Conditional generative adversarial network for dose distribution prediction in hepatocellular carcinoma proton beam therapy: Improved accuracy in steep dose gradient regions

Shuta Ogawa, M.S.[1, 2], Noriyuki Kadoya, Ph.D.[2], Takahiro Kato, Ph.D.[3], Ryohei Kato, M.S.[1, 2], Ryota Tozuka, M.S.[2, 4], Yuki Narita, Ph.D.[1], Sho Oyama, M.S.[1], Masao Murakami, M.D., Ph.D.[5], Keiichi Jingu, M.D., Ph.D.[2]

1. Department of Radiation Physics and Technology, Southern Tohoku Proton Therapy Center, Japan 2. Department of Radiation Oncology, Tohoku University Graduate School of Medicine, Japan
3. Department of Radiological Sciences, School of Health Sciences, Fukushima Medical University, Japan 4. Department of Therapeutic Radiology, University of Yamanashi, Japan
5. Department of Radiation Oncology, Southern Tohoku Proton Therapy Center, Japan

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

ABSTRACT

Purpose:
Accurate deep-learning dose prediction in proton beam therapy (PBT) is challenging because of the steep dose gradients inherent to particle beams. Conventional pixel-wise losses yield over-smoothed predictions, and fixed dose-gradient penalties help only marginally. We developed a conditional generative adversarial network (cGAN) that dynamically penalizes high-frequency component discrepancies.

Methods:
We developed a modified Pix2Pix model for three-dimensional dose prediction in PBT for hepatocellular carcinoma (HCC). The generator was a U-Net with attention gates and instance normalization; a PatchGAN discriminator preserved high-frequency detail through adversarial training. A baseline U-Net with composite losses served as the comparison. Both models predicted prescription-normalized dose from CT images and target/organs-at-risk structures alone. Using 172 HCC patients, accuracy was evaluated as the mean absolute error (MAE) of dose-volume metrics for the clinical target volume (CTV), planning target volume (PTV), and normal liver, plus the 10% isodose volume representing the beam path. Models were compared by paired t-tests.

Results:
The cGAN significantly improved CTV $D_{95\%}$ (MAE, 0.48 vs. 1.08 Gy(RBE); p = 0.034), with the same non-significant trend for $D_{98\%}$ and $D_{2\%}$. No significant differences were found for the PTV or normal liver. For the 10% isodose volume, the U-Net gave a lower MAE than the cGAN (51.0 vs. 61.3 cm³; p = 0.48), reflecting higher low-dose accuracy.

Conclusions:
By capturing the steep dose gradients characteristic of PBT, the cGAN improved CTV coverage prediction, though the U-Net remained better in the gradually varying low-dose region. Predicting dose from CT and structures alone, the framework may help standardize planning and enable virtual dose evaluation where PBT is unavailable. Future work will improve low-dose accuracy for balanced predictions across all dose levels.

CONTACT

[**name**] Shuta Ogawa
[**organization**] Southern Tohoku Proton Therapy Center
[**address**] 172, Yatsuyamada, Koriyama, Fukushima 963-8052, Japan.
[**email**] shuta.ogawa125@gmail.com

INTRODUCTION

In proton beam therapy (PBT) for hepatocellular carcinoma (HCC), beam arrangement depends on tumor anatomy and respiratory motion, making planning time-consuming and planner-dependent.

Conventional CNNs with pixel-wise losses over-smooth the steep dose gradients of PBT, losing high-frequency detail (Figure 1).

Recent 3D U-Nets with dose-gradient-aware composite losses reduce this over-smoothing, yet accurately predicting these steep gradients remains challenging.

We therefore developed a conditional generative adversarial network (cGAN) that dynamically penalizes high-frequency errors and compared it with such a 3D U-Net, with both models predicting dose from CT images and structures alone.

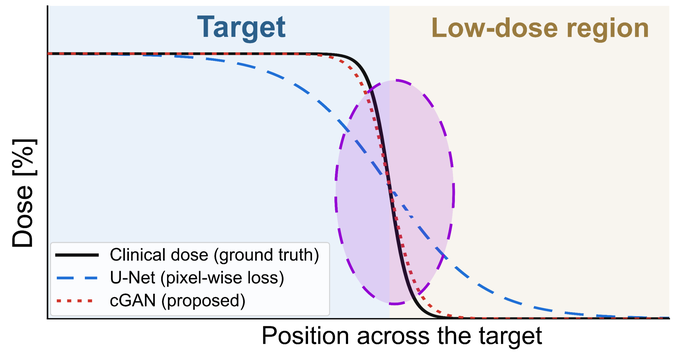

Figure 1. Schematic dose profile across the target. Pixel-wise losses smooth out the characteristic sharp fall-off, averaging away high-frequency components (U-Net, blue). Adversarial training preserves this steep gradient (cGAN, red).

METHODS AND MATERIALS

Both a Pix2Pix-based cGAN and a baseline 3D U-Net without adversarial training were trained under identical conditions (Figure 2). The cGAN generator was a 3D U-Net with attention gates; a 3D PatchGAN discriminator preserved high-frequency detail through adversarial training. Both used composite reconstruction losses (global and per-ROI L1, a dose-gradient loss, and medium-/high-dose L1 losses); the cGAN added an adversarial loss.

Both models predicted prescription-normalized dose from the 7-channel input. Using 172 HCC patients, accuracy was evaluated as the MAE of dose-volume metrics for the CTV, PTV, and normal liver (Liver − CTV), plus the 10% isodose volume; models were compared by paired t-tests.

