Optimizing the Combination of Radiotherapy with Immunotherapy *via* Deep Reinforcement Learning

MaryLena Bleile[1, 2], Yixun Xing[1,3], Dan Nguyen[1, 6], Hein Van Nguyen[4], Casey Moore[5], Debabrata Saha[6], Benjamin Chen [6], Michael Story [6], Daniel F. Heitjan[2,7], Robert Timmerman[6], Steve Jiang[1, 6]

(1) Medical Artificial Intelligence and Automation Laboratory, UT Southwestern Medical Center
(2) Department of Statistical Sciences, Southern Methodist University
(3) Department of Advanced Data Analytics, University of North Texas
(4) Department of Electrical and Computer Engineering, University of Houston
(5) Department of Pathology, UT Southwestern Medical Center
(6) Department of Radiation Oncology, UT Southwestern Medical Center
(7) Department of Population and Data Science, UT Southwestern Medical Center

BACKGROUND

- Aim: optimize the combination of radiotherapy (RT) with immunotherapy (IO)
- Amount benefit IO adds depends on RT dose spacing

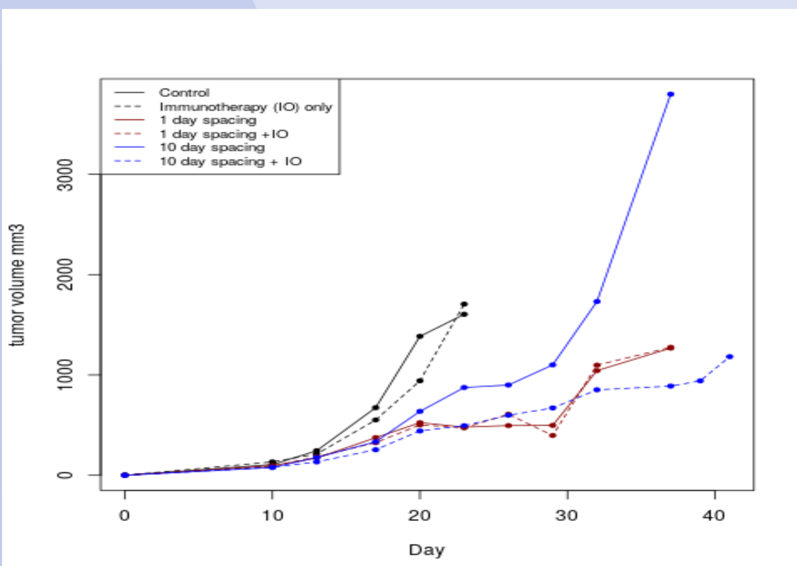

Fig.1: Immunotherapy has a greater effect on mice with normal immune systems when two fractions of radiotherapy are applied 10 days apart (blue) than when they are applied on consecutive days (red).

Goal: Given a fixed number of radiation pulses, find the optimal **adaptive** spacing for administration **for the individual subject**.

- Best RT spacing depends on the **individual's immune system** and may **change with time** for a single individual
- **Reinforcement Learning** is an approach to this problem which requires formulation of the problem as a **Markov Decision Process** (MDP): collection of potential states, actions, and a reward function.

