GRADIENT: Generalized Radiotherapy Dose Prediction via Text Conditioning

Sangwook Kim, B.Eng.[1,2,5]; Aly Khalifa, PhD[1,2,3]; Yuan Gao, MSc[1,2,4,5]; Thomas G Purdie, PhD[1,2,3]; Chris McIntosh, PhD[1,2,3,4,5]
[1]University of Toronto, [2]University Health Network, [3]Princess Margaret Cancer Centre,
[4]Peter Munk Cardiac Centre, [5]Vector Institute

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

ABSTRACT

Purpose:
To develop GRADIENT, a unified protocol-aware radiotherapy dose prediction model that generalizes across heterogeneous treatment protocols.

Methods:
GRADIENT incorporates text-derived protocol conditioning and image-text pretraining to align CT-based anatomical features with protocol-specific representations. The model was evaluated on 2,296 patients from 17 protocols and 7 anatomical sites across seen, zero-shot unseen, few-shot, and full-shot adaptation settings using DVH-MAE, HI-MAE, and voxel-wise MAE.

Results:
GRADIENT consistently outperformed anatomy-only and conventional conditioning baselines. Compared with the NoCondition baseline, it reduced DVH-MAE, HI-MAE, and voxel-wise MAE by 20.9%, 25.0%, and 10.8% for seen protocols, and by 12.0%, 27.1%, and 10.6% in zero-shot unseen protocols. It also achieved lower aggregate ranks across few-shot settings.

Conclusion:
Protocol-aware image-text pretraining improves generalized dose prediction across protocol shifts and limited-data scenarios, supporting GRADIENT as a scalable framework for multi-protocol radiotherapy dose prediction.

CONTACT

Sangwook Kim
University of Toronto
sangwook.kim@mail.utoronto.ca
+1 437 980 5986

INTRODUCTION

- Radiotherapy dose prediction models are commonly trained for specific anatomical sites or treatment protocols, limiting their scalability across heterogeneous clinical workflows.
- Clinical dose distributions depend not only on patient anatomy, but also on protocol-level factors such as treatment site, prescription dose, and fractionation.
- We propose **GRADIENT**, **a unified protocol-aware dose prediction framework** that **incorporates text-derived treatment protocol information** to improve generalization across seen, unseen, and low-data protocols.

MATERIALS AND METHODS

- **GRADIENT** uses CT images, PTV/OAR masks, and structured protocol text to predict 3D radiotherapy dose distributions.
- Protocol information was encoded using text-derived conditioning vectors and integrated into a conditional 3D dose prediction network.
- Image-text pretraining was used to align anatomical representations with protocol-specific embeddings before supervised dose prediction.
- The model was trained and evaluated on **2,296 patients** from **17 treatment protocols** across **7 anatomical sites**.
- Performance was assessed in seen-protocol, zero-shot unseen-protocol, few-shot, and full-shot adaptation settings using normalized DVH MAE, HI MAE, and foreground voxel-wise MAE.

RESULTS

- **GRADIENT** consistently outperformed anatomy-only and conventional conditioning baselines across evaluation settings.
- In seen protocols, GRADIENT reduced normalized DVH MAE by 20.9%, HI MAE by 25.0%, and voxel-wise MAE by 10.8% compared with NoCondition.
- In zero-shot unseen protocols, GRADIENT reduced normalized DVH MAE by 12.0%, HI MAE by 27.1%, and voxel-wise MAE by 10.6%.
- Across few-shot and full-shot adaptation settings, **GRADIENT achieved lower aggregate ranks than baseline models**, suggesting improved robustness under protocol shift and limited-data conditions.
- Qualitative results showed closer agreement with clinical dose distributions and reduced spatial error compared with NoCondition.

