ground truth M_{true}. The second term is the L_1 loss for the predicted monitor units MU_{pred} compared to the ground truth MU_{true}. The parameter λ balances the two loss components and is set to 100.

Second Training Stage. A key limitation of direct supervision in the first stage is the non-uniqueness of optimal solutions, as multiple MLC and MU configurations can yield clinically equivalent dose distributions, making a single ground truth plan ambiguous. While our second training stage—incorporating a dose-based loss term—does not fully resolve this ambiguity, it mitigates multi-arc redundancy by guiding the network toward a consistent dose representation. Furthermore, dose supervision evaluates the network's output in a clinically relevant domain, aligning the optimization process with actual treatment objectives.

As shown in Fig. 1, the RT plan predicted by the *Deep RT Planner* serves as input to the *RT Dose Predictor* in a cascaded manner. This process remains fully end-to-end differentiable, allowing dose supervision to backpropagate through the pipeline and optimize the *Deep RT Planner* accordingly. The second-stage loss function is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{BCE}}(M_{\text{pred}}, M_{\text{true}}) + \lambda_1 \cdot \|\text{MU}_{\text{pred}} - \text{MU}_{\text{true}}\|_1 + \lambda_2 \cdot \|D_{\text{pred}} - D_{\text{true}}\|_2^2$$

where D_{pred} and D_{true} represent the predicted and ground truth 3D dose distributions, respectively. The parameters λ_1 and λ_2 control the relative weighting of the MU and dose terms and were set to 100 and 10, respectively.

The first training stage ran for approximately 400 epochs, while the second stage lasted 100 epochs, with early stopping applied if the validation loss did not improve for 10 consecutive epochs. We used the AdamW optimizer [21] with cosine annealing, a learning rate of 10^{-4}, a weight decay of 10^{-3}, and a batch size of 4.

3 Results

3.1 Dosimetric Comparison with Clinical Plans

We evaluated our two-stage methodology using two different neural network architectures as the *Deep RT Planner*: a 3D U-Net and a UNETR. Both networks have comparable model sizes, and this comparison aims to demonstrate that the proposed second-stage physics-guided training enhances performance independently of the *Deep RT Planner* architecture. Results were averaged over a test set of 13 patients. To ensure efficient evaluation, we used CUDA-accelerated preprocessing [7], including rotation and projection operations, and performed model inference on a single NVIDIA RTX A6000 GPU in under one second per patient.

Table 1 presents the differences between predicted and ground truth (GT) dose-volume histogram (DVH) characteristics for the PTV, CTV, and OARs [5]. Each value in the table represents the difference Predicted − GT, where negative values indicate that the predicted DVH characteristics are lower than those of the ground truth. For PTV and CTV, optimal performance corresponds to values closest to zero, minimizing deviation from the ground truth. In contrast, for

