Deep Learning-Based QA Plan-to-Portal Image Prediction for Proactive IMRT Quality Assurance

Xiuxiu He, Jie Yang, Licheng Kuo, Åse M. Ballangrud, Laura I. Cervino, Jean M. Moran, Pengpeng Zhang, Xiang Li, Tianfang Li

Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY

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

INTRODUCTION

Current IMRT pre-treatment QA is reactive: measurement-based EPID QA is scheduled only after treatment approval and depends heavily on machine availability. A failed measurement triggers the physics investigation, plan revision, MD re-approval, and repeat measurement — delaying treatment by days. We developed a deep learning model using data from 20+ LINACs that predicts the measured portal image directly from the QA plan during the planning stage. Because the predicted image reproduces the measurement, predicted-image QA can flag potential failures early and transform EPID QA from reactive to proactive.

METHODS AND MATERIALS

- **Model:** Residual U-Net (768×768 input) that predicts the residual added to the reference image (RI + network = PI). Attention gates and squeeze-and-excitation channel attention emphasize field-specific features; group normalization stabilizes small-batch training.
- **Data:** 1118 IMRT fields from 223 patients; QA-plan reference image (RI) as input, measured portal image (PI) as ground truth. Split 894 training / 224 validation, no augmentation.
- **Test set:** Independent 26-patient cohort (146 fields).
- **Training loss:** Combined L1 + SSIM with a top-K worst-pixel penalty, augmented by two edge-aware terms — gradient (edge-slope) matching and a penumbra-weighted L1 — added to sharpen field edges, where prediction error was largest.

$$\mathcal{L} = \alpha\mathcal{L}_1 + (1-\alpha)\mathcal{L}_{SSIM} + \lambda_{max}\mathcal{L}_{max} + \lambda_{grad}\mathcal{L}_{grad} + \lambda_{edge}\mathcal{L}_{edge}$$

- **Registration:** Rigid alignment with translation and rotation matching predictions to measurements.
- **Evaluation:** Prediction accuracy from Pred-vs-PI MAE (jaw-masked) and Pred-vs-PI gamma (3%/2mm, 10% threshold). QA equivalence assessed by comparing predicted-portal gamma (Pred vs RI) against that of measured (PI vs RI).

DISCUSSION

The predicted portal closely matches the measurement: mean MAE 1.09% of peak signal and Pred-vs-PI gamma 98.8% (3%/2mm) across 146 fields.

Because the prediction is accurate, predicted-portal gamma (Pred vs RI) reproduces conventional measured QA (PI vs RI): Pearson r = 0.932, AUC 0.983, and 95.2% verdict agreement at the 95% criterion, with 86.7% of failing fields flagged. Prediction takes ~0.76 s per field on an RTX 6000 GPU.

CONTACT

Tianfang Li, PhD and Xiuxiu He, PhD
Department of Medical Physics
Memorial Sloan Kettering Cancer Center, New York, NY

RESULTS

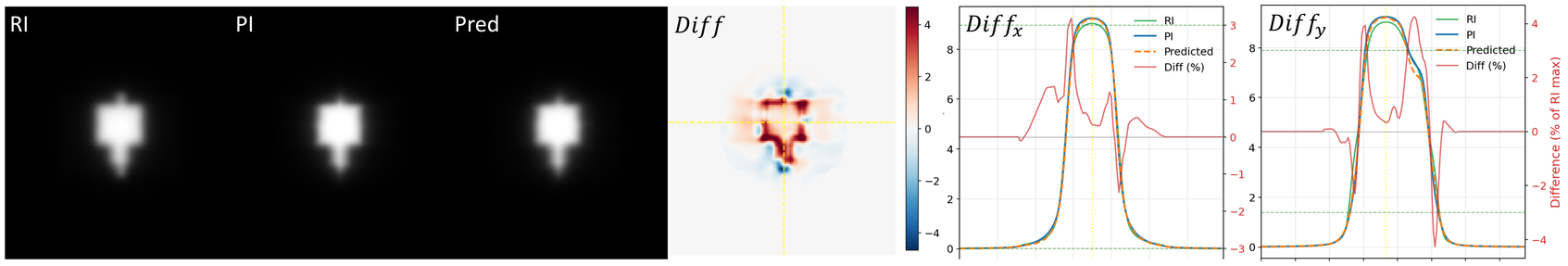

Figure 1. A representative field of a single lesion brain SRS plan. RI: Reference image from an EPID QA plan. PI: Ground truth portal image from measurement. Pred: Model predicted portal image. Diff: Difference image between the predicted and ground truth portal image. $Diff_x$ and $Diff_y$: line profiles.

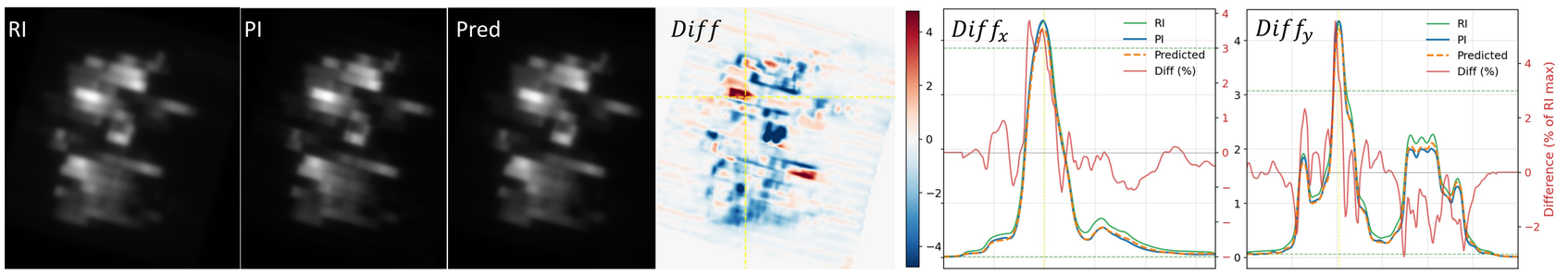

Figure 2. A representative field of a 19-lesion hypo-brain plan.

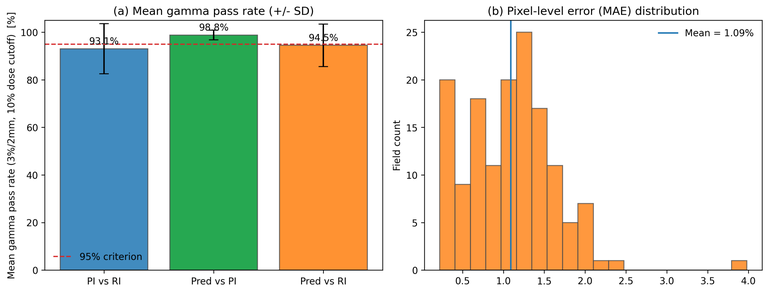

Figure 3. Model accuracy summary: mean gamma pass rate (±SD) and per-field MAE distribution (146 fields).

Predicted portal matches the measurement (Pred vs PI: 98.8% gamma, 1.09% MAE); predicted QA (Pred vs RI, 94.5%) reproduces measured QA (PI vs RI, 93.1%).

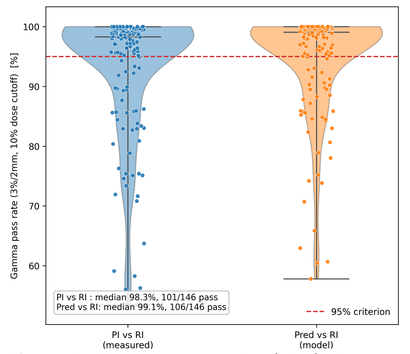

Figure 4. Gamma pass-rate distribution: measured (PI vs RI) vs model (Pred vs RI).

Predicted QA (Pred vs RI) reproduces measured QA (PI vs RI): medians 98.3% vs 99.1%, 101 vs 106 fields ≥95%.

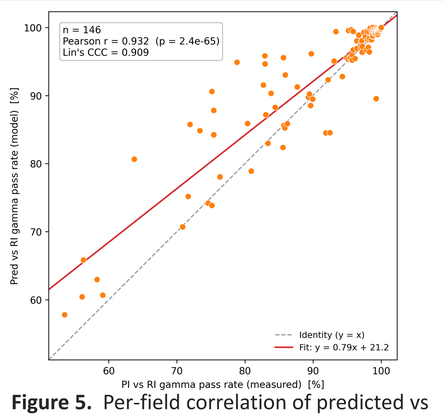

Figure 5. Per-field correlation of predicted vs measured gamma pass rate.

Predicted and measured QA agree per field: Pearson r = 0.932, Lin's CCC = 0.909.

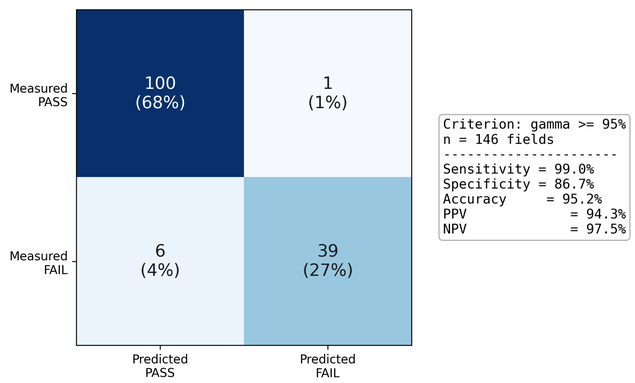

Figure 6. QA verdict agreement at the 95% gamma criterion (n = 146).

Predicted and measured QA verdicts agree for 95.2% of fields; 86.7% of true fails flagged.

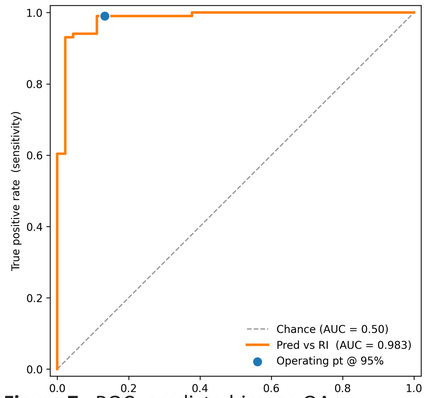

Figure 7. ROC: predicted-image QA vs measurement-based QA verdict.

AUC = 0.983 — predicted QA separates passing from failing fields.

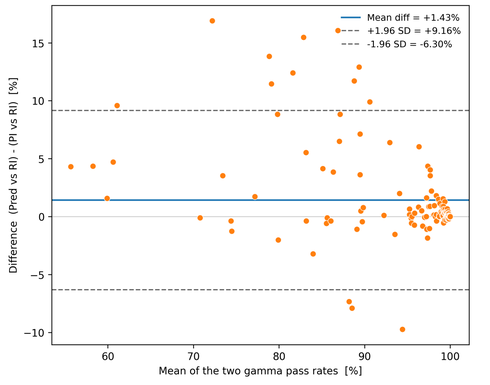

Figure 8. Bland–Altman agreement of predicted and measured QA gamma pass rates.

Predicted minus measured QA: +1.43% bias; 95% limits −6.30% to +9.16%.

CONCLUSIONS

A Residual Attention U-Net predicts the measured portal image from the QA plan with high accuracy (Pred vs PI: 98.8% gamma, 1.09% MAE) before any measurement.

Because the predicted portal reproduces the measurement, predicted-image gamma QA (Pred vs RI) agrees with conventional measured QA (PI vs RI) — AUC 0.983, 95.2% accuracy — enabling proactive plan optimization and fewer QA-driven delays.