Generative AI helps Radiotherapy Planning with User Preference (DEMO)

Riqiang Gao[1] Simon Arberet[1] Martin Kraus[1] Han Liu[1] Wilko F.A.R. Verbakel[2] Dorin Comaniciu[1]
Florin C. Ghesu[1] Ali Kamen[1]

[1]Digital Technology and Innovation, Siemens Healthineers
[2]Varian Medical Systems, Siemens Healthineers

NeurIPS 2025 GenAI4Health

Paper link:

Motivation: Background and Challenges in RT Planning

RT Planning Background:

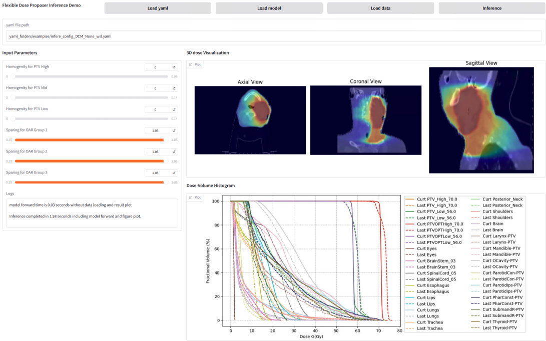
- Radiotherapy (RT) is one of the most common cancer treatments (suitable for about 50% cancer patients)
- RT planning is complex and time-consuming
- RT planning involves a multidisciplinary team and can be subjective

Current Limitations:

- RT planning varies significantly across institutions
- Existing models lack user interaction to balance PTV/OAR trade-offs
- RapidPlan™ is DVH-based, missing spatial dose details
- RapidPlan requires institution-specific models

Our Solution:

- Interactive AI dose prediction
- User-defined preferences via sliders
- Integration in treatment planning system



SIEMENS Healthineers

Key Metrics:

- Homogeneity Index (HI):

$$HI = \frac{D_{05} - D_{95}}{D_{50}}$$

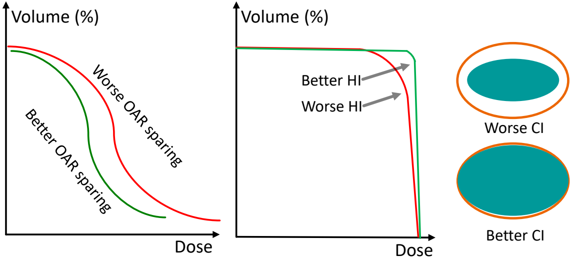
- Conformity Index (CI):

$$CI = \frac{V_{\text{covered}}}{V_{\text{PTV}}}$$

- OAR Sparing: Mean doses to organs at risk

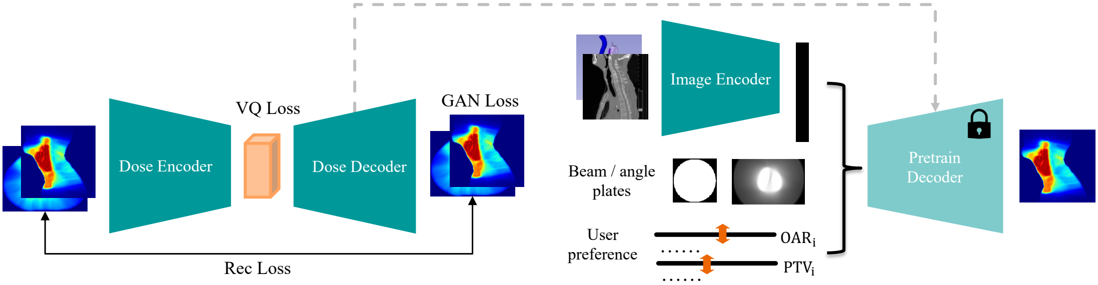
The Trade-off:

- Better PTV homogeneity **or** Better OAR sparing
- Different planners have different preferences
- Traditional models cannot adapt



Innovation:

Pioneering dose prediction model with interactive sliders enabling real-time customization of trade-offs



(a) Stage 1: Decoder Pretraining

(b) Stage 2: Flexible Dose Prediction

Stage I: Foundational Decoder

- VQ-VAE architecture
- Pre-trained on 31K doses
- Stabilizes training with realistic dose distributions

Stage II: Flexible Prediction

- Multi-conditional inputs (CT, structures, preferences)
- random sampling for preferences during training
- One-step generation via GANs (fast inference)

Training Strategy: Two-Stage Loss Functions

SIEMENS Healthineers

Stage I: Foundational Decoder

$$\mathcal{L}_{\text{stage1}} = \underbrace{\mathbb{E}_i[\|x_i - \hat{x}_i\|]}_{\text{Reconstruction}} + \beta L_{vq} + L_{adv}(x, \hat{x}) + \underbrace{\lambda \cdot \log\left(\mathbb{E}_{i<j}\left[\exp(-t\|\hat{z}_i - \hat{z}_j\|^2)\right]\right)}_{\text{Uniformity}}$$