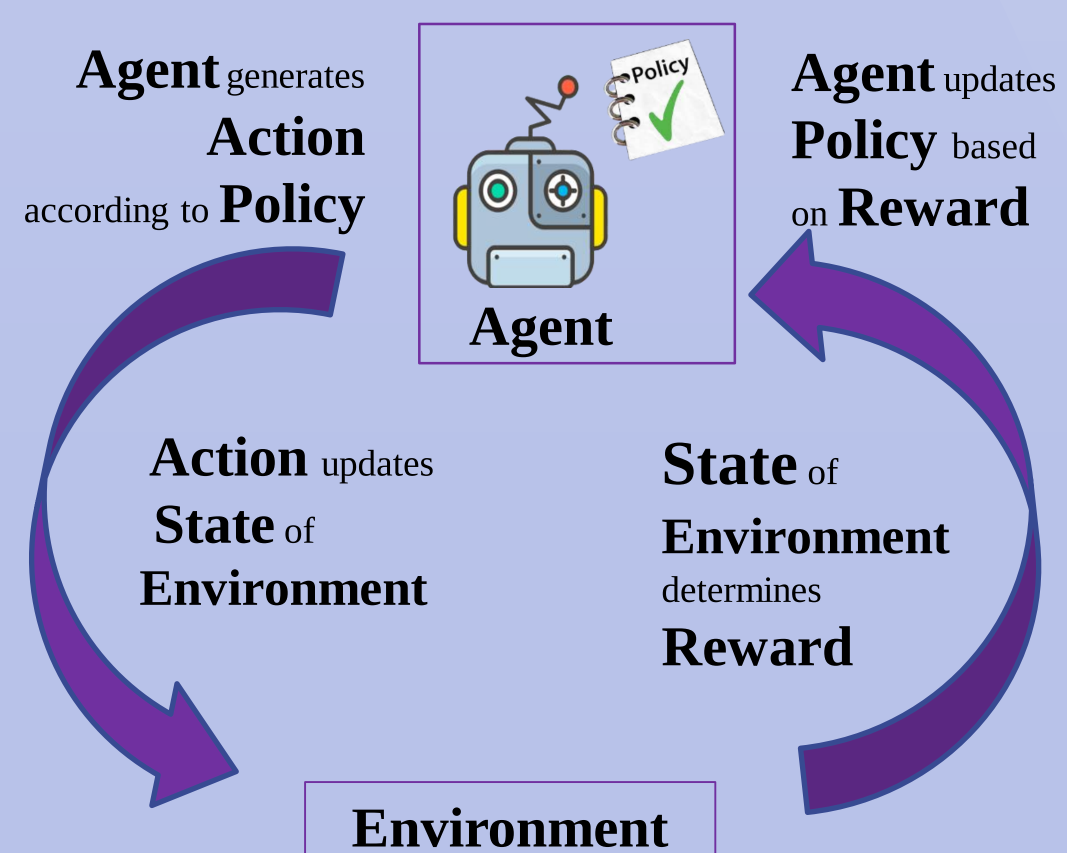

Fig.2: Illustration of the Reinforcement Learning paradigm

METHOD

- Train in simulated environment based on biologically-motivated dynamical systems model which can reproduce the synergistic effects between RT and IO

