Physics and Imaging in Radiation Oncology 16 (2020) 99–102

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Short Communication

Using prediction models to evaluate magnetic resonance image guided radiation therapy plans

M. Allan Thomas [a,c], Joshua Olick-Gibson [a], Yabo Fu [a,d], Parag J. Parikh [b], Olga Green [a], Deshan Yang [a,*]

[a] Department of Radiation Oncology, Washington University in St. Louis, St. Louis, MO 63108, United States
[b] Department of Radiation Oncology, Henry Ford Cancer Institute, Detroit, MI 48202, United States
[c] Department of Imaging Physics, UT MD Anderson Cancer Center, Houston, TX 77030, United States
[d] Department of Radiation Oncology, Emory University, Atlanta, GA 30332, United States

ARTICLE INFO

Keywords:
Neural network
Adaptive radiation therapy
Treatment plan quality
Magnetic resonance image guidance

ABSTRACT

Comprehensive analysis of daily, online adaptive plan quality and safety in magnetic resonance imaging (MRI) guided radiation therapy is critical to its widespread use. Artificial neural network models developed with offline plans created after simulation were used to analyze and compare online plans that were adapted and reoptimized in real time prior to treatment. Roughly one third of ^{60}Co adapted plans were of inferior quality relative to fully optimized, offline plans, but MRI-linac adapted plans were essentially equivalent to offline plans. The models also enabled clear justification that MRI-linac plans are superior to ^{60}Co in an overwhelming majority of cases.

1. Introduction

The use of daily, online adaptive magnetic resonance imaging guided radiation therapy (MRgRT) has grown recently across a variety of clinics. As a result, the potential benefits and practical difficulties of online adaptive MRgRT are beginning to be understood [1–7]. Developing and assessing treatment planning processes and workflows for MRgRT remains a challenge. Daily changes in patient anatomy up to 3 cm in magnitude are possible [8,9,10]. Adapted plans cannot be optimized and scrutinized with the same level of time and effort as plans developed offline because daily adaptive decisions are being made while the patient is on table [1,2]. Furthermore, using plan-specific optimization parameters to create high quality offline plans at patient simulation can lead to subpar adapted plans with substantionally reduced target coverage [7],.

Even with the growth of online adaptive MRgRT, it remains difficult to assess overall online adapted plan quality relative to fully optimized, offline plans [1,2],[11]. Because it is difficult to simulate the inherent complexities and timely decisions associated with MRgRT adaptive workflows [12], actual approved and treated plans offer the best opportunity to assess and improve online adaptive RT. There is very limited previous work where real, clinically treated plans were used to compare online adaptive and offline MRgRT [13], or ^{60}Co and MRI-linac capabilities. The main objective in this study was therefore to use artificial neural network (ANN) models to analyze a wide variety of previously approved and treated MRgRT plans in order to achieve two primary goals: 1) explore online adaptive plan variability and quality relative to high-quality, offline plans; and 2) compare and contrast ^{60}Co- and linac-based MRgRT.

2. Materials and methods

2.1. Patient characteristics

A total of 125 patients with abdominal cancers treated at our institution with high biologically effective dose (BED), online adaptive MRgRT were used for analysis. Online adaptive MRgRT has been used for abdominal, lung, pelvis, and breast cancers [11], but this study focused on abdominal cancers for two primary reasons: 1) natural alignment with the benefits of MRgRT in terms of enhanced soft tissue contrast imaging and daily anatomy changes, and 2) abdominal cancer cases produced the highest percentage of plans requiring daily adaptation at our institution [1,2,11]. Various treatment sites were included such as pancreas, liver, adrenal, bile duct, etc. but most cases (67%) were pancreas cancer. The patients were stratified based on the type of MRgRT: ^{60}Co (n = 70) or MRI-linac (n = 55). All patients were treated

* Corresponding author at: Department of Radiation Oncology, Washington University, School of Medicine, St. Louis, MO 63110, United States.
E-mail address: yangdeshan@wustl.edu (D. Yang).

https://doi.org/10.1016/j.phro.2020.10.002
Received 3 June 2020; Received in revised form 1 October 2020; Accepted 2 October 2020
Available online 28 October 2020

with one of two high BED protocols as discussed in detail previously [14],. Overall, 781 of 975 (80%) treated plans were adapted, so a total of 125 offline and 781 online adapted plans were included. Offline plans were created on the patient's simulation image, received standard planning time, optimization, and checks just like traditional IMRT plans, and served as the starting point of plan adaptation for the first fraction.