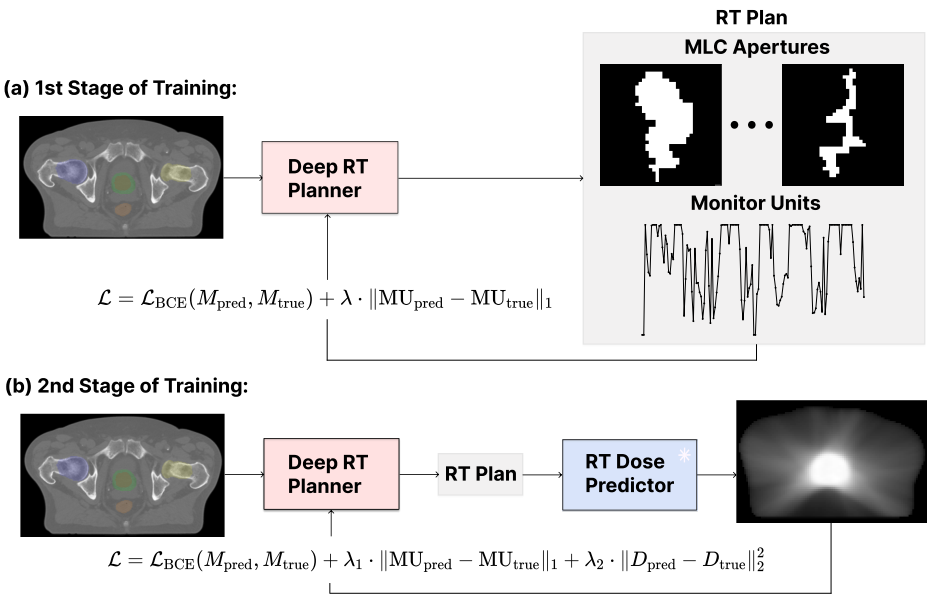

Fig. 1. Two-stage physics-guided training framework: (a) The *Deep RT Planner* is trained with ground truth RT plans. (b) Dose supervision via the *RT Dose Predictor*.

OARs, lower dose values are preferable, meaning a negative difference relative to the ground truth indicates better sparing of healthy tissue.

The second training stage significantly improved key dosimetric metrics for the PTV/CTV, including D98, D95, and V95%. The best performance was achieved using the 3D U-Net, which yielded a mean absolute difference of $D_{95\%} = 0.59 \pm 2.23$ Gy, $D_{98\%} = 0.70 \pm 2.14$ Gy, and $V_{95\%} = -0.42 \pm 1.12\%$, demonstrating strong alignment between the predicted and ground truth treatment objectives.

Physics-guided training reduced OAR toxicity compared to first-stage models. A detailed analysis revealed that for the rectum, one of the most radiosensitive structures, the 3D U-Net achieved greater dose reductions than the UNETR. This trend aligns with previous findings [14]. For the femoral heads, the UNETR performed better on average, though both models exhibited variability in dose metrics.

Fig. 2 presents DVH curves for a single patient before and after the physics-guided stage for the 3D U-Net, where we observe that the physics guidance step brings the dose objectives closer to the ground truth.

3.2 Gamma Pass Rate Analysis

To further evaluate the predicted radiotherapy plans, we computed the gamma pass rate [31] using 3%/3 mm criteria for dose values exceeding 10% of the maximum to assess overall dose distribution and for values above 90% to focus on high-dose regions critical for target coverage.

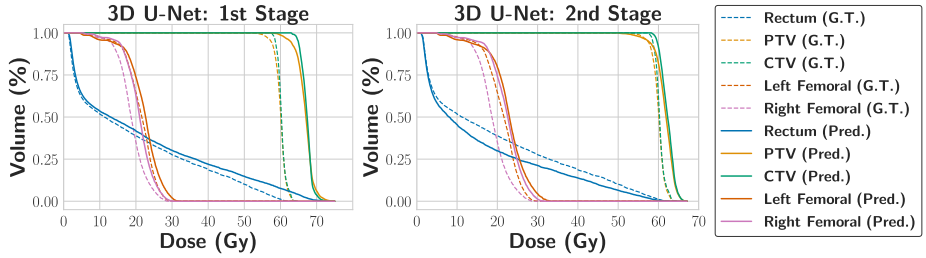

Fig. 2. DVHs comparing predicted dose distributions after the 1st (left) and 2nd (right) training stages of the 3D U-Net for the same patient. Ground truth (G.T.) doses are shown as dashed lines, while predicted (Pred.) doses are shown as solid lines.

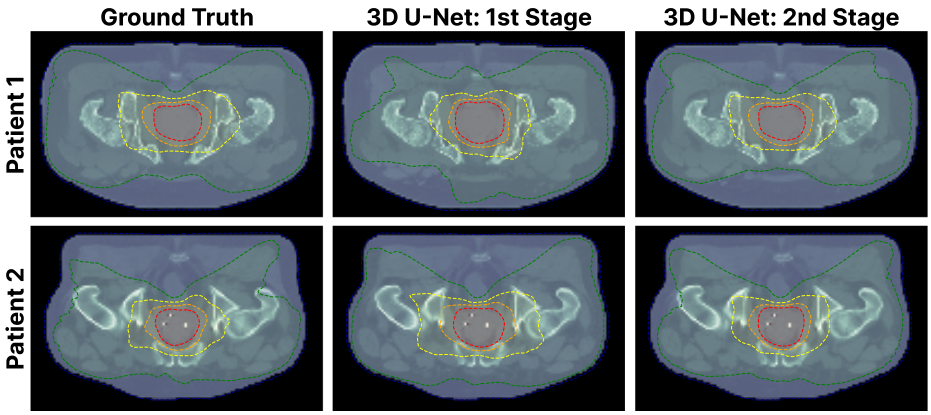

Fig. 3. Isodose distributions (25%, 50%, 75%, 90% of max dose) for two patients. The ground truth (left) is compared to predictions from the 1st (middle) and 2nd (right) training stages of the 3D U-Net.

