Development and multi-institutional external validation of an automatic IMRT planning system for highly accurate beam delivery using deep learning-predicted fluence maps

T Matsuura [1], D Kawahara [1], S Utsunomiya [2], M Sakai [2], H Nakano [3], T Yamada [4], Y Murakami [1]

1. Hiroshima University, Hiroshima, Japan 2. Niigata University, Niigata, Japan 3. Niigata University of Health and Welfare, Niigata, Japan 4. Niigata University Medical and Dental Hospital, Niigata, Japan

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

INTRODUCTION

Intensity-modulated radiation therapy (IMRT) enables highly conformal dose delivery but requires patient-specific quality assurance (QA) because of complex multileaf collimator motion. Although measurement-based QA is essential for ensuring accurate beam delivery, it remains labor-intensive and time-consuming.

AI-based QA prediction can estimate delivery accuracy before measurement and may improve QA workflow efficiency[1]. However, QA prediction alone does not directly improve treatment plans when low delivery accuracy is predicted, and plan modification with QA re-evaluation is still required.

Improving delivery accuracy at the planning stage may represent a more fundamental strategy for reducing QA-related uncertainty, particularly in time-critical workflows such as online adaptive radiotherapy, where conventional measurement-based QA is difficult to perform.

This study aimed to develop a deep learning-based fluence synthesis model as a component of accuracy-focused automatic IMRT planning, designed to improve beam delivery accuracy using predicted fluence maps. Its dosimetric performance and generalizability were evaluated through internal validation and multi-institutional external validation without model retraining.

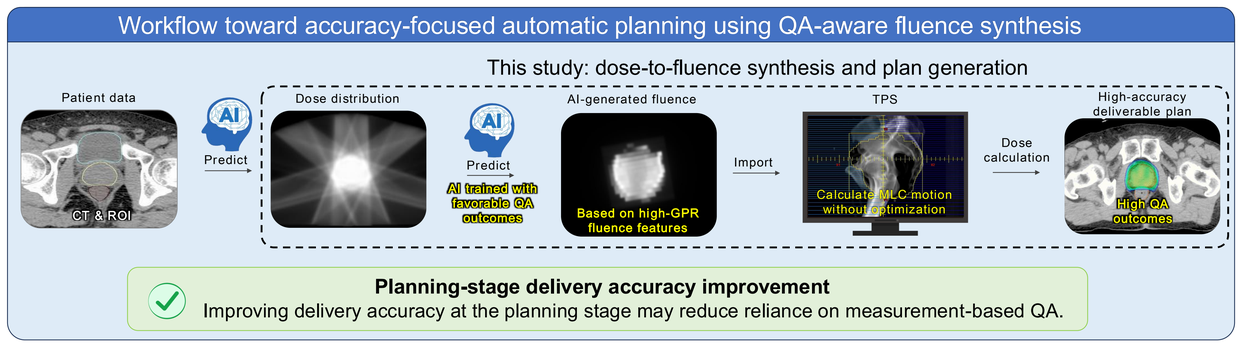

Figure 1. Conceptual workflow toward accuracy-focused automatic planning and the main scope of this study.

MATERIAL & METHODS

Figure 2 summarizes the workflow from model development to synthetic plan generation and validation. A total of 101 prostate IMRT plans were retrospectively collected for model development. All plans were generated in Eclipse for a TrueBeam linac using 10-MV X-rays with seven fixed fields and dynamic MLC delivery. EPID-based measurements and gamma analyses were performed for all beams. Beams in the upper 50% of gamma passing rates (GPRs) at each beam angle were selected as modeling data to represent fluence characteristics associated with favorable QA outcomes.

In this study, field dose distributions derived from CT and contour information were assumed to be predictable at clinically acceptable levels based on previous dose prediction studies[2,3]. Thus, the corresponding portal dose distributions derived from clinical plans were used as inputs to the fluence synthesis model.

A pix2pix generative adversarial network (GAN) with a U-Net generator was trained, after data augmentation, using paired datasets of portal dose distributions and actual optimal fluence maps. The developed fluence synthesis model automatically generated optimal fluence maps from portal dose distributions without TPS-based optimization. The AI-generated optimal fluence maps were imported into the treatment planning system, where only leaf sequencing and dose calculation were performed to generate deliverable synthetic plans (SPs). The network was implemented in TensorFlow 2.19.0 / Python 3.12 and trained for 2000 epochs.

Internal validation was performed using 20 independent test cases that were not included in model training. For external validation, the trained model was applied to 20 prostate IMRT cases from two external institutions without model retraining. SPs were delivered and re-measured using EPID. Clinical plans (CPs) and SPs were compared in terms of EPID-based GPRs and dose–volume metrics for the PTV, rectal wall, and bladder wall.