Overall framework

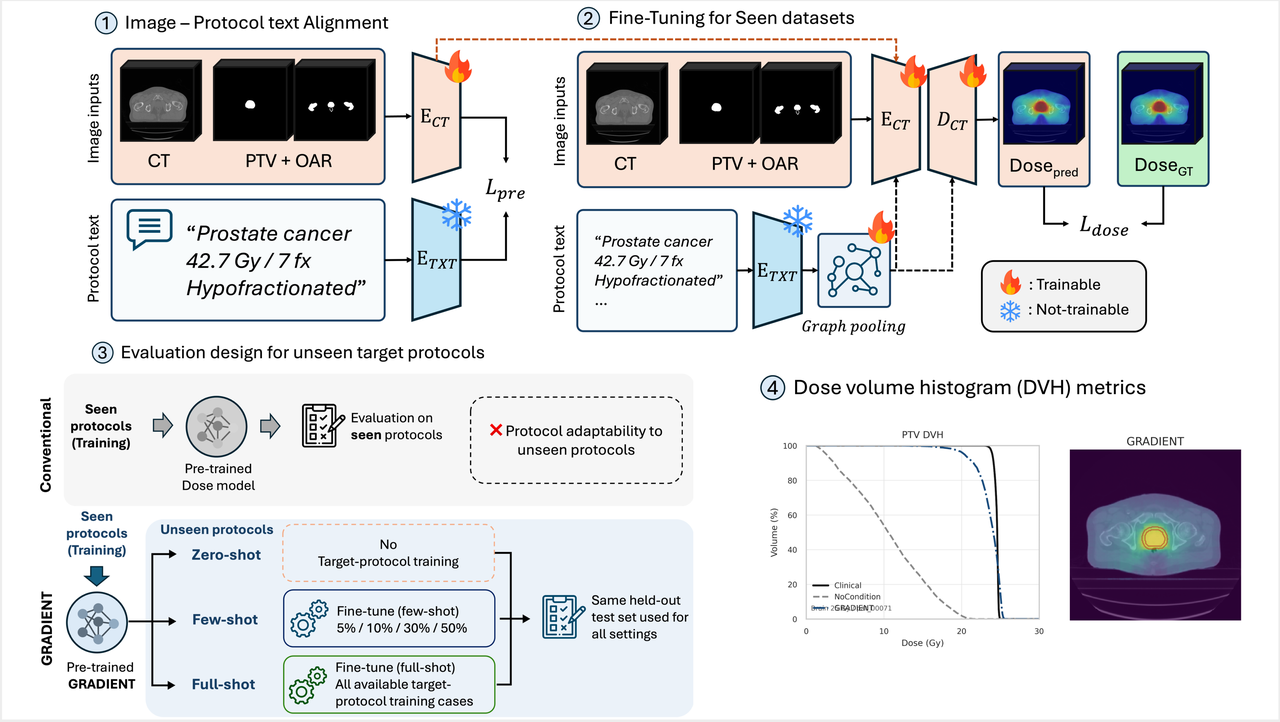

Table 1. Aggregate normalized DVH-MAE across evaluation settings. Values show mean and 95% bootstrap confidence intervals over protocols. Lower is better; bold indicates the best model.

	Separate	NoCondition	GRADIENT
Seen	0.376 (0.295, 0.456)	0.090 (0.051, 0.138)	**0.072 (0.041, 0.105)**
Zero-shot	N/A	0.148 (0.103, 0.190)	**0.130 (0.092, 0.168)**
5%	0.301 (0.213, 0.413)	0.187 (0.121, 0.256)	**0.161 (0.099, 0.223)**
10%	0.302 (0.213, 0.415)	0.186 (0.117, 0.262)	**0.167 (0.096, 0.245)**
30%	0.302 (0.214, 0.414)	0.141 (0.086, 0.201)	**0.135 (0.079, 0.196)**
50%	0.301 (0.214, 0.411)	0.137 (0.082, 0.200)	**0.118 (0.067, 0.179)**
Full	0.297 (0.212, 0.402)	0.106 (0.070, 0.142)	**0.091 (0.059, 0.125)**

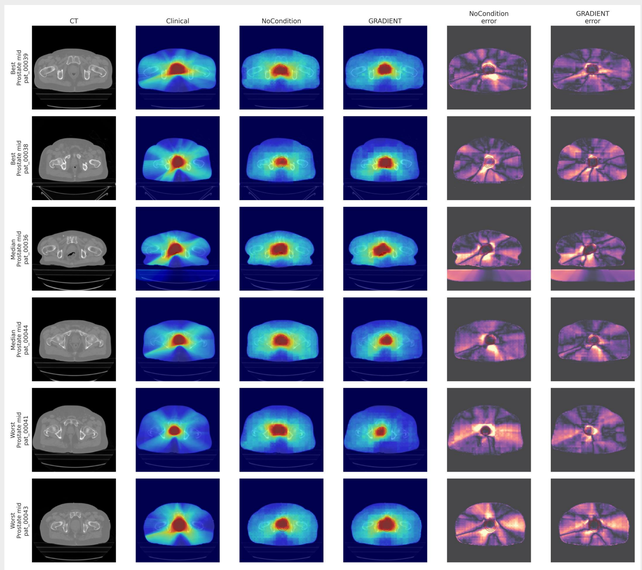

Figure 1. Representative qualitative results for prostate 3600 cGy / 6 fractions. Clinical dose, NoCondition, GRADIENT, and error maps are shown for best, median, and worst cases. GRADIENT shows improved agreement with clinical dose and reduced spatial error.

DISCUSSION

- **GRADIENT** suggests that treatment protocol information provides **complementary guidance beyond patient anatomy alone for radiotherapy dose prediction.**
- Text-derived protocol conditioning helped the model adapt across heterogeneous prescriptions, fractionation schemes, anatomical sites, and protocol-shifted settings.
- The strongest benefits were observed in **unseen and limited-data settings, supporting the role of protocol-aware pretraining for improving generalization**.
- Qualitative error maps further suggest that GRADIENT produces dose distributions that more closely resemble clinical plans compared with anatomy-only prediction.

CONCLUSIONS

- **GRADIENT** improves generalized radiotherapy dose prediction by integrating anatomical information with structured treatment protocol text.
- Protocol-aware image-text pretraining enables more robust prediction across **seen, unseen, and low-data treatment protocols**, supporting scalable multi-protocol AI-assisted treatment planning.

REFERENCES

1. Kim S, Gao Y, Purdie TG, McIntosh C. TREAT: A Unified Text-Guided Conditioned Deep Learning Model for Generalized Radiotherapy Treatment Planning. MICCAI 2025; 614-624.
2. Cai Y, et al. Swin UNETR: A three-dimensional medical image segmentation network combining vision transformer and convolution. BMC Med Inform. Decis. Mak. 2023;23(1):33.