Table 2 shows that two-stage training consistently improved gamma pass rates, with the largest gain in the high-dose region, nearly doubling performance, and the 3D U-Net model achieved $90.5\% \pm 7.3\%$.

Fig. 3 compares isodose curves for two patients, showing the ground truth alongside predictions from the 1st and 2nd training stages of the 3D U-Net.

4 Discussion

We developed a physics-guided training pipeline that generates treatment plans adhering to clinical DVH criteria in under one second, whereas contemporary GPU-based planning systems require minutes [20]. Our approach directly addresses a key bottleneck in ART: time constraints.

In the first training stage, supervised on ground truth RT plans following approaches similar to those proposed in the literature [28], the generated plans

ROI	Metric	UNETR		3D U-Net	
		1st Stage	2nd Stage	1st Stage	2nd Stage
CTV	$D_{95\%}$ (Gy)	3.32 ± 4.34	**-1.12 ± 2.30**	2.94 ± 3.58	**0.59 ± 2.23**
	$D_{98\%}$ (Gy)	2.79 ± 4.32	**-1.30 ± 2.35**	2.67 ± 3.59	**0.70 ± 2.14**
	$V_{95\%}$ (%)	-1.79 ± 4.00	**-1.68 ± 1.82**	-1.09 ± 2.41	**-0.42 ± 1.12**
PTV	$D_{95\%}$ (Gy)	2.59 ± 3.82	**-1.70 ± 2.21**	1.75 ± 3.55	**0.42 ± 1.83**
	$D_{98\%}$ (Gy)	2.20 ± 3.96	**-1.95 ± 2.77**	0.92 ± 4.72	**-0.71 ± 2.12**
	$V_{95\%}$ (%)	-1.72 ± 6.60	**-1.63 ± 1.71**	-0.81 ± 2.55	**-0.22 ± 1.87**
Rectum	D_{mean} (Gy)	0.48 ± 6.03	**-0.69 ± 5.67**	0.71 ± 5.14	**-0.82 ± 4.45**
	D_{max} (Gy)	6.08 ± 4.56	**1.57 ± 3.16**	5.86 ± 4.51	**-0.07 ± 3.19**
	$V_{40\text{Gy}}$ (%)	0.76 ± 6.49	**-0.47 ± 5.68**	1.08 ± 6.61	**-1.12 ± 5.13**
Left Femoral	D_{mean} (Gy)	0.16 ± 10.22	**-0.72 ± 6.34**	0.35 ± 10.10	**-0.15 ± 6.60**
	D_{max} (Gy)	1.24 ± 7.65	**-0.71 ± 5.46**	1.19 ± 6.64	**0.79 ± 6.21**
	$V_{30\text{Gy}}$ (%)	0.13 ± 1.01	**-0.71 ± 3.21**	0.45 ± 4.57	**-0.32 ± 3.22**
Right Femoral	D_{mean} (Gy)	1.30 ± 7.96	**-0.73 ± 6.96**	1.38 ± 8.03	**-0.69 ± 5.96**
	D_{max} (Gy)	2.74 ± 5.62	**-0.09 ± 5.11**	1.85 ± 4.90	**1.23 ± 4.14**
	$V_{30\text{Gy}}$ (%)	0.88 ± 2.40	**-0.32 ± 1.15**	0.12 ± 1.74	**-0.21 ± 1.16**

Table 1. Comparison of dose metrics for UNETR and 3D U-Net across training stages. Bold values indicate the best performance for each model. Each value represents the difference Predicted − Ground Truth.

