A Machine Learning Model for Brain V12Gy/V60% Prediction of LINAC-based Single-Iso-Multiple-Targets (SIMT) Stereotactic Radiosurgery (SRS): a Longitudinal Study

Xinyi Li, Jiahan Zhang, Yang Sheng, Kyle Lafata, Neville Eclov, Yunfeng Cui, William Giles, Justus Adamson, Anna Rodrigues, Zhiheng Wang, Sua Yoo, Fang-Fang Yin, Q. Jackie Wu, Chunhao Wang
Department of Radiation Oncology, Duke University Medical Center, Durham, NC

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

Introduction

In SIMT SRS, V12Gy is an important dosimetric index as normal tissue toxicity (radiation-induced brain radionecrosis) indicator. A prediction of achievable V12Gy/V60% (assuming 20Gy x 1fx) can assist physicians in the determination of fractionation schemes (single fx vs. multiple fx). Such predictions can also assist planners to generate a SIMT plan more rapidly with a clear planning goal, and hence fewer trials rerunning the time-consuming inverse planning. The developed models have been compiled as a script-based GUI in a commercial TPS, and it is easy to use as a 'one-click execution'. In the long term, the use of the model could reduce V12Gy/V60% intra-planner and inter-planner variability.

Methods

A. Overall Study Design: As shown in Figure 1, four versions of the model were trained and studied. $Model_{v1}$ was trained on 71 plans made by multiple planners (2~25 targets, average=6). The dose prescriptions of these plans include 18/20/22Gy in 1 fraction, 24 Gy in 3 fractions, and 25/27.5Gy in 5 fractions. $Model_{v1\text{-}Alpha}$ was trained on 27 out of the 71 plans (2~14 target, average=5) which were made by the same planner, and this planner will be referred to as Alpha. $Model_{v1}$ and $Model_{v1\text{-}Alpha}$ were implemented in the clinical treatment planning system as one-click scripting-based GUI execution. During the following 3-month study period, the V12Gy/V60% predictions from both models were provided to the planners. If the V12Gy/V60% was higher than the prediction range, the case would be recommended to be tentatively replanned. Subsequently, 17 accrued plans (2~11 targets, average=4), including 8 by Alpha, including 8 by Alpha were added to the training data, while 17 oldest cases were removed, including 8 by Alpha. $Model_{v2}$/$Model_{v2\text{-}Alpha}$ were thus trained using the same methodology and the same number of training cases with $Model_{v1}$/$Model_{v1\text{-}Alpha}$.

B. Machine Learning Model: The prediction of V12Gy/V60% (PTV volumes included) utilized a Gradient Boosted Trees (GBT) Regression model. The hyperparameters of the GBT model, which are user-defined parameters that define search tree depths and sizes, were fine-tuned by a random search optimizer.

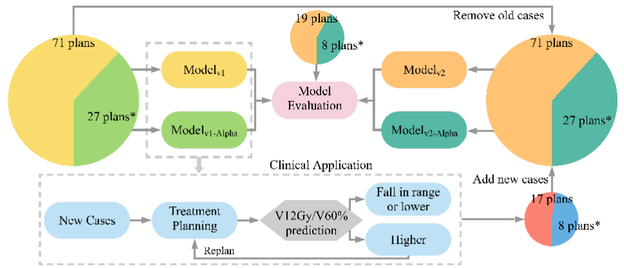

*Figure 1. Overall workflow. * plans were made by Alpha.*

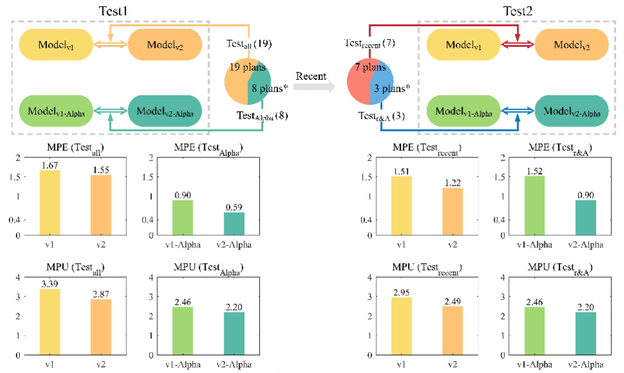

Figure 2. Comparison of models' performance. MPE: mean prediction error. MPU: mean prediction uncertainty. r&A: test cases made recently by Alpha.

The prediction uncertainty of a test case is estimated by the mean prediction error in the validation group.

C. Evaluation: During the 3-month study period, a new plan would be identified as suboptimal if its V12Gy/V60% is higher than one of the model predictions. Suboptimal plans were replanned to see if V12Gy/V60% could be further reduced. The predictions of the test cases from all 4 models were compared with the ground-truth values by mean prediction error (MPE, the average of prediction error) and mean prediction uncertainty (MPU, the average of prediction uncertainty). Nineteen plans were spared for the independent test.