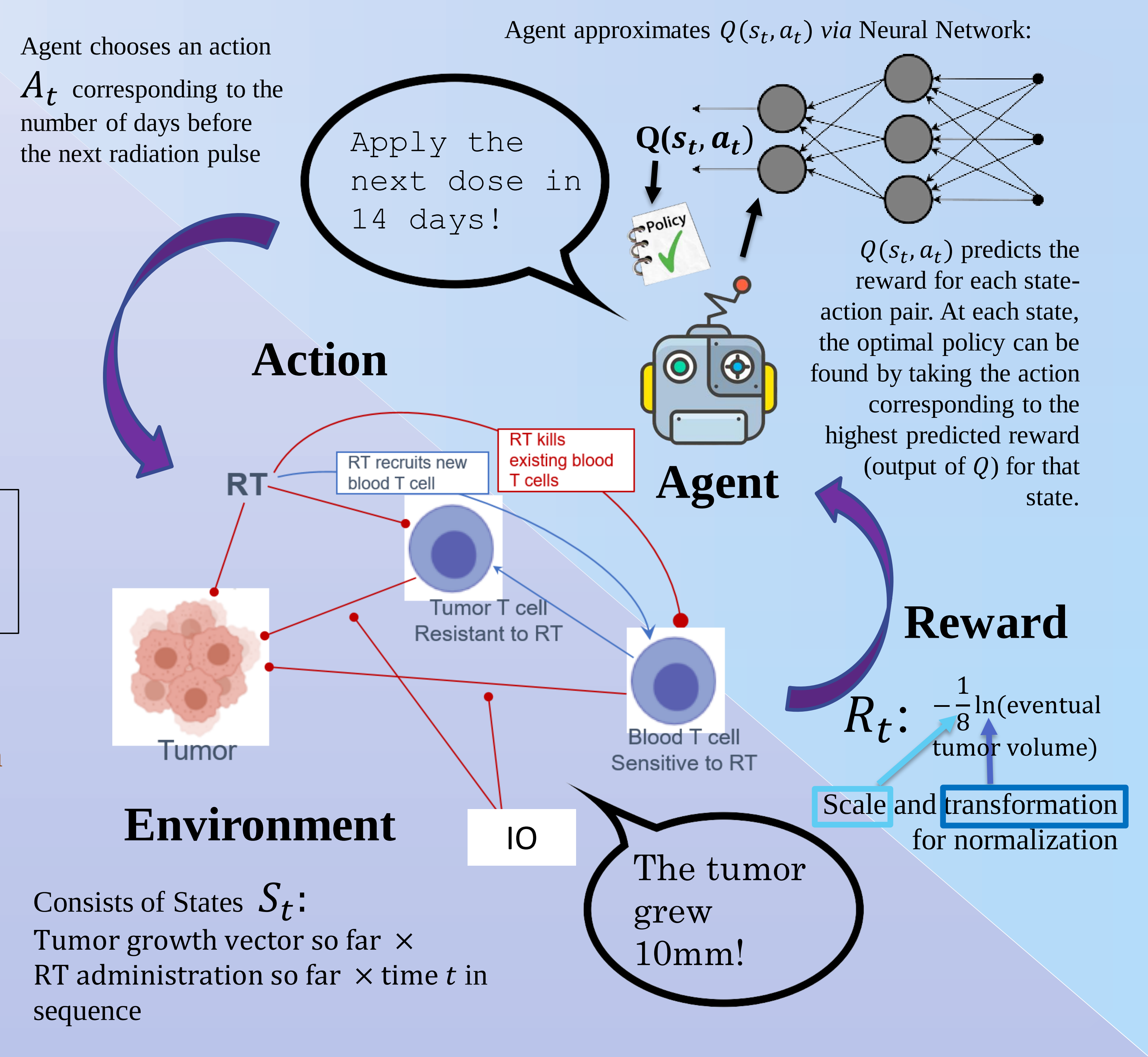

Fig.3: Reinforcement Learning framework for radiation policy determination

Novelty:

- **Mathematical formulation** of the optimal spacing problem as a tractable Markov Decision Process
- **Reinforcement Learning algorithm** with novel specifications which can obtain **response-adapted** optimal treatment policies in a simulated environment
- **Computational framework** which contributes to the body of literature paving the way for AI-driven tumor xenograft experiments

- Simulate real data by applying noise to the parameters of the dynamical systems model that governs the environment.