Dose Threshold	UNETR		3D U-Net	
	1st Stage	2nd Stage	1st Stage	2nd Stage
10%	82.85 ± 7.30	**85.08 ± 6.37**	84.70 ± 5.93	**86.78 ± 4.63**
90%	41.98 ± 25.74	**80.40 ± 10.72**	56.46 ± 32.68	**90.50 ± 7.28**

Table 2. Gamma pass rates (%) for different dose thresholds across models. Bold values indicate the best performance for each architecture.

exhibited deviations from clinical goals. The physics-guided stage introduced a clinically relevant training objective, mitigating multi-arc redundancy and guiding the network toward a consistent dose representation. As a result, PTV and CTV dose predictions aligned more closely with the ground truth, while OAR doses remained slightly below or marginally above it, demonstrating the clinical feasibility of our approach.

Regarding the *Deep RT Planner*, the 3D U-Net outperformed UNETR, likely due to the latter's need for larger training datasets [9]. Both architectures had similar parameter counts, but the relatively shallow transformer may have limited UNETR's generalization.

The gamma pass rate at the 10% dose threshold showed minimal differences between the two physics-based models, possibly explaining the variability in DVH metrics for femoral heads and the UNETR's slight advantage in this specific metric. Conversely, the superior gamma pass rate in the high-dose region supports the observed improvement in PTV/CTV DVH metrics, underscoring the value of dose-aware supervision.

Scalability remains a challenge for broader clinical adoption. With more diverse training data, this approach could generalize to an adaptive RT agent for multiple cancer sites and treatment stages. Additionally, integrating DVH-aware loss functions [17] or leveraging geometry-aware architectures [4] may help address these challenges.

References

1. Ahnesjö, A.: Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media. Medical Physics **22**(3), 379–388 (1995)
2. Bedford, J.L.: Treatment planning for volumetric modulated arc therapy. Medical Physics **36**(11), 5128–5138 (2009). https://doi.org/10.1118/1.3240488
3. Bortfeld, T.: Imrt: a review and preview. Physics in Medicine Biology **51**(13), R363–R379 (2006). https://doi.org/10.1088/0031-9155/51/13/R21
4. Bronstein, M.M., Bruna, J., Cohen, T., Veličković, P.: Geometric deep learning: Grids, groups, graphs, geodesics, and gauges (2021), https://arxiv.org/abs/2104.13478
5. Charles M. Washington, Dennis T. Leaver, M.T.: Washington & Leaver's Principles and Practice of Radiation Therapy. Mosby, 5th edn. (2020)
6. Cho, K., van Merrienboer, B., Bahdanau, D., Bengio, Y.: On the properties of neural machine translation: Encoder-decoder approaches (2014), https://arxiv.org/abs/1409.1259
7. Cook, S.: CUDA Programming: A Developer's Guide to Parallel Computing with GPUs. Elsevier (2013)
8. Delaney, G., Jacob, S., Featherstone, C., Barton, M.: The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines. Cancer **104**(6), 1129–1137 (2005)
9. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2021), https://arxiv.org/abs/2010.11929
10. Elekta: Monaco treatment planning system
11. Faroughi, S.A., Pawar, N., Fernandes, C., Raissi, M., Das, S., Kalantari, N.K., Mahjour, S.K.: Physics-guided, physics-informed, and physics-encoded neural networks in scientific computing (2023), https://arxiv.org/abs/2211.07377
12. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016)
13. Hatamizadeh, A., Yin, H., Kautz, J., Molchanov, P.: Unetr: Transformers for 3d medical image segmentation. arXiv preprint arXiv:2103.10504 (2022)
14. Heilemann, G., Zimmermann, L., Schotola, R., Lechner, W., Peer, M., Widder, J., Goldner, G., Georg, D., Kuess, P.: Generating deliverable dicom rt treatment plans for prostate vmat by predicting mlc motion sequences with an encoder-decoder network. Medical Physics **50**(8), 5088–5094 (Aug 2023). https://doi.org/10.1002/mp.16545
15. Hrinivich, W.T., Bhattacharya, M., Mekki, L., McNutt, T., Jia, X., Li, H., Song, D.Y., Lee, J.: Clinical vmat machine parameter optimization for localized prostate cancer using deep reinforcement learning. Medical Physics **51**(6), 3972–3984 (2024). https://doi.org/10.1002/mp.17100
16. Hrinivich, W.T., Lee, J.: Artificial intelligence-based radiotherapy machine parameter optimization using reinforcement learning. Medical Physics **47**(12), 6140–6150 (2020). https://doi.org/10.1002/mp.14544

