

Large-scale automatic carbon ion treatment planning for head and neck cancers via parallel multi-agent reinforcement learning

Jueye Zhang^{1,2,3}, Chao Yang³, Youfang Lai³, Kai-Wen Li³, Wenting Yan³, Yunzhou Xia³, Haimei Zhang³, Jingjing Zhou³, Gen Yang¹, Chen Lin¹, Tian Li⁴, Yibao Zhang^{2,*}

 1.Peking University 2.Peking University Cancer Hospital 3.CAS Ion Medical Technology 4.Hong Kong Polytechnic University

ABSTRACT

Purpose. Head-and-neck cancer (HNC) treatment planning is challenging due to complex target volumes and closely located organs-at-risk (OARs). Although intensity-modulated carbon-ion therapy (IMCT) improves dose conformity and OAR sparing, its planning remains time-consuming due to extensive treatment-planning parameter (TPP) tuning. Existing learning-based approaches face limitations in plan optimality and efficient exploration of the high-dimensional TPP space.

Methods. We developed a scalable multi-agent reinforcement learning (MARL) framework for automated IMCT planning, enabling simultaneous optimization of 45 TPPs. A centralized-training decentralized-execution QMIX architecture was integrated with improved value estimation and recurrent state encoding. Compact DVH-based states, a clinically informed reward function, and a parallel TPS-interaction framework were introduced to improve optimization efficiency.

Results. On 20 head-and-neck cases (10 training and 10 testing), the proposed method generated clinically competitive plans compared with expert manual plans. The relative plan score was $85.93 \pm 7.85\%$ for MARL and $85.02 \pm 7.92\%$ for manual planning, with significant improvements in five OARs ($p < 0.05$).

Conclusions. The proposed MARL framework enables efficient high-dimensional TPP optimization and demonstrates the potential for automated IMCT planning with improved OAR sparing.

CONTACT

Zhang Yibao

 Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Radiation Oncology, Peking University Cancer Hospital & Institute

 Beijing 100142 China

 zhangyibao@pku.edu.com

INTRODUCTION

Multiple OARs + Additional Physical & Biological Modelling $\Rightarrow$ IMCT Planning Challenging

Limitations:

1. Manual planning: labor-intensive, experience-dependent
 2. DL methods: training-data bias, feasibility issues
 3. Existing RL: tune only 4–5 TPPs, or subsets (1–16)
- High-dim action space (3^{45}) hard to explore

Comparison with Existing Works

Name	Algorithm	# TPPs	# Parallel
WTPN, VTPN, KgDRL	IQL	4–5	4–5
HieVTPN, VTP	HieDRL	5–48	1
Qingqing W, et al.	PPO	76	16
Ours	QMIX	45	45

