A deep reinforcement learning based neural network for beam orientation selection problem in the treatment planning of the IMRT prostate cancer.

JULY 12–16 2020 VIRTUAL
JOINT AAPM | COMP MEETING
EASTERN TIME [GMT-4]

A. Sadeghnejad Barkousaraie, G. Bohara, S. Jiang, D. Nguyen
Medical Artificial Intelligence and Automation (MAIA) Laboratory, Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, TX, USA

MAIA Laboratory
UT Southwestern Medical Center
Radiation Oncology

INTRODUCTION

- Beam orientation optimization (BOO) finds a suitable set of IMRT beam angles.
- Selection of beams affects the quality of the treatment dramatically.
- Traditional BOO algorithms:
  - requires pre-dose calculation for many candidate beams (time consuming)
  - Difficulty to explore the huge solution space

AIM

To introduce **a self-learning deep neural network** to predict a set of beam orientations for intensity modulated radiation therapy (IMRT) that can:

- **outperform a current state-of-the-art** optimization method, column generation (CG)
- develop a high-quality plan that can **be extended to complex problems** such as 4π radiation therapy and proton therapy.

METHOD

A deep reinforcement learning based neural network that can self-improve over time.

The proposed method starts with a previously trained deep neural network (DNN)[1], that has been trained to mimic CG performance by iteratively selecting one beam at a time.

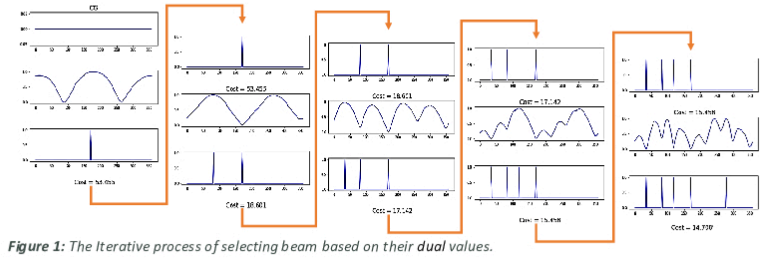


Figure 1: The iterative process of selecting beam based on their dual values.

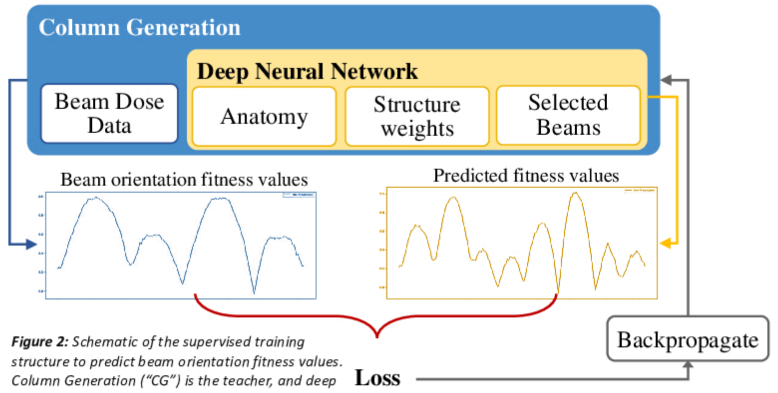


Figure 2: Schematic of the supervised training structure to predict beam orientation fitness values. Column Generation ("CG") is the teacher, and deep neural network ("DNN") is the trainee.

- *A self-improving tree structure is embedded in the model after the selection of each set of beam.*
- *A tree-based structure is added to the model that can learn and improve its own beam selection policy.*
- *Each tree will use the current model to traverse through the decision space and later update the current model based on the improved results.*