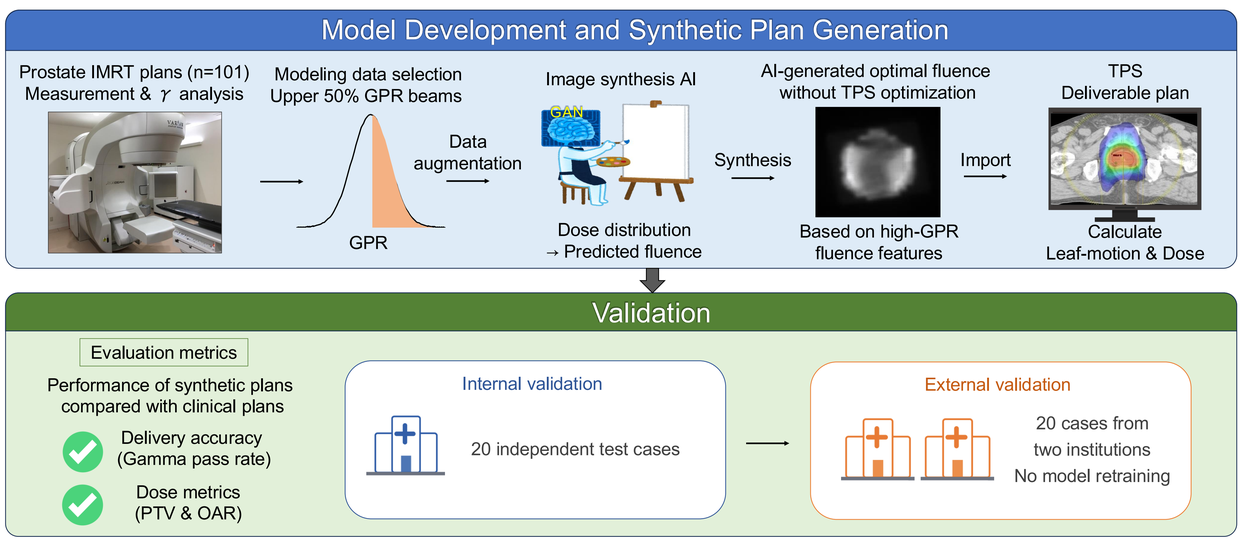

Figure 2. Workflow of model development and synthetic plan generation with internal and external validation.

RESULTS & DISCUSSION

Figure 3 shows box plot comparisons of GPRs between CPs and SPs. In both internal and external validation, SPs showed significantly higher GPRs than CPs at 3%/2 mm and 1%/1 mm. GPRs increased from 97.1% to 98.8% internally and from 97.0% to 98.5% externally at 3%/2 mm, with larger improvements at 1%/1 mm: 68.6% to 80.2% internally and 67.2% to 74.6% externally.

At the TG-218–recommended 3%/2-mm criterion, beams below the 95% GPR threshold decreased from 9.1% to 0% internally and from 7.1% to 1.7% externally. The statistically estimated probability of SP beams falling below this threshold was < 0.1% in internal validation and 0.7% in external validation.

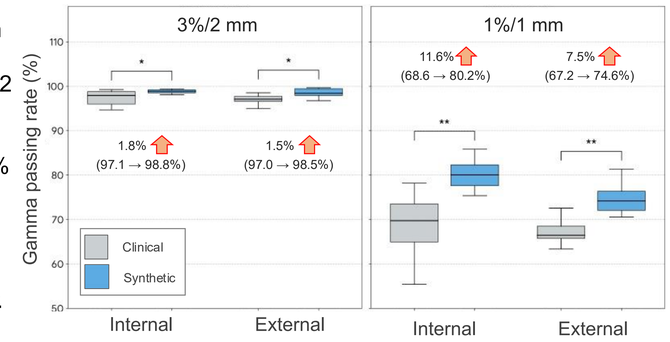

Figure 3. Box plots of the GPRs for CPs and SPs at 3%/2-mm and 1%/1-mm tolerances. *$p < 0.05$, **$p < 0.01$

The GPR improvement may be partly attributed to reduced plan complexity. Modulation complexity scores increased significantly from CPs to SPs in both internal validation (0.171 ± 0.014 to 0.252 ± 0.020) and external validation (0.220 ± 0.069 to 0.290 ± 0.073), indicating less complex modulation in SPs. Compared with CPs, MUs in SPs were reduced by 23.2% in internal validation and 19.9% in external validation, consistent with the reduced modulation complexity of SPs. Although previous work showed that the aperture shape controller function improved GPRs by reducing modulation complexity[4], the GPR improvement observed in this study was larger than that reported previously. These findings suggest that the GAN-based model learned fluence characteristics associated with high-GPR beams used as modeling data, in addition to reducing modulation complexity.