2.2. Treatment techniques

Detailed descriptions of the specific workflows and treatment planning methods for adaptive MRgRT using the MRIdian ^{60}Co and MRI-linac systems (Viewray, Cleveland, OH) have been presented previously [1,2,11],[14–17]. The following characteristics were particularly relevant to this work. Both offline and online adaptive plans were developed with *OAR isotoxicity prioritized over target coverage*. Essentially, the OAR constraints discussed in detail previously [14], were hard constraints that could not be exceeded, regardless of the effect on target coverage. Treatment plan deviations manifested mainly in changes in target coverage, so the % of the GTV volume receiving >=95% of the prescription dose (V95) was the main plan quality metric. The prescription dose and dose constraints for OARs were used to guide the plan's optimization. The four critical OARs of stomach, duodenum, small bowel, and large bowel (OAR_{CRIT}) were used in nearly all plans, with other OARs (aorta, esophagus, spinal cord, liver, one or both kidneys) also potentially used but with different dose constraints. The target used for optimization was not the PTV (5 mm isotropic expansion of GTV), but rather the PTV_{OPT} (PTV minus OAR_{5mm}). OAR_{5mm} was OAR_{CRIT} expanded by a 5–8 mm isotropic margin. The majority of patients (~80% in this study) had a 5 mm OAR structure expansion for producing PTV_{OPT} and it was held constant for all plans for each patient.

2.3. ANN prediction models

ANN models to predict voxelized dose inside the GTV were developed using patient anatomy/geometry information only. The model input variables included GTV size/shape, distance relationships between GTV and OARs, and patient size information [18–23]. Additional details of the model development and testing have been published previously [14],. The prediction models were developed using input variables extracted only from offline plans because they received normal planning time, attention, and analysis prior to their approval and use in patients. In contrast, online adaptive plans were not afforded the time to pursue detailed optimization, so their overall quality *a priori* was not known.

A cross validation process like that described in our previous work [14], was used to test the ANN models and assess their accuracy and precision. For each iteration of the cross validation, V95 values for the test group of plans were determined from the 3D dose predictions and raw V95 prediction errors were calculated: $\Delta V95 = V95_{clinical} - V95_{predicted}$. Then the mean error, 95% prediction intervals (PI, ±1.96σ), and 95% confidence intervals (CI) of the mean error and 95% PI were all determined as outlined in Bland-Altman analysis [24],. Limits of agreement (LoA) for each model were calculated as the mean error ± 95% PI. In order to minimize the effect of potential outlier plans (plans both inferior and superior to the average) on the trained models, a model refinement process was also incorporated [14,19,20]. Any plans with $\Delta V95$ outside of the model LoA were excluded, the models were re-trained, and new prediction errors and model metrics were calculated. Model refinement excluded 10 out of the 70 ^{60}Co offline plans and 5 of the 55 linac plans. The refined models were then used for all plan comparison analysis, with inferior, superior, and acceptable plans identified as described in Fig. 2.

Two separate models were developed (^{60}Co and linac) and both used the exact same types of patient anatomy and geometry input variables – those optimized in our previous work on ANN dose prediction models [14],. Adapted plan quality relative to offline plans was determined by inputting the parameters extracted from adapted plans into the models trained with offline plans. The adapted plan predictions from the offline model outputs were then compared to the clinical plan metric. ^{60}Co and linac MRgRT were compared by inputting parameters from ^{60}Co plans into the linac ANN model. Effectively, then, the model outputs reflected the predicted 3D dose distribution that would have been achievable had the ^{60}Co plan actually been planned using the linac.

3. Results

Dose prediction errors for both ^{60}Co and linac models were ~0.2 ± 3.0 Gy when averaged across all plans. Absolute dose errors were ~3.0 ± 2.0 Gy. As shown in Fig. 1, both models produced V95 predictions that strongly correlated with their respective clinical values, maintained minimal bias, and possessed precision within ±6%. In both models ~95% of plans had $\Delta V95$ within the LoA. As seen in Fig. 2(a), nearly one third (157 plans, 30%) of ^{60}Co online adapted plans were deemed inferior, with clinical V95 values outside the lower prediction range of the ANN model. This observation strongly indicates these adapted plans could have achieved improved target coverage if they were developed and optimized offline. Larger deviations were observed as the clinical V95 decreased, showing that more intrinsically difficult cases tended to produce plans that were more inferior. The ^{60}Co plans identified as inferior had statistically significantly lower mean and max OAR_{CRIT} doses relative to those established as adequate quality. Fig. 2(b) shows the overwhelmingly majority (91%) of adapted linac plans had clinical V95 values that fit within the prediction range of the offline model, with only 2% and 7% identified as inferior and superior, respectively. These observations demonstrate that target coverage in linac online adapted plans was essentially equivalent to the expectations set by the offline model.

