Evaluation of Patient-Specific Auto-segmentation via Intentional Overfitting: Assessing Lightweight Architectures for Accelerated Online Workflow in MRgART for Prostate Cancer

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

Toshitaka Takahashi[1)], Syoma Ide[1)], Kota Abe[2)], Masato Tsuneda[2)], Makoto Saito[3)], Yujiro Nakajima[4)], Naoki Tohyama[4)], Takashi Uno[5)], Yukio Fujita[3,4)]

1) Komazawa University, Tokyo, Japan, 2) Graduate School of Medicine, Chiba University, Chiba, Japan, 3) Department of Radiology, Chiba University Hospital, Chiba, Japan , 4) Department of Radiation Sciences, Komazawa University, Tokyo, Japan, 5) Diagnostic Radiology and Radiation Oncology, Graduate School of Medicine, Chiba University, Chiba, Japan

INTRODUCTION

- **Challenges in MRgART:**
  Online contouring is time-consuming due to daily anatomical variations.
- **Limitations of Generalized Models:**
  Conventional models struggle to adapt to patient-specific changes.
- **Computational Barrier:**
  Highly accurate complex models require impractical computational time for clinical use.

AIM

To evaluate whether patient-specific models (PSMs) generated by fine-tuning the base model on treatment-planning MRI to the point of near-overfitting with lightweight networks can reduce training time while maintaining high segmentation accuracy for online MRgART.

METHODS

Phase 1: Base Model (BM) Training

- To establish generalized organ segmentation capabilities, three types of networks were trained: a standard **3D-Unet**, **LW-Unet** (LightWeight-Unet : a lightweight 3D-Unet with separable convolutions,~42.6K parameters) and **MM-Unet** (MixedMamba-Unet: a 3D-Unet variant with Res2Net encoding, Mamba bottleneck, and adaptive skip fusion,~73M parameters).
- These models were initially trained on the LUND-PROBE dataset[(1)], comprising 425 prostate cancer T2-weighted MRI cases, to serve as Base Models (BMs). The training was continued until the models fully converged.

Phase 2: Patient-Specific Model (PSM) Generation via Intentional Overfitting

- We evaluated 41 clinical prostate cancer cases treated with a 1.5T MR-Linac (Unity, Elekta) at Chiba University. Each patient's dataset consisted of a planning MRI and 5 daily pre-treatment fraction MRIs.
- For each patient, an individualized PSM was generated by fine-tuning the BM using exclusively the planning MRI and corresponding ground-truth contours to the point of near-overfitting.
- This training process was deliberately extended beyond full model convergence (Intentional Overfitting) to effectively adapt to the patient's unique anatomical characteristics.

Phase 3: Evaluation of Clinical Feasibility

- To simulate an online adaptive workflow, the generated PSMs were applied to the daily fraction MR images (5 fractions per patient).
- Segmentation accuracy for the rectum, bladder, and CTV was quantitatively evaluated using the Dice similarity coefficient (DSC).
- Furthermore, the required training durations were compared among the 3D-Unet, LW-Unet, and MM-Unet to assess whether the lightweight models could sufficiently reduce training time to meet strict clinical time constraints.