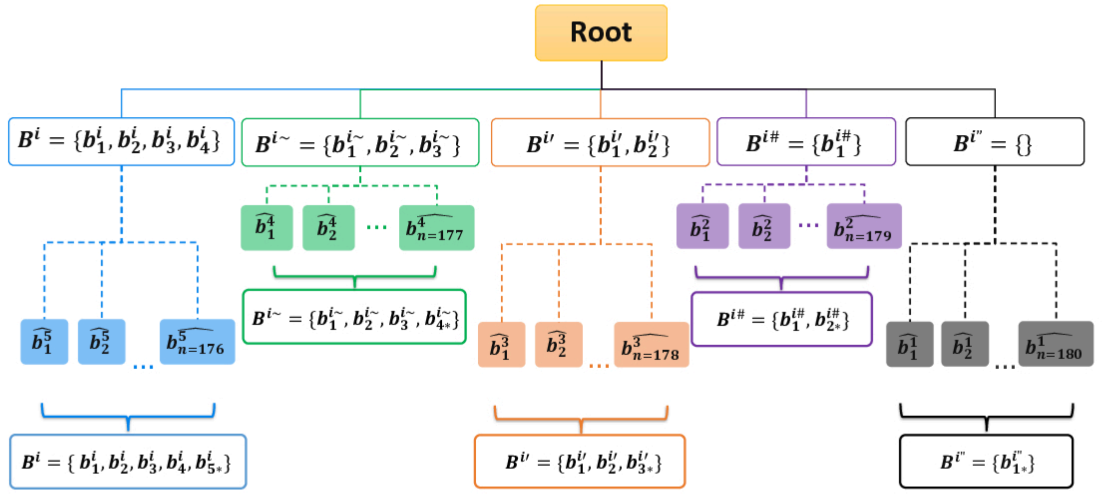


Figure 3. *Self improving tree structure, with total simulation of the next move exploration strategy. For each input all possible children will be predicted (using the DNN model), then, for each of these children the reward value will be calculated. These reward values will be used to train The DNN model. The reward value is usually a function of the fluence map optimization problem(FMO) [2] of a selected set of beams. e.g. for zero input beams, 180 possible children are predicted and their reward value is calculated to update the predicted value of zero input beam. For 3 input beams, 177 possible children are predicted and their reward values are used to further train DNN model.*

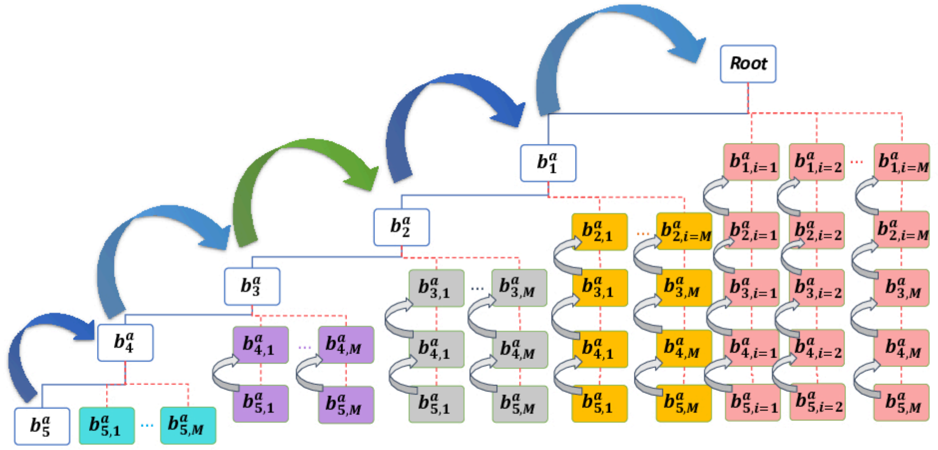


Figure 4 . *Fully explored updating strategy.*
For each child generated using Figure 3, a whole plan of 5 beams will be created by iterative selection of argmax beam values predicted by current DNN model. To save calculation time, instead of calculating FMO directly, the reward value is calculated by using a beam tunable neural network[3] to predict the near optimal treatment plan (dose prediction) for the selected 5 beams. From the predicted dose of the plan the value of the objective function $-\min_x \frac{1}{2}\sum_{\forall s\in S} w_s^2\|D_s x - p\|_2^2 \quad s.t.\ x \geq 0$, *where* $D_s x$ *is the dose delivered to structure s, and p is the prescribed dose—is calculated. The reverse value of the objective function value is used to retrain the network.*

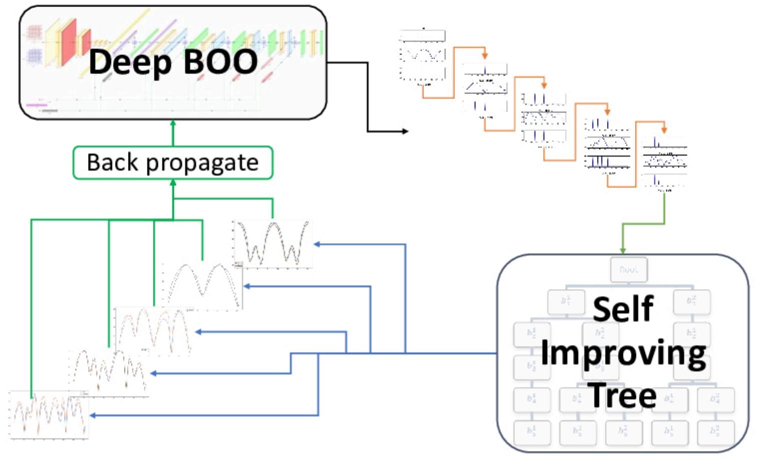


Figure 5 . *General Structure of the method. The DNN model will be dynamically trained during the process by using the updating strategy for input beams from 0 to 4.*