Figure 4 shows representative dose distributions and DVHs for CPs and SPs. The dose distributions of SPs were generally comparable to those of CPs in both internal and external validation. As shown in Table 1, SPs showed slight reductions in target coverage and increases in high-dose target metrics compared with CPs. Nevertheless, all internal cases and 90% of external cases satisfied the clinical dose constraints for the target and OARs. Although the external validation supports the potential generalizability of the proposed approach, further refinement is required to minimize the slight reduction in target coverage while preserving the improved delivery accuracy.

This study demonstrated that selecting training data based on favorable QA outcomes can improve delivery accuracy in fixed-beam prostate IMRT. Because portal dose distributions derived from clinical plans were used in this study, integration with clinically validated dose prediction workflows will be required for a fully automated planning system. Future work should include this integration, further evaluation of inter-institutional generalizability, and extension of this QA-aware fluence synthesis approach to VMAT delivery and non-prostate treatment sites.

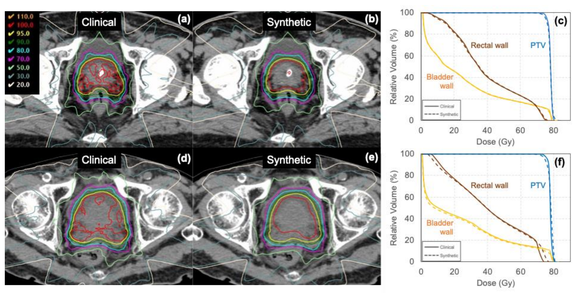

Figure 4. Representative dose distributions and DVHs for internal (a–c) and external (d–f) validation cases.

Table 1. Dosimetric comparison between CPs and SPs. Values are mean ± SD (median). *$p < 0.05$.

Structure	Parameters	Internal		External		Constraints
		Clinical	Synthetic	Clinical	Synthetic	
PTV	D2% (Gy)	79.9 ± 0.1 (79.9)	82.4 ± 0.7 * (82.6)	80.1 ± 0.4 (80.1)	81.8 ± 1.1 * (81.7)	≤ 85.8
	D95% (Gy)	76.1 ± 0.3 (76.2)	73.3 ± 0.4 * (73.3)	75.4 ± 1.0 (75.3)	71.8 ± 1.4 * (72.1)	≥ 70.2
Rectal wall	V70 Gy (%)	14.3 ± 4.3 (13.3)	15.7 ± 3.4 (15.2)	14.2 ± 2.6 (14.4)	14.7 ± 2.8 (13.5)	≤ 25%
	V40 Gy (%)	39.6 ± 7.6 (41.0)	41.7 ± 7.9 * (43.0)	49.8 ± 5.0 (48.9)	50.1 ± 6.0 (49.0)	≤ 65%
Bladder wall	V70 Gy (%)	20.1 ± 8.1 (17.1)	19.5 ± 7.7 * (17.2)	16.6 ± 5.7 (17.0)	15.5 ± 5.5 (15.7)	≤ 35%
	V40 Gy (%)	33.7 ± 11.1 (28.9)	34.9 ± 11.4 * (30.3)	33.4 ± 11.4 (35.0)	32.4 ± 10.9 (32.9)	≤ 65%

CONCLUSION

The proposed QA-aware fluence synthesis model, trained on data selected from high-GPR beams, improved beam delivery accuracy across internal and multi-institutional external datasets without model retraining. Most synthetic plans maintained clinically acceptable dosimetric performance, although slight reductions in target coverage were observed. These findings suggest that selecting training data based on favorable QA outcomes is an effective strategy for generating fluence maps associated with improved delivery accuracy. This approach may contribute to future accuracy-focused automatic IMRT planning workflows, particularly during stable linac operation periods and in time-critical workflows such as online ART.

REFERENCES

1) T Ono, H Iramina, H Hirashima, et al. Journal of Radiation Research. 65, 421-432 (2024)
2) H Lee, H Kim, J Kwak et al. Scientific Reports 9, 15671 (2019).
3) D Nguyen, L Troy, J Xun et al. arXiv: 1709.09233 (2017).
4) D Binny, M Spalding and S.B. Crowe et al. Medical Dosimetry 45, 284–292 (2020).

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant No. 26K18957.

CONTACT INFORMATION

Takaaki Matsuura, Ph.D.
Department of Radiation Oncology, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan
E-mail: matsuura-t@hiroshima-u.ac.jp