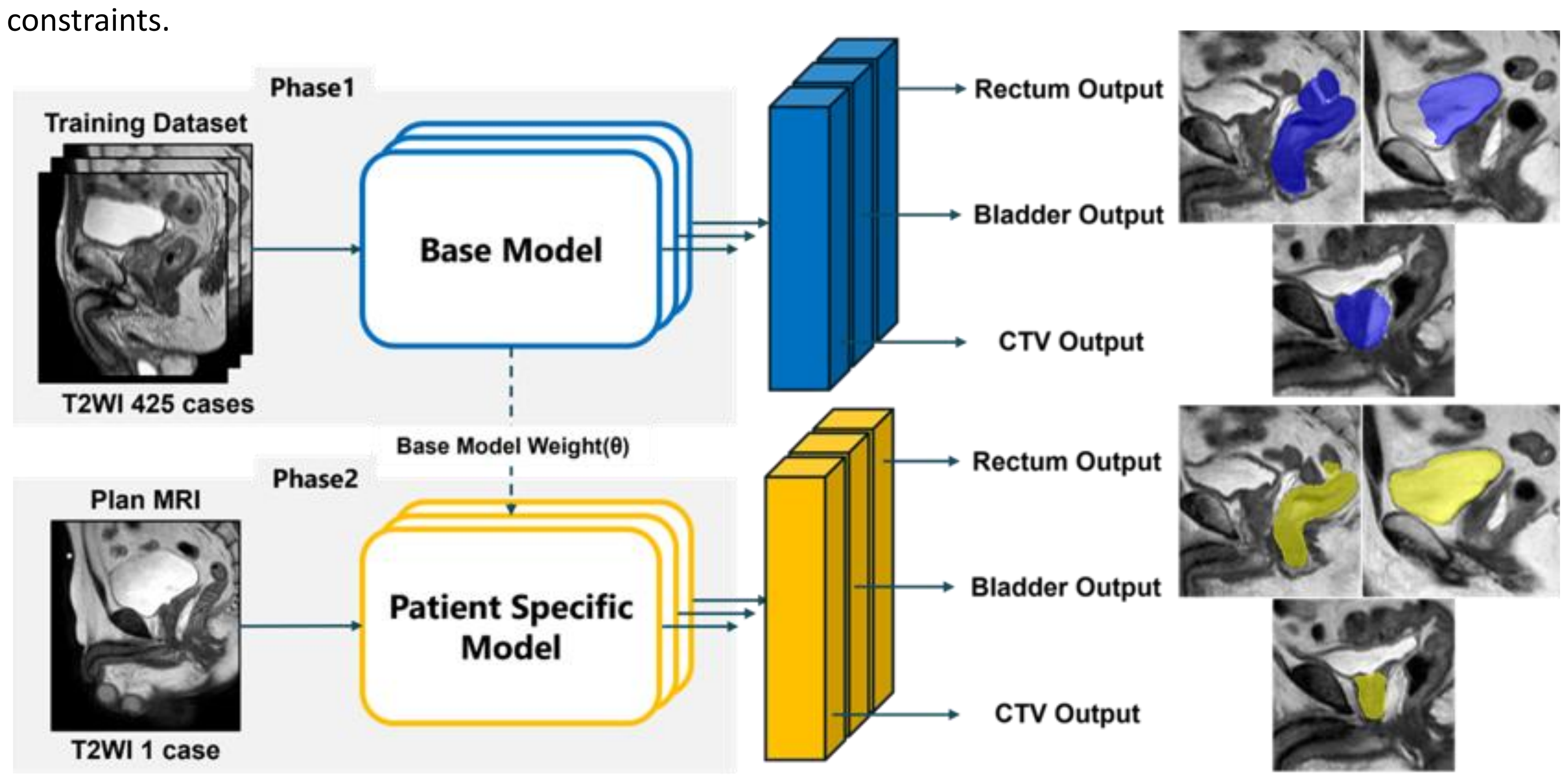

Figure 1: Workflow for Patient-Specific Model (PSM) Generation

REFERENCES

1) Rogowski V, et al. LUND-PROBE – LUND Prostate Radiotherapy Open Benchmarking and Evaluation dataset. Sci Data. 2025;12:611. doi:10.1038/s41597-025-04954-5.
2) Li Z, et al. Patient-specific daily updated deep learning auto-segmentation for MRI-guided adaptive radiotherapy. Radiother Oncol. 2022;177:222-230.
3) Yang B, et al. Pretreatment information–aided automatic segmentation for online magnetic resonance imaging-guided prostate radiotherapy. Med Phys. 2024;51:922-932.
4) Chen X, et al. Personalized auto-segmentation for magnetic resonance imaging-guided adaptive radiotherapy of prostate cancer. Med Phys. 2022;49(8):4971-4979.