RESULTS

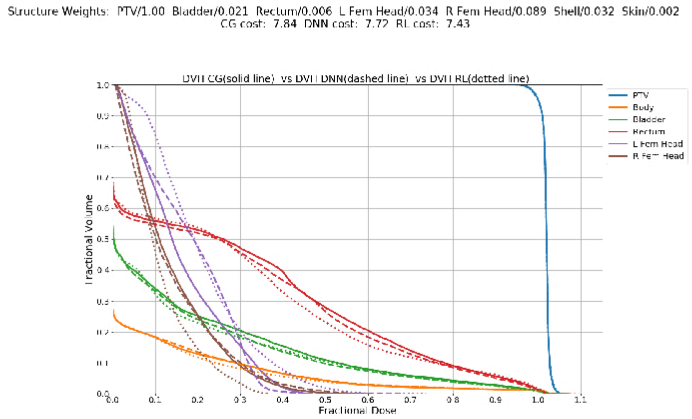


Figure 6. *DVH Graph of CG, DNN and RL (proposed) methods*

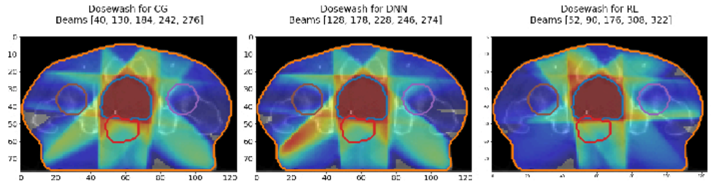
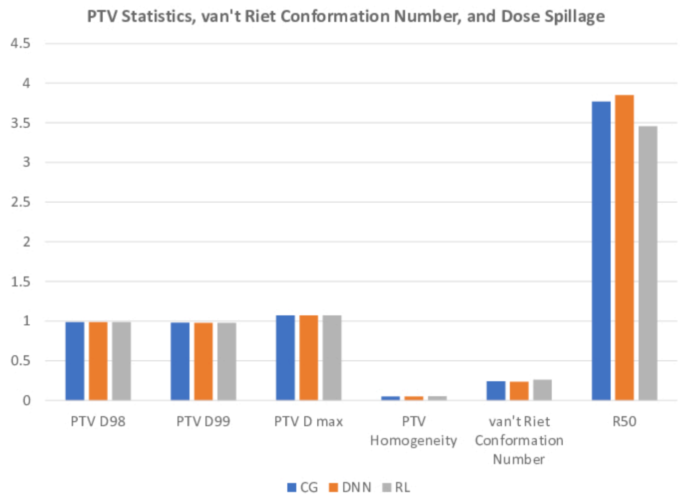
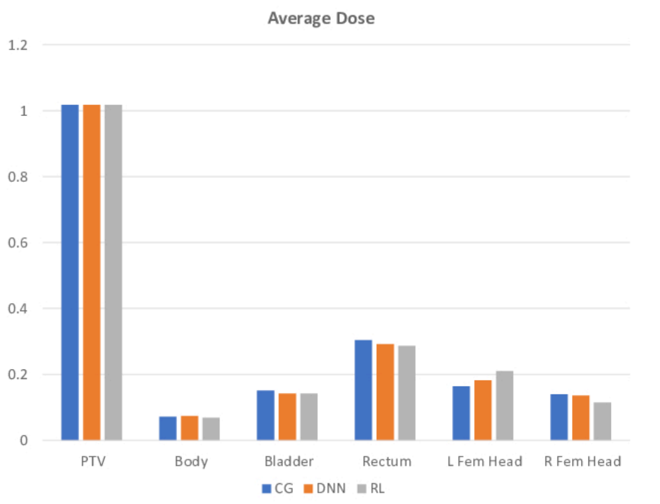


Figure 7. *Dosewash of the treatment plan generated by CG, DNN and RL*





Our preliminary results show more than 20% and 33% improvement in the objective function of the plan generated based on the beams chosen by RL compare to CG and DNN respectively.

CONCLUSION

We propose a strong and fast reinforcement learning method to help for the selection of beam orientations in prostate cancer patients. This model can find this set of beam orientations in at most 2 seconds.

This is an ongoing project. More efficient updating strategy can be used in this work, which is currently being studied.

REFERENCES

1. Sadeghnejad Barkousaraie, A., Ogunmolu, O., Jiang, S. and Nguyen, D. (2020), A fast deep learning approach for beam orientation optimization for prostate cancer treated with intensity-modulated radiation therapy. Med. Phys., 47: 880-897.
2. Cabrera G, G., et al., A metaheuristic approach to solve the multiobjective beam angle optimization problem in intensity-modulated radiation therapy. International Transactions in Operational Research, 2018. 25(1): p. 243-268.
3. Bohara G, Sadeghnejad-Barkousaraie AS, Jiang S, Nguyen D. Using Deep Learning to Predict Beam-Tunable Pareto Optimal Dose Distribution for Intensity Modulated Radiation Therapy. arXiv preprint arXiv:2006.11236. 2020 Jun 19.

CONTACT INFORMATION

Azar.Barkousaraie@UTSouthwestern.edu, Dan.Nguyen@UTSouthwestern.edu