Fig. 2(c) shows that a large majority (78%) of ^{60}Co adapted plans were identified as inferior to the expectations from the linac model. Furthermore, Table 1 shows nearly 40% of ^{60}Co adapted plans had clinical V95 values > 10% lower, and roughly 7% > 20% lower, than the linac model predictions. Finally, Fig. 2(d) demonstrates that the median (mean) V95 values of the three groups of plans compared progressed from 77.5 (77.4) to 81.6 (81.4) to 87.9 (86.8). Although not shown explicitly in Fig. 2, offline ^{60}Co plans were also deemed to be inferior to the expectations of the linac model in terms of target coverage, but at a slightly reduced rate of ~60%.

4. Discussion

This study used ANN prediction models, bolstered by patient- and plan-specific parameters, to comprehensively compare offline, online adapted, ^{60}Co, and linac plans in MRgRT. Our results showed that many ^{60}Co online adaptive plans, roughly one third, were not able to maintain the same level of target coverage as offline plans. These observations indicate that for one third of ^{60}Co adapted plans, a tradeoff of reduced target coverage relative to the benchmark established by comparable offline plans was required in order to ensure meeting all OAR constraints. The statistically significantly lower mean and max OAR_{CRIT} dose metrics in inferior ^{60}Co (30%) adapted plans suggest the online re-optimization was not able to push OAR doses sufficiently in order to achieve improved target coverage in all ^{60}Co adapted plans. Unlike ^{60}Co, our results showed that MRI-linac adapted plans were able to maintain target coverage expectations that were equivalent to offline plans with comparable intrinsic difficulty. These results establish that linac-based online adaptive MRgRT can maintain important plan quality metrics equivalent to offline plans that received the requisite time and attention to be fully optimized. This is a key observation to boost the clinical confidence in online plan adaptations with linac-MRgRT. Linac plans also outperformed ^{60}Co plans at a rate of nearly 4 out of 5 and the average increase in target coverage (V95) had the plans been developed with the linac was ~10%. This provided further evidence that linac hardware was better able to produce high quality plans.