RESULTS

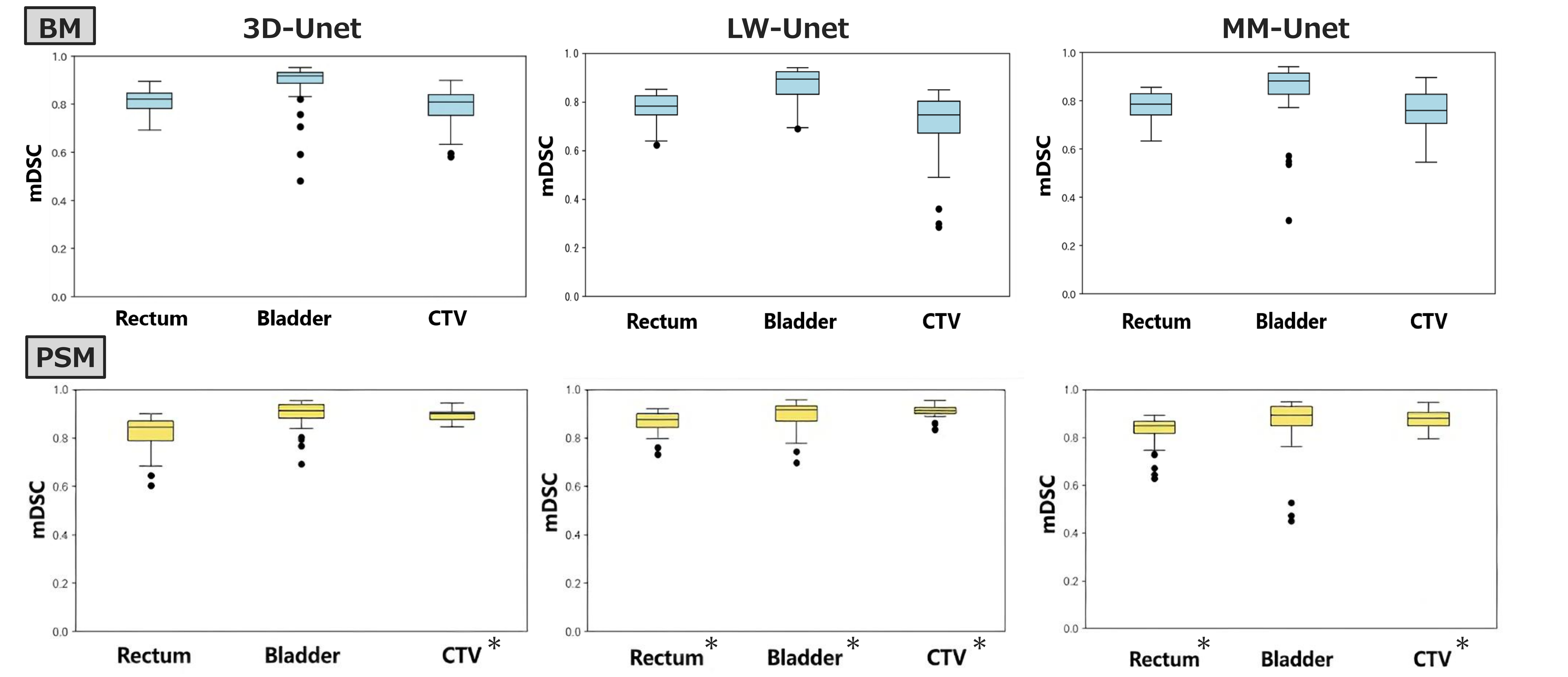

Fig. 2. Comparison of Mean Dice Similarity Coefficient (mDSC) across Three Network Architectures (* $p < 0.05$ vs. BM)

- Across all architectures, **patient-specific models (PSMs) improved or maintained the mean DSC for each organ compared to the base models (BMs)**. The mean DSCs (BM / PSM) were 0.79/0.82 (Rectum), 0.86/0.90 (Bladder), and 0.76/0.90 (CTV) for 3D-Unet; 0.78/0.87, 0.86/0.89, and 0.67/0.91 for LW-Unet; and 0.78/0.83, 0.84/0.86, and 0.76/0.88 for MM-Unet.