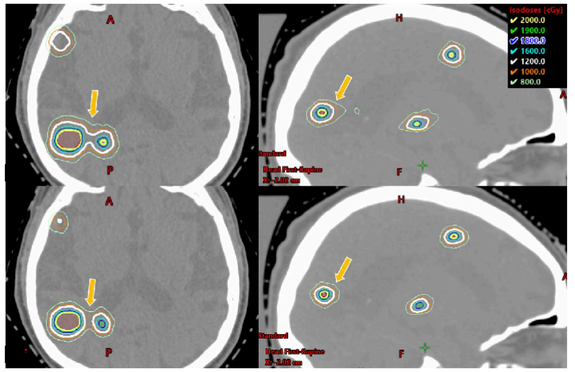

Figure 3. Isodose distribution comparison before replanning (left column) and after replanning (right column) of case #4 (upper row, transversal view) and case #2 (lower row, sagittal view) in Table 1.

Table 1. V12Gy/V60% and its prediction before and after replanning.

Case #	Before / After	V12Gy/ V60% (cc)	$Model_{v1}$ (cc)	In/out range	$Model_{v1\text{-}Alpha}$ (cc)	In/out range
1	Before	58.35	54.58±9.20	In	49.34±7.14	Higher
	After	50.91		In		In
2	Before	13.85	14.84±2.69	In	10.85±2.09	Higher
	After	12.04		Lower		In
3	Before	32.85	33.40±6.30	In	27.69±4.89	Higher
	After	28.70		In		In
4	Before	28.33	20.62±3.19	Higher	19.21±2.48	Higher
	After	17.37		Lower		In

For Test1, patient cohort $Test_{all}$, including all 19 test cases, was used to compare the performance between $Model_{v1}$ and $Model_{v2}$. Patient cohort $Test_{Alpha}$, including all 8 test cases planned by Alpha, was used to compare the performance between $Model_{v1\text{-}Alpha}$ and $Model_{v2\text{-}Alpha}$. For Test2, patient cohort $Test_{recent}$, including all 7 test cases planned during the 3-month study period, was used to compare multi-planner plan quality consistency before and after model application. Patient cohort $Test_{r\&A}$, including all 3 test cases planned by Alpha during the 3-month study period, was used to compare single-planner (Alpha) plan quality consistency before and after model application.

Results

A. Identify suboptimal plans: Table 1 listed 4 cases that were replanned during the 3-month study period. Replanning results indicated that the V12Gy/V60% could be further reduced to the prediction ranges or lower. Figure 3 shows the isodose distribution of two replanned cases before and after replanning. After replanning, the 12Gy/60% isodose line (white) is more conformal to the PTVs (red) where the yellow arrows are pointing in both cases. These results suggest that $Model_{v1}$ and $Model_{v1\text{-}Alpha}$ successfully identified the outlier plans that can be further improved.

B. Test1: ML model performance has been improved: As shown in Figure 2, in Test1, $Model_{v2}$ has lower MPE and MPU compared to $Model_{v1}$ on $Test_{all}$. $Model_{v1}$/$Model_{v2}$ accurately predicted 14/17 cases out of 19 cases in $Test_{all}$. $Model_{v2\text{-}Alpha}$ has lower MPE and MPU compared to $Model_{v1\text{-}Alpha}$ on $Test_{Alpha}$. $Model_{v1\text{-}Alpha}$/$Model_{v2\text{-}Alpha}$ accurately predicted 6/8 cases out of 8 in $Test_{Alpha}$. After the model's clinical application, MPE and MPU decreased, while the number of accurately predicted cases increased.

Test1 results indicate that the model's performance was improved after updating training cases.

C. Test2: SIMT plan quality consistency has been improved: As shown in Figure 2, in Test2, $Model_{v1}$ has lower MPE and MPU compared to $Model_{v2}$ on $Test_{recent}$. $Model_{v1}$/$Model_{v2}$ accurately predicted 5/7 cases out of 7 cases in $Test_{recent}$. $Model_{v1\text{-}Alpha}$ has lower MPE and MPU compared to $Model_{v2\text{-}Alpha}$ on $Test_{r\&A}$. $Model_{v1\text{-}Alpha}$/$Model_{v2\text{-}Alpha}$ accurately predicted 1/3 cases out of 3 in $Test_{r\&A}$. In other words, $Model_{v2}$/$Model_{v2\text{-}Alpha}$ has better performance on recent test cases compared to $Model_{v1}$/$Model_{v1\text{-}Alpha}$.

These results indicate that both inter- and intra- planner plan quality consistency improved over the 3-month study period.

Conclusion

An ML model for SIMT SRS V12Gy/V60% prediction was successfully developed. The presented longitudinal study suggests the great value of the model's application in SIMT planning quality consistency improvement.