17. Jhanwar, G., Dahiya, N., Ghahremani, P., Zarepisheh, M., Nadeem, S.: Domain knowledge driven 3d dose prediction using moment-based loss function. Physics in Medicine amp; Biology **67**(18), 185017 (2022). https://doi.org/10.1088/1361-6560/ac8d45
18. Lemus, O.M.D., Cao, M., Cai, B., Cummings, M., Zheng, D.: Adaptive radiotherapy: Next-generation radiotherapy. Cancers (Basel) **16**(6) (2024). https://doi.org/10.3390/cancers16061206
19. Lillicrap, T., Hunt, J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D.: Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971 (2019)
20. Liu, F., Jansson, N., Podobas, A., Fredriksson, A., Markidis, S.: Accelerating radiation therapy dose calculation with nvidia gpus (2021), https://arxiv.org/abs/2103.09683
21. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization (2019), https://arxiv.org/abs/1711.05101
22. Ni, Y., Chen, S., Hibbard, L., Voet, P.: Fast vmat planning for prostate radiotherapy: dosimetric validation of a deep learning-based initial segment generation method. Physics in Medicine Biology **67**(15) (jul 2022). https://doi.org/10.1088/1361-6560/ac80e5
23. Otto, K.: Volumetric modulated arc therapy: Imrt in a single gantry arc. Medical Physics **35**(1), 310–317 (2008). https://doi.org/10.1118/1.2818738
24. Philips Healthcare: Pinnacle treatment planning system
25. Shen, C., Nguyen, D., Chen, L., Gonzalez, Y., McBeth, R., Qin, N., Jiang, S.B., Jia, X.: Operating a treatment planning system using a deep-reinforcement learning-based virtual treatment planner for prostate cancer intensity-modulated radiation therapy treatment planning. Medical Physics **47**(6), 2329–2336 (2020). https://doi.org/10.1002/mp.14114
26. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, 2nd edn. (2018)
27. Unkelbach, J., Chan, T., Bortfeld, T.: Accounting for range uncertainties in the optimization of intensity modulated proton therapy. Physics in Medicine & Biology **60**(2), 623 (2015)
28. Vandewinckele, L., Reynders, T., Weltens, C., Maes, F., Crijns, W.: Deep learning based mlc aperture and monitor unit prediction as a warm start for breast vmat optimisation. Physics in Medicine & Biology (2023)
29. Verhaegen, F., Seuntjens, J.: Monte carlo modelling of external radiotherapy photon beams. Physics in Medicine & Biology **48**(21), R107–R164 (2003)
30. de Vries, L., Van Herten, R.L.M., Hoving, J.W., Isgum, I., Emmer, B., Majoie, C.B., Marquering, H., Gavves, S.: Accelerating physics-informed neural fields for fast ct perfusion analysis in acute ischemic stroke. In: Burgos, N., Petitjean, C., Vakalopoulou, M., Christodoulidis, S., Coupe, P., Delingette, H., Lartizien, C., Mateus, D. (eds.) Proceedings of The 7th International Conference on Medical Imaging with Deep Learning. Proceedings of Machine Learning Research, vol. 250, pp. 1606–1626. PMLR (03–05 Jul 2024), https://proceedings.mlr.press/v250/vries24a.html
31. Wendling, M., Zijp, L.J., McDermott, L.N., Smit, E.J., Sonke, J.J., Mijnheer, B.J., van Herk, M.: A fast algorithm for gamma evaluation in 3d. Medical Physics **34**(5), 1647–1654 (2007). https://doi.org/10.1118/1.2721657
32. Weng, L.: Reward hacking in reinforcement learning. Lil'Log (2024), https://lilianweng.github.io/posts/2024-11-28-reward-hacking/

33. Witte, M., Sonke, J.J.: A deep learning based dynamic arc radiotherapy photon dose engine trained on monte carlo dose distributions. Physics and Imaging in Radiation Oncology **30**, 100575 (2024). https://doi.org/10.1016/j.phro.2024.100575
34. Özgün Çiçek, Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation (2016), https://arxiv.org/abs/1606.06650