Stage II: Flexible Prediction

$$\mathcal{L}^{(i)}_{\text{stage2}} = \underbrace{\|x_i - \hat{x}_i\|}_{\text{Image Recon.}} + \underbrace{\|z_i - \hat{z}_i\|}_{\text{Latent Recon.}} + L_{adv}(x_i, \hat{x}_i) + \mathcal{L}^{(i)}_{\text{obj}}$$

$$\mathcal{L}^{(i)}_{\text{obj}} = \underbrace{\|\tilde{h} - \hat{h}\|}_{\text{ptv HI preference}} + \underbrace{\|p - \hat{p}\|}_{\text{ptv dose alignment}} + \underbrace{\|\tilde{w} \cdot u_{\text{oar}} - \hat{u}_{\text{oar}}\|}_{\text{oar-sparing preference}}$$

SIEMENS Healthineers

Dataset:

- 6 cohorts of head-and-neck cancer
- Total: 820 training, 103 validation, 113 test cases
- Stage I pre-trained on 31K doses

Baseline:

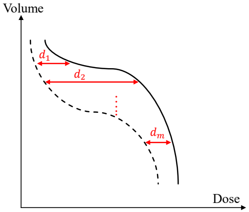
- Varian RapidPlan^{TM} (high-quality model)

Evaluation Metrics:

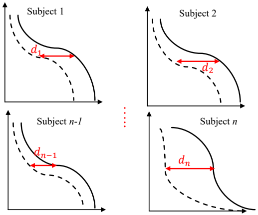
- **DVH Estimation Accuracy:** Expected vs. achieved DVH differences
- **Inter-patient** and **Inter-patient** highlight different perspectives
- Quality of deliverable plans

Data Distribution:

Cohort	0	1	2	3	4	5
Train	370	147	128	103	52	20
Valid	48	17	15	14	8	1
Test	54	19	17	12	7	4



$D_{intra} = [d_1, d_2, \ldots, d_m]$



$D_{inter} = [d_1, d_2, \ldots, d_n]$

Results: Superior DVH Estimation Accuracy

Intra-patient Differences:

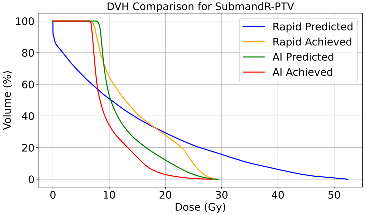
- **15/15** OARs: FDP outperforms RapidPlan
- (check tables in manuscript)

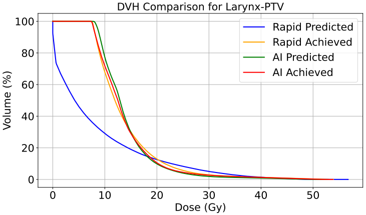
Inter-patient Differences:

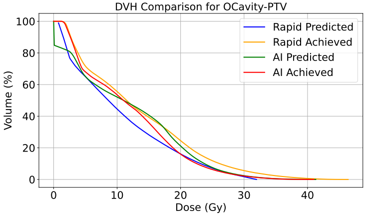
- **12/15** OARs: FDP shows lower variability
- (check tables in manuscript)

Key Insight:

FDP provides more robust and reliable DVH estimations







Some examples about expected vs. achieved DVHs for RapidPlan and our FDP model

Results: Structure-wise Plan Quality Comparison

Table: Percentages of better, worse, similar when compare the FDP to RapidPlan (per structure)

OAR	SpinalCor 05	Larynx-PTV	Lips	Mandible-PTV	OCavity-PTV	ParotidCon-PTV	ParotidIps-PTV	Esophagus
better	47.50	30.00	47.50	31.25	64.56	32.50	48.68	30.23
worse	5.00	21.43	0.00	1.25	1.27	6.25	5.26	4.65
similar	47.50	48.57	52.50	67.50	34.18	61.25	46.05	65.12

OAR	SubmandL-PTV	Shoulders	SubmandR-PTV	Posterior Neck	PharConst-PTV	BrainStem 03	Trachea	**OAR count**
better	59.57	0.00	71.11	12.50	56.16	7.50	51.16	**14**
worse	2.13	0.00	6.67	6.25	2.74	3.75	0.00	**0**
similar	38.30	100.00	22.22	81.25	41.10	88.75	48.84	-

PTV	HI (PTVHigh)	CI (PTVHigh)	HI (PTVMid)	CI (PTVMid)	HI (PTVLow)	CI (PTVLow)	**PTV count**
better	0.00	0.00	0.00	4.00	1.43	4.29	**1**
worse	0.00	0.00	0.00	4.00	0.00	4.29	**0**
similar	100.00	100.00	100.00	92.00	98.57	91.43	-

Thresholds: 1 Gy for OARs; 0.015 for PTV indices (HI & CI). **bold** *= category counts.*

Summary: OAR: 14 better / 0 worse PTV: 1 better / 0 worse

SIEMENS Healthineers

Top OAR Improvements:

- SubmandR-PTV: **71%** better
- OCavity-PTV: **65%** better
- SubmandL-PTV: **60%** better