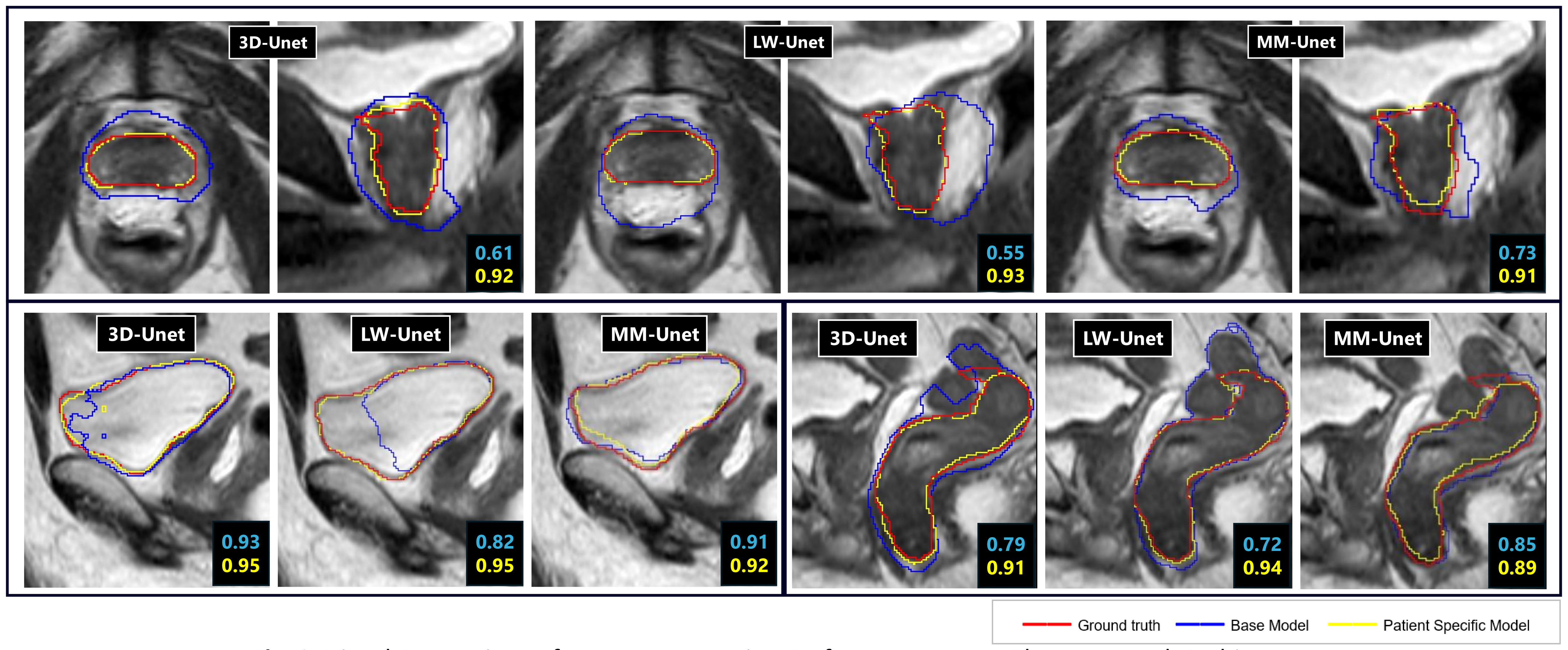

Fig. 3. Visual Comparison of Auto-segmentation Performance across Three Network Architectures

- Visual assessments confirmed that **PSMs more closely matched the ground truth masks than BMs, successfully adapting to the patients' daily anatomical variations**. The DSC for each prediction is displayed in the lower right corner of the image (blue: BM, yellow: PSM).

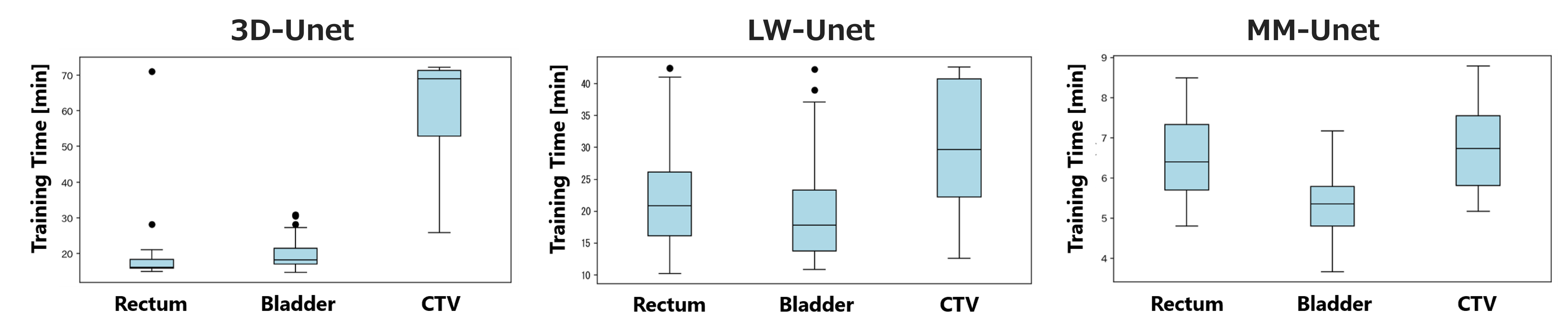

Fig. 4. Comparison of Training Time across Three Network Architectures

- The average training time required for intentional overfitting of the CTV was substantially reduced from **61 min with 3D-Unet** to approximately **31 min with LW-Unet** and **7 min with MM-Unet**. MM-Unet also achieved the shortest training times for both the rectum and bladder.

CONCLUSIONS

- Combining intentional overfitting of pre-treatment planning MRI with lightweight architectures maintained high segmentation accuracy while substantially reducing the patient-specific **model training time to less than 7 minutes when all organs are trained in parallel with MM-Unet**.
- This optimal balance of computational efficiency and precision suggests that real-time, personalized daily auto-segmentation is clinically feasible for online MRgART workflows.

CONTACT INFORMATION

Name：Toshitaka Takahashi
E-mail：3615108t@komazawa-u.ac.jp
Address： Komazawa University Graduate School, Tokyo, Japan