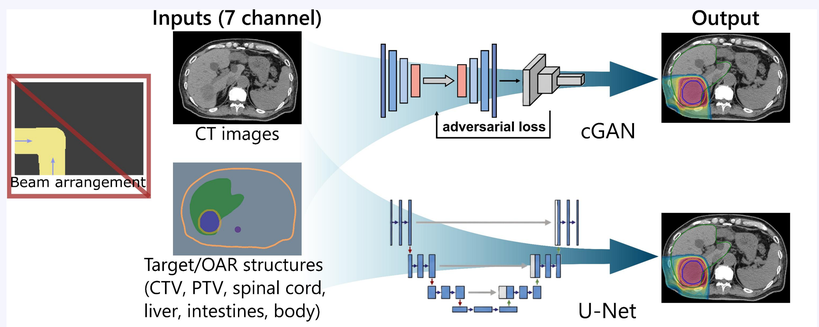

Figure 2. Overview of the proposed framework. The inputs are only CT images and target/OAR contours. The cGAN and U-Net each predict a dose distribution, which is compared with the clinical dose.

RESULTS

For target-volume dose prediction, the cGAN significantly outperformed the U-Net in CTV $D_{95\%}$. The MAE was 0.48 vs. 1.08 Gy(RBE) (p = 0.034; Table 1). The same trend was seen for CTV $D_{98\%}$ and $D_{2\%}$, though it was not significant. No significant differences were observed for the PTV or normal liver.

For the 10% isodose volume, the U-Net gave a lower MAE than the cGAN. The values were 51.0 vs. 61.3 cm³ (p = 0.48). This reflects numerically higher accuracy in the low-dose region.

In the representative case (Figure 3), the cGAN reproduced the dose within the CTV more accurately. This is consistent with its better $D_{95\%}$. At the target margin, however, the U-Net was closer to the clinical dose. The difference maps showed larger cGAN errors in the peripheral, low-dose region.

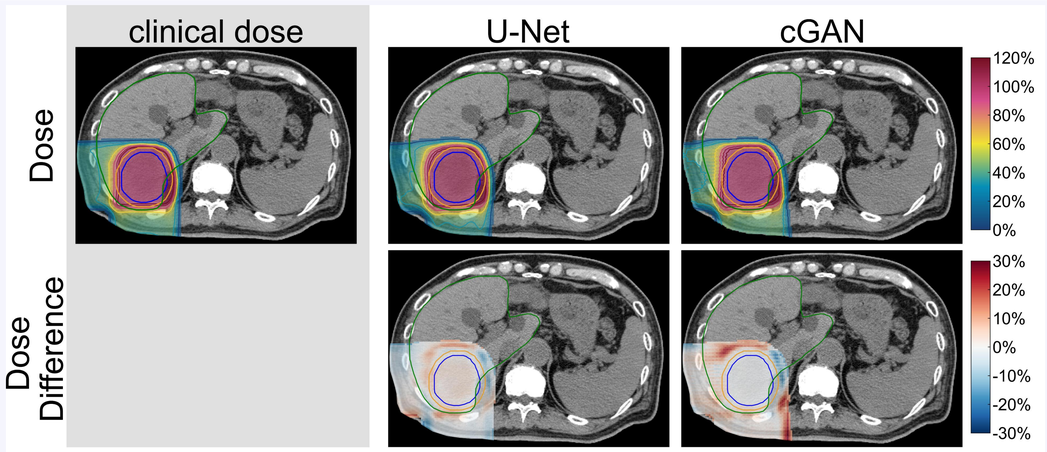

Figure 3. Dose distributions and their differences from the clinical dose for a representative case. Left: clinical dose; middle: U-Net; right: cGAN.

Table 1. Mean absolute error (MAE ± SD) of dose-volume metrics for the CTV, PTV, and normal liver. p-values are from paired t-tests.

Structure	Dose Metric	cGAN (MAE ± SD)	U-Net (MAE ± SD)	p-value
CTV	$D_{98\%}$ [Gy(RBE)]	1.00 ± 1.17	1.34 ± 0.89	0.363
	$D_{95\%}$ [Gy(RBE)]	0.48 ± 0.46	1.08 ± 0.82	0.034
	$D_{2\%}$ [Gy(RBE)]	0.74 ± 0.59	0.87 ± 0.63	0.594
PTV	$D_{98\%}$ [Gy(RBE)]	3.49 ± 4.68	3.00 ± 3.98	0.289
	$D_{95\%}$ [Gy(RBE)]	2.09 ± 2.04	1.96 ± 1.40	0.734
	$D_{2\%}$ [Gy(RBE)]	0.73 ± 0.55	0.75 ± 0.52	0.933
Normal liver (Liver − CTV)	D_{mean} [Gy(RBE)]	0.54 ± 0.58	0.57 ± 0.42	0.797
	$V_{30\ Gy(RBE)}$ [%]	1.17 ± 1.38	0.88 ± 1.31	0.234

DISCUSSION

The cGAN's superior target-volume performance likely arises because adversarial training with the PatchGAN discriminator dynamically penalized high-frequency mismatches, enabling reproduction of the steep dose gradients and overcoming the over-smoothing typical of pixel-wise-loss CNNs. In the low-dose region, however, the U-Net performed better: by prioritizing high-frequency components, the cGAN lost accuracy where the dose varies gradually. These results suggest a trade-off between steep dose-gradient and gradually changing low-dose regions.

Because it predicts dose from CT and structures alone, the framework could provide an early reference dose that standardizes plan quality, reduces inter-planner variability, and shortens planning time. It could also enable virtual PBT dose evaluation at facilities without PBT, supporting treatment-selection decisions.

CONCLUSIONS

The cGAN framework captures the steep dose gradients characteristic of PBT and significantly improves target-volume dose-prediction accuracy. Predicting dose from CT images and structures alone, it holds promise for standardizing treatment planning and for virtual dose evaluation at facilities without PBT. Future work will improve low-dose-region accuracy for balanced predictions across the entire dose range.