METHODS AND MATERIALS

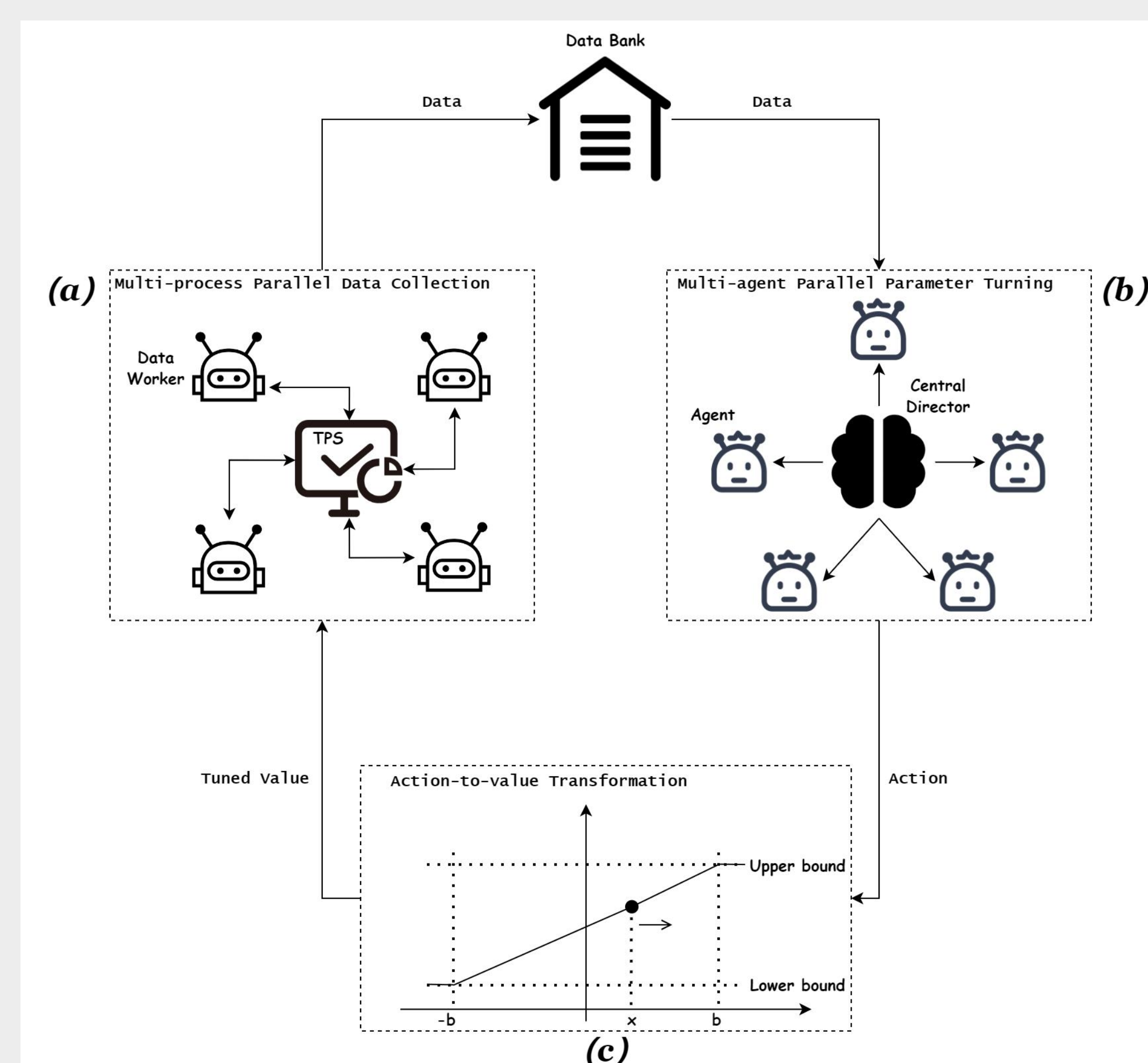
Key Design Choices

- State:** LSTM-encoded DVHs of 19 organs, $s_t \in \mathbb{R}^{19 \times 150}$

Temporal Awareness
- Action:** 45 agents, $a_i \in \{-1, 0, 1\}$; linear transform
- Reward:** $r_t = \phi(s_{t+1}) - 0.5 \times \max \phi(*)$

Clinical Prior
- QMIX:** $Q_{\text{tot}} = f(Q_1, \dots, Q_n)$, $\frac{\partial Q_{\text{tot}}}{\partial Q_i} \geq 0$
- Dueling:** $Q = V + A$
- DDQN:** $y = r + \gamma \hat{Q}(s', \arg \max_{a'} Q(s', a'))$