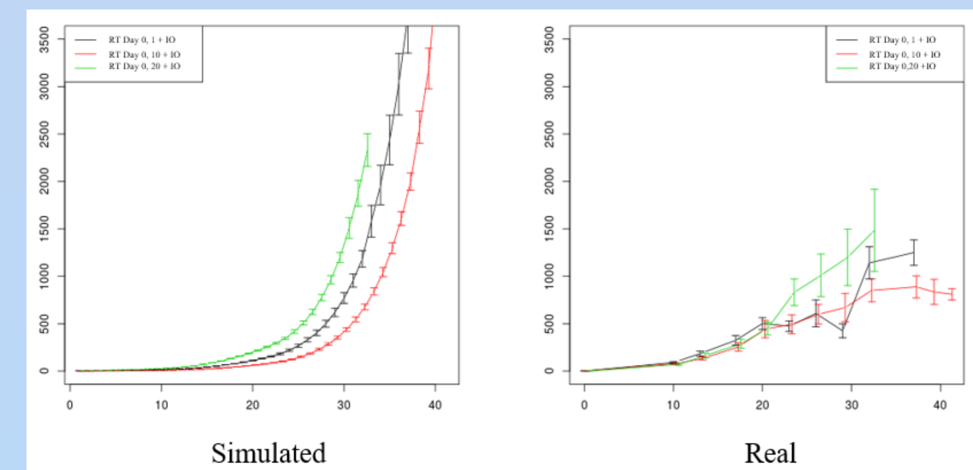

Fig.4: Error bars on simulated vs real data are of comparable length.

Results

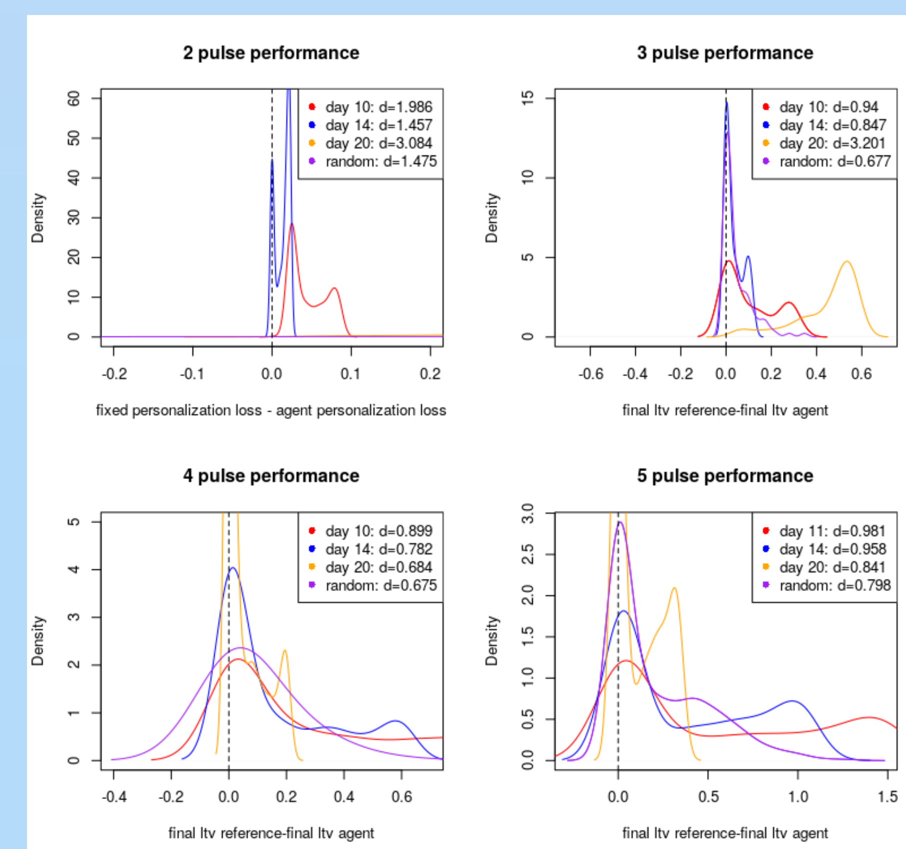

Fig.5: Using the agent results in lower eventual tumor volumes than using fixed or random spacing. LTV = log (final day) tumor volume, and d is the effect size.

CONCLUSIONS

Q-learning shows promise as a tool for learning response-adaptive personalized fractionation plans.

SELECTED REFERENCES

Moore, Casey, et al. Personalized Ultrafractionated Stereotactic Adaptive Radiotherapy (PULSAR) in Preclinical Models Enhances Single-Agent Immune Checkpoint Blockade. *International Journal of Radiation Oncology* Biology* Physics* (2021).

Yauney, G., & Shah, P. Reinforcement learning with action-derived rewards for chemotherapy and clinical trial dosing regimen selection. *Machine Learning for Healthcare Conference* (2018).