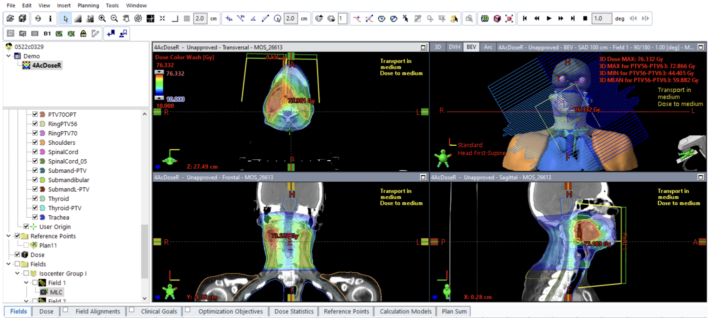
Clinical Significance:

- Reduced toxicity for patients
- Maintained PTV coverage
- Better quality of life outcomes

Key Takeaway

FDP delivers superior quality plans in Eclipse treatment planning system.

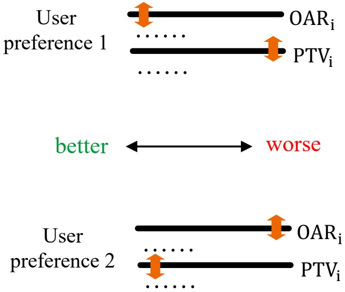
Eclipse Treatment Planning System Integration

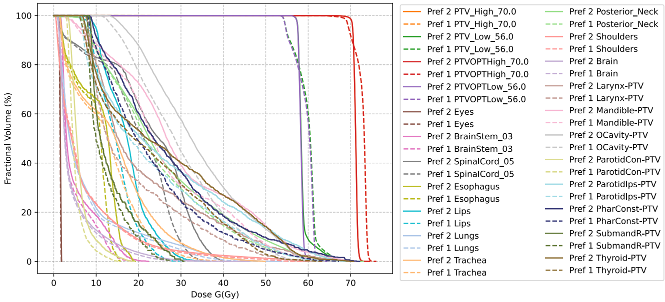


Group	Field ID	Technique	Machine/Energy	MLC	Field Weight	Scale	Gantry Rtn [deg]	Coll Rtn [deg]	Couch Rtn [deg]	Wedge	Field X [cm]	X1 [cm]	X2 [cm]	Field Y [cm]	Y1 [cm]	Y2 [cm]	X [cm]	Y [cm]	Z [cm]	Calculated SSD [cm]	MU	Ref. D [Gy]
I	Field 1	Arc Therapy-I	TrueBeam - 6X	VMAT	0.814	IEC61217	179.0 CCW 181.0	30.0	0.0	None	15.0	-8.5	+6.5	16.2	-9.5	+6.7	-0.94	-6.43	27.49	93.7	164	
I	Field 2	Arc Therapy-I	TrueBeam - 6X	VMAT	0.981	IEC61217	181.0 CW 179.0	30.0	0.0	None	15.0	-10.5	+4.5	16.2	-9.5	+6.7	-0.94	-6.43	27.49	93.7	198	
I	Field 3	Arc Therapy-I	TrueBeam - 6X	VMAT	0.886	IEC61217	179.0 CCW 181.0	330.0	0.0	None	15.0	-4.5	+10.5	16.2	-9.5	+6.7	-0.94	-6.43	27.49	93.7	179	
I	Field 4	Arc Therapy-I	TrueBeam - 6X	VMAT	0.822	IEC61217	181.0 CW 179.0	330.0	0.0	None	15.0	-6.4	+8.6	16.2	-9.5	+6.7	-0.94	-6.43	27.49	93.7	166	

FDP predictions are optimized in Eclipse to generate deliverable clinical plans

Demo Feature: Interactive Preference Control





Preference 1 (P1): OAR-focused

- Prioritize OAR sparing
- Accept higher PTV dose heterogeneity
- Lower OAR mean/max doses

Preference 2 (P2): PTV-focused

- Prioritize PTV homogeneity
- Accept higher OAR doses
- Better PTV dose distribution

Real-time adaptation: Model responds to slider adjustments within seconds

Key Contributions:

1. Novel two-stage framework with foundational decoder
2. First interactive dose prediction model with real-time user preference sliders.
3. Clinical integration with treatment planning systems
4. Superior performance vs. RapidPlan

Limitations:

- Currently focused on head-and-neck cancer
- clinical validation is limited

Future Directions:

- Extend to other cancer treatment sites
- More rigorous clinical validation
- Integration with automated planning pipelines
- Explore diffusion-based alternatives

Impact

Significant step toward personalized AI-assisted radiotherapy planning

Demo Video:

https://huggingface.co/Jungle15/DoseProposerDemo

Disclaimer: The information in the slides is based on research results that are not commercially available. Future commercial availability cannot be guaranteed.

Thank You!

Questions?

Contact: riqiang.gao@siemens-healthineers.com

Demo Video:

Paper: