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

Simultaneous Radiobiological Response-Guided Automatic Treatment Planning and Physician Preference Quantification In High-Dose-Rate Brachytherapy of Cervical Cancer via Inverse Deep Reinforcement Learning

Yin Gao, PhD[1]; Ethan Wang, MD[3]; Kevin Albuquerque, MD[3]; Xun Jia, PhD[2]

[1]Department of Radiation Oncology, University of Pennsylvania, [2]Department of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins University, [3]Department of Radiation Oncology, UT Southwestern Medical Center

Abstract

Purpose: Automatic treatment planning and physician's plan preference quantification are both critical tasks in High-dose-rate brachytherapy (HDRBT) of cervical cancer for rapidly generating high-quality plans meeting physician's intent. This study employs inverse deep reinforcement learning (iDRL) to simultaneously address two tasks by learning interpretable physician preferences based on radiobiological responses and using the learned preference model to guide automatic treatment planning.

Methods: The iDRL framework integrated DRL-based auto-planning with active reward learning. A planning agent trained using a DQN-based scheme iteratively adjusted organ weighting factors in an optimization engine based on observed plan to improve plan quality. Tumor control probability (TCP) and normal tissue complication probabilities (NTCPs) for bladder, rectum, sigmoid, and small bowel were computed using established radiobiological models. A logistic regression model was used to distinguish clinically approved and agent-generated plans in TCP/NTCP domain, generating plan approval probability that served as the reward function for the DRL agent. The reward model and planning agent were jointly trained. The framework was trained on 60 patients and evaluated on an independent cohort of 15 patients treated by the same physician.

Results: The iDRL framework was successfully trained. The weighting factors in the logistic regression model for TCP and NTCPs of bladder, rectum, sigmoid, and small bowel were 0.41, −0.11, −0.27, −0.30, −0.12. Odds-ratio analysis identified TCP as the dominant driver of plan acceptance and greatest sensitivity to sigmoid and rectum toxicity. On the independent test cohort, plans generated by the trained agent achieved radiobiological quality comparable to clinical plans, with TCP 0.92±0.02 (clinical: 0.93±0.01), $NTCP_{bladder}$ 0.05±0.01 (0.05±0.01), $NTCP_{rectum}$ 0.09±0.01 (0.11±0.02), $NTCP_{sigmoid}$ 0.10±0.03 (0.11±0.03), and $NTCP_{bowel}$ 0.06±0.03 (0.07±0.02).

Conclusion: The iDRL framework for HDRBT treatment planning in cervical cancer effectively bridges physician preference modeling and automatic plan generation, enabling these two tasks to be addressed simultaneously within a clinically meaningful framework.

Contact

Yin Gao, PhD
Assistant Professor
Department of Radiation Oncology,
University of Pennsylvania
Email: yin.gao@pennmedicine.upenn.edu

INTRODUCTION

Clinical challenge

- Cervical HDR brachytherapy planning begins after applicator placement, while the patient remains on the treatment table.
- Planners and physicians must rapidly balance target coverage, organs at risk (OARs) sparing, and cumulative dose from external beam radiation therapy (EBRT) and prior brachytherapy fractions.
- Competing dose objectives often require repeated plan adjustment and review.

Why it matters

- Planning is time-sensitive and strongly dependent on clinical experience.
- Limited time may prevent full exploration of the best tumor control–toxicity trade-off (TCP/NTCP).
- Prolonged planning increases discomfort and the risk of patient or applicator movement.

Current gap

- Most auto-planning methods rely on predefined dosimetric objectives.
- These objectives do not directly capture physician plan-approval preferences and biological response.
- Preference modeling and automated planning remain largely separate.

Study objective

- Develop an interpretable Inverse Deep Reinforcement Learning (iDRL) framework that jointly learns physician preference from TCP/NTCP and uses it to guide automated HDR brachytherapy planning.

METHODS AND MATERIALS

Clinical dataset

- 20 patients; 75 fraction-level cervical HDRBT plans
- Tandem-and-ovoid applicators; plans approved by one physician
- 60 plans for training; 15 held-out plans for testing
- Clinical plans served as the physician-accepted reference

The iDRL framework (Figure 1)

- Planning agent: Iteratively generates HDRBT plans from reward feedback
- Preference model: Learns interpretable physician preferences in TCP/NTCP space

DRL Planning Agent

- Learns how to improve a plan through repeated interaction with the inverse optimizer.
- State: DVHs of the CTV, bladder, rectum, sigmoid, and small bowel
- Action: increase, maintain, or decrease one of five structure-specific optimization weights
- Five parameter-specific deep Q-networks estimate the expected benefit of each action
- The highest-valued action is applied to generate the next plan: $a_t = \arg\max_a Q_\theta(s_t, a)$

Radiobiological Preference Model

Learns what the physician considers an acceptable plan.

- Each plan is represented by its predicted tumor control and toxicity:
- Logistic regression converts these endpoints into a predicted physician-approval probability:

$$\mathbf{x}_t = [TCP, NTCP_{bladder}, NTCP_{rectum}, NTCP_{sigmoid}, NTCP_{bowel}]$$

- The planning agent is rewarded when an action increases the predicted approval probability:

$$p_t = P(\text{approve} \mid \mathbf{x}_t) = \frac{1}{1+\exp[-(\beta_0 + \boldsymbol{\beta}^T \mathbf{x}_t)]}$$

$$r_t = p_{t+1} - p_t$$

- The preference model is iteratively updated using balanced sets of clinically approved and agent-generated plans.

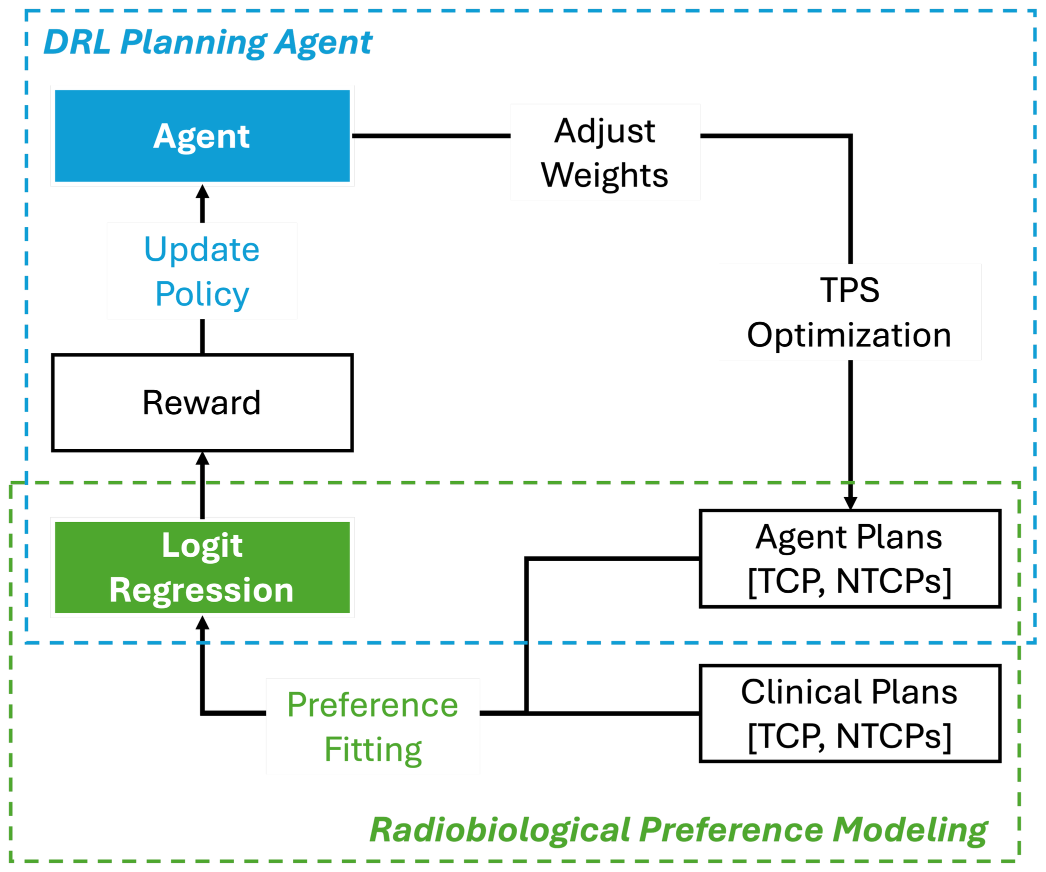

Figure 1. Overview of the iDRL framework, consisting of a DRL-based planning agent and an interpretable radiobiological preference model implemented via logistic regression.

RESULTS

The model learned an interpretable physician preference (Table 1)

- TCP was the dominant positive driver of plan approval.
- Increasing NTCP reduced approval probability for all OARs.
- Sigmoid and rectum toxicity had the strongest negative influence

Table 1. Learned decision model coefficients β and odds ratios (OR) for radiobiological endpoints used to model physician plan acceptance. OR=exp(β) indicates whether increasing that endpoint increases (>1) or decreases (<1) the odds of approval.

	TCP	$NTCP_{bladder}$	$NTCP_{rectum}$	$NTCP_{sigmoid}$	$NTCP_{bowel}$
β	0.4	−0.1	−0.3	−0.3	−0.1
OR	1.5	0.9	0.8	0.7	0.9

iDRL guided plans toward physician-preferred radiobiological trade-offs

- Initial agent plans formed a distinct cluster from clinically approved plans, shown in Figure 2.
- After optimization, final plans shifted toward the clinical-plan cluster and the learned preference boundary.
- The centroid movement of final plans demonstrated convergence toward physician-accepted TCP–NTCP trade-offs.

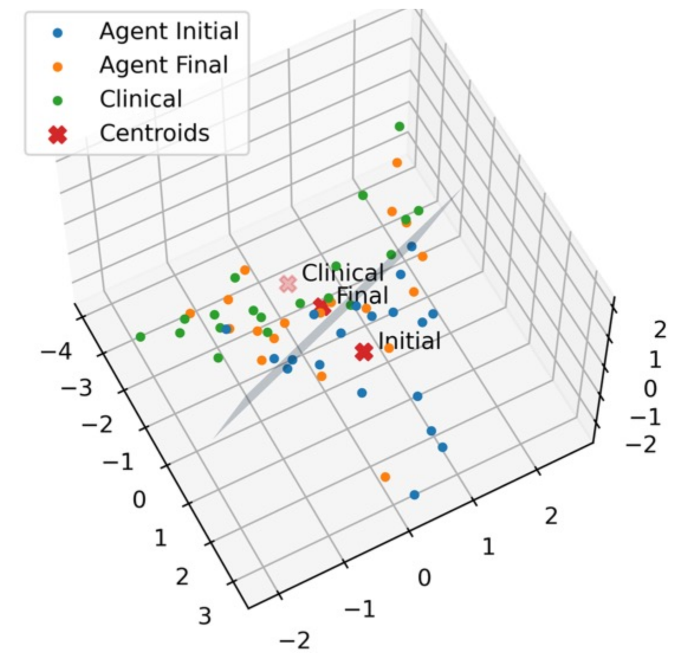

Figure 2. Clinical, agent-initial, and agent-final plans in the PCA representation of TCP/NTCP space.

Automated plans achieved clinical-comparable quality within minutes

- iDRL plans achieved TCP near the clinical plans, with comparable OAR NTCP.
- Plans were generated on average within 5 minutes, substantially reducing planning time compared with the conventional workflow.

Table 2. Radiobiological comparison of iDRL-generated and clinical plans in 15 test cases. Values are mea ±SD.

	iDRL plans	Clinical plans
TCP	0.92 ± 0.02	0.93 ± 0.01
NTCPbladder	0.05 ± 0.01	0.05 ± 0.01
NTCPrectum	0.09 ± 0.01	0.11 ± 0.02
NTCPsigmoid	0.10 ± 0.03	0.11 ± 0.03
NTCPbowel	0.06 ± 0.03	0.07 ± 0.02

- In an example case (Figure 3), iDRL improved TCP from 0.89 to 0.92 while preserving OAR risks comparable to the physician-approved plan.

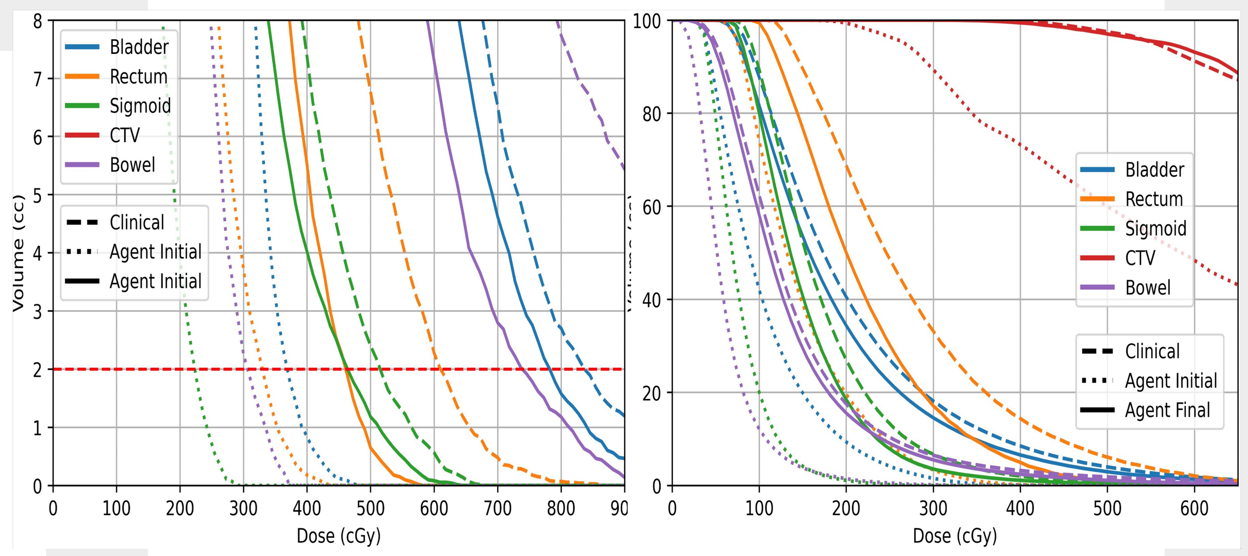

Figure 3. Absolute-volume (left) and percent-volume (right) DVHs for clinical (dashed), agent initial-step (dotted), and agent final-step (solid) plans, with the D2cc threshold shown as a red line.

CONCLUSIONS

This study introduces a unified iDRL framework that learns interpretable physician preferences and directly uses them to guide automated cervical HDRBT planning. The method generated clinically comparable plans within approximately 5 minutes, supporting faster, more consistent, and physician-aligned treatment planning.