Training Stabilization



RESULTS

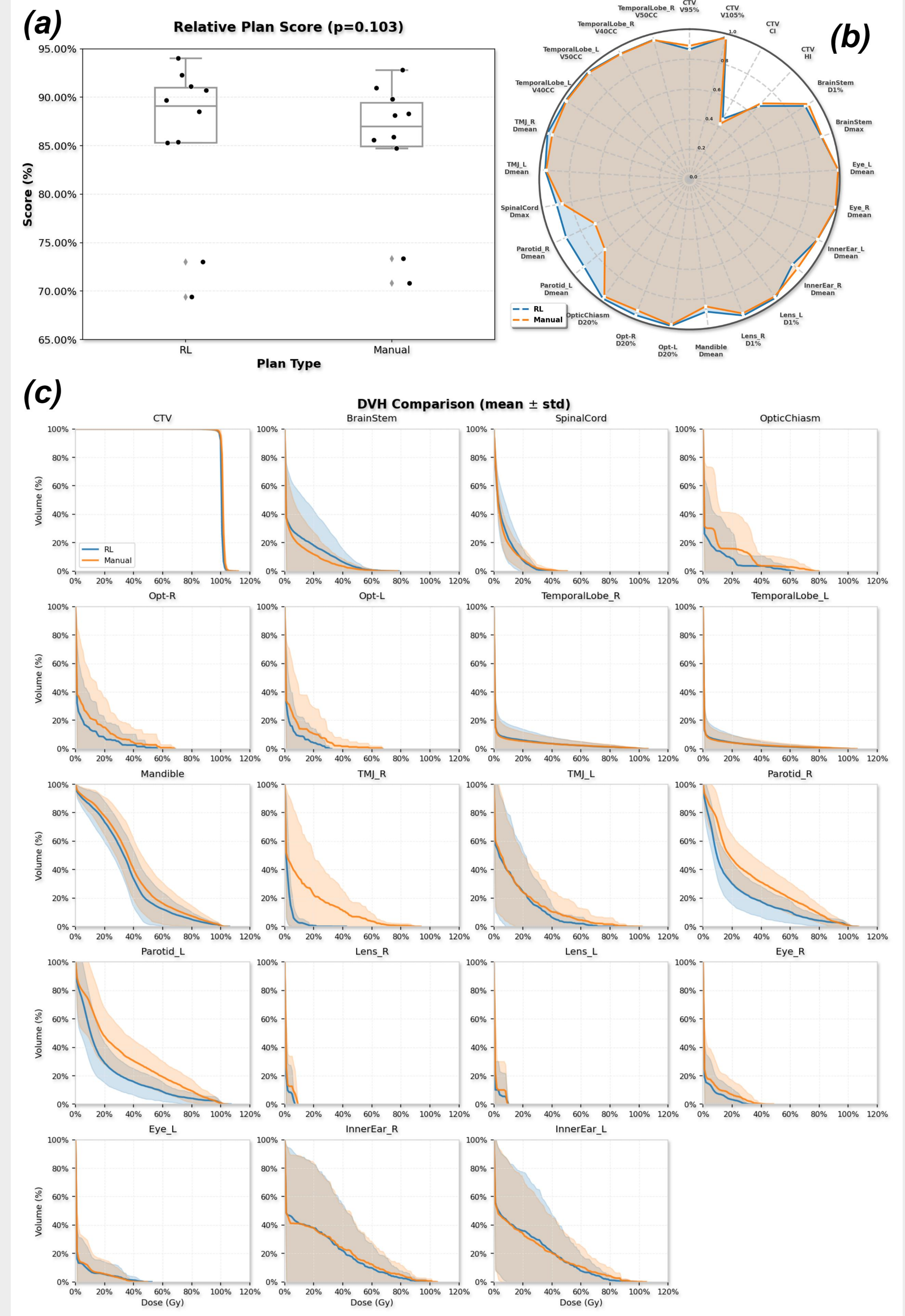


Figure 1. Statistical comparison between RL and Manual plans. (a) Box plot of relative plan score. (b) Radar chart of relative score. (c) Mean DVHs.

CONCLUSIONS

1. Largest parallel scale: tuning 45 TPPs simultaneously
2. 5 OARs significantly better spared than manual plans