Technical Report from CWRU-UH-Cornell on GDP-HMM Challenge

Justin Namuk Kim[1], Sunwoo Kwak[2], Theodore Arsenault[1,3]
[1] Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA
[2] Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA
[3] University Hospitals, Seidman Cancer Center, Cleveland, OH, USA

1. Motivation

Current state-of-the-art dose-distribution prediction falls into two broad families:

1. 3D segmentation backbone – e.g. cascaded U-net in the OpenKBP "C3D" [1] method; the organizer's baseline for GDP-HMM [2] uses a MedNeXt-Base network [3].
2. 2D diffusion-based generators – DoseDiff [4], DiffDP [5], etc.

Given the limited challenge timeline and the lack of decisive evidence demonstrating one method is better than the other due to nature of stacking multiple innovations in a single approach, we chose a straightforward 3D segmentation backbone strategy. Our goal was to identify a strong segmentation backbone architecture, keep the data preprocessing and loading *identical as provided in the challenge*.

2. Experiments and Findings

2.1 Experiment 1 – SegMamba backbone [6]

Hypothesis: With larger parameters, network learns richer representations of complex inputs, thus SegMamba (66M) segmentation network should outperform the MedNeXt-base (10M)

Outcome: best validation loss 0.2231 (MAE=2.561, ID: 2475) at epoch 388 – worse than the organizer's baseline. Limited time precluded an in-depth ablation, but two explanations are plausible: (i) Specialized segmentation network may be ill-suited to sharply peaked dose distribution prediction; (ii) Heterogenous multi-input information is not well-utilized by SegMamba

2.2 Experiment 2 – MedNeXt-Large (kernel 5)

We reverted to MedNeXt and selected "Large, k=5" variant, strongest configuration in the original paper.

Hypothesis: MedNeXt-Large (k5) should outperform the MedNeXt-base (10M)

Outcome: best validation loss 0.2090 (MAE=2.411, ID: 2566) at epoch 308 – a clear improvement over both the baseline (MAE=2.54) and SegMamba.

2.3 Experiment 3 – Physics-informed composite loss

We prototyped a combination of Dice, L1, MAE loss to improve PTV coverage and OAR sparing. Implementation bugs and time pressure prevented a stable run; the idea was shelved.

2.4 Experiment 4 – Smooth-L1 loss

Observation: validation L1 plateaued near 0.20 with smaller and slower decrease. We replaced L1 with Smooth-L1(beta=0.25) to emphasize sub-2.4 Gy voxels while retaining robustness to larger errors.

Outcome*:* best validation L1 loss 0.2116, Smooth-L1 loss 0.1115 (MAE=2.405). We expect a lower beta (=0.15) would have pushed MAE further down, but the validation phase has already been completed. This experiment is our team's final submission for the challenge.

3. Future work

Our immediate next step is to prototype a **cascaded MedNeXt** workflow, mirroring the two-stage C3D design: a coarse-resolution MedNeXt generates an initial dose volume, which a second, finer-resolution MedNeXt then refines. Because MedNeXt accepts an arbitrary number of channels, this cascade can be assembled with minimal engineering effort and, judging from C3D's success, should push voxel-wise MAE below our single-stage result.

Yet MAE alone is not a perfect proxy for clinical plan quality. In our own study the **SegMamba model recorded the top plan-quality score (Metric-2 = 118.44) despite a higher MAE (2.561)**, underscoring that backbone choice—and perhaps the broader CNN-versus-diffusion paradigm—affects deliverability metrics differently from mean absolute error. A systematic sweep that disentangles MAE from Metric-2 across MedNeXt cascades and representative diffusion models is therefore warranted.

Another promising lever is the **use of distance maps** in place of binary masks. DoseDiff attributes much of its gain to this richer geometric encoding. In a sub-study of Experiment 2 we replaced OARs with distance maps and reached a best validation loss of 0.2246 (epoch 322)—encouraging but not yet superior to the binary baseline. A more thorough integration of distance-encoded PTV and OAR channels into MedNeXt can be further investigated.

Finally, future work should examine the eight-channel input stack itself—some elements may be redundant—and expand **physics-informed loss** so the optimizer is driven not only by MAE but also by structural fidelity and plan-quality surrogates. Together, these directions aim to close the remaining gap between algorithmic accuracy and clinically actionable dose plans while keeping inference costs modest compared with diffusion-based pipelines

4. Implementation Details

All experiments were executed on a Linux HPC cluster, Python 3.11.3, PyTorch 2.3.1 compiled with CUDA 12.1. The job used 4 A40 GPUs (48GB VRAM each). Apart from our efficient dataloader, the training pipeline is the same as organizer's implementation ("train_lightning.py"). Memory limited train batch size to 2, but with efficient dataloader, was able to increase it to 4. The default train budget was set to 120 hours or max epoch 400; for the final run ("Experiment 5") extended to 144 hours or max epoch 600.

5. References

[1] S. Liu, J. Zhang, T. Li, H. Yan, and J. Liu, "Technical Note: A cascade 3D U-Net for dose prediction in radiotherapy," *Medical Physics*, vol. 48, no. 9, pp. 5574–5582, 2021, doi: 10.1002/mp.15034.

[2] R. Gao *et al.*, "Automating High Quality RT Planning at Scale," Mar. 11, 2025, *arXiv*: arXiv:2501.11803. doi: 10.48550/arXiv.2501.11803.

[3] S. Roy *et al.*, "MedNeXt: Transformer-driven Scaling of ConvNets for Medical Image Segmentation," Jun. 02, 2024, *arXiv*: arXiv:2303.09975. doi: 10.48550/arXiv.2303.09975.

[4] Y. Zhang, C. Li, L. Zhong, Z. Chen, W. Yang, and X. Wang, "DoseDiff: Distance-aware Diffusion Model for Dose Prediction in Radiotherapy," Mar. 28, 2024, *arXiv*: arXiv:2306.16324. doi: 10.48550/arXiv.2306.16324.

[5] Z. Feng *et al.*, "DiffDP: Radiotherapy Dose Prediction via a Diffusion Model," vol. 14225, 2023, pp. 191–201. doi: 10.1007/978-3-031-43987-2_19.

[6] Z. Xing, T. Ye, Y. Yang, G. Liu, and L. Zhu, "SegMamba: Long-range Sequential Modeling Mamba For 3D Medical Image Segmentation," Sep. 15, 2024, *arXiv*: arXiv:2401.13560. doi: 10.48550/arXiv.2401.13560.