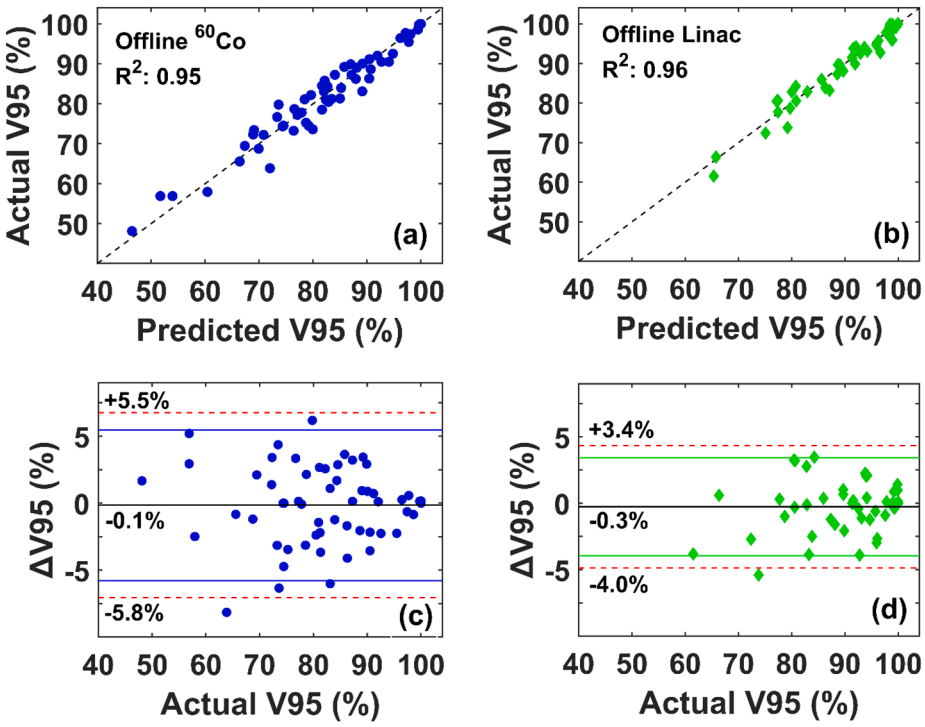

Fig. 1. Results for clinical vs. predicted V95 values for (a) ^{60}Co and (b) linac offline plans. The predicted V95 values come from the respective ANN model 3D dose predictions. The R^2 values of the clinical vs. predicted comparisons are included in (a) and (b). Bland-Altman plots of the V95 prediction errors are shown for (c) ^{60}Co and (d) linac models. The values for mean bias and LoA are indicated in (c) and (d). The mean bias, LoA (solid line), and LoA + 95% CI (dotted line) are also plotted in (c) and (d) to show the precision of the model predictions.

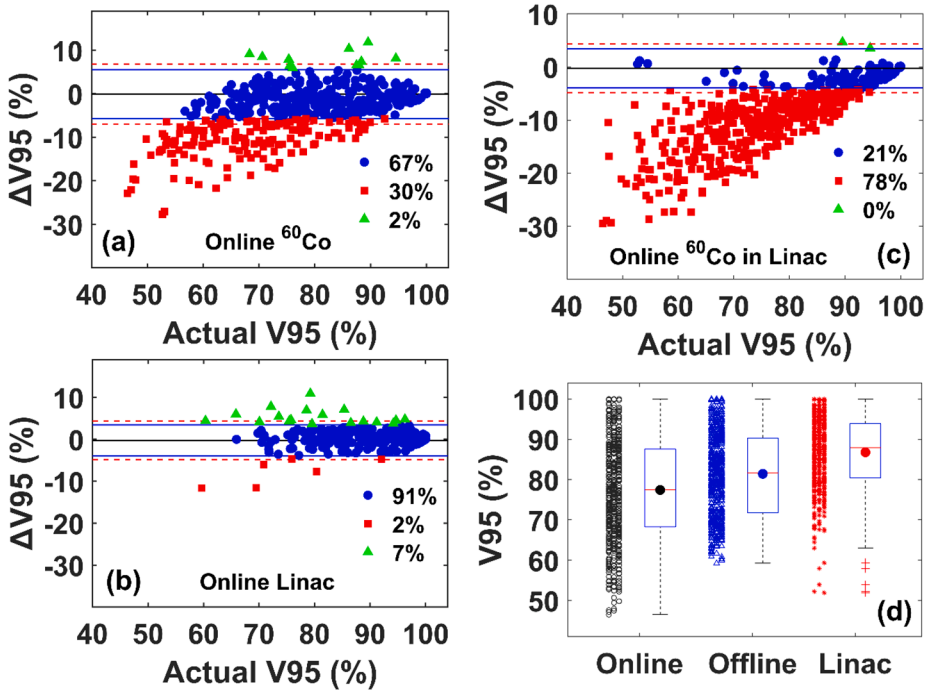

Fig. 2. Bland-Altman plots of V95 predictions for online adapted plans from ANN models developed using offline plans: (a) ^{60}Co, (b) linac, (c) ^{60}Co in linac model. The mean bias and LoA are plotted much like in Fig. 1. In each plot, the data points are defined based on their comparison to the model's LoA: 1) blue circle = acceptable: $\Delta V95$ within LoA, 2) red square = inferior: $\Delta V95 < LoA_{lower}$, 3) green diamond = superior: $\Delta V95 > LoA_{upper}$. The percentages of how many plans fell into the three categories are also indicated. The entire distribution of V95 values as well as boxplots for each distribution are shown in (d): online = clinical ^{60}Co adapted plans, offline = offline ^{60}Co ANN model predictions for ^{60}Co adapted plans, linac = linac ANN model predictions for ^{60}Co adapted plans. The mean values of each distribution are shown with solid dots, the median values are shown as a solid line. The boxes show interquartile ranges of the distributions, while the whiskers include all values up to ± 2.7σ, with any outliers indicated by a + sign.

The details of the ViewRay MRI-linac system hardware have been outlined previously [25],. As discussed in a recent study [12], the distinction between ^{60}Co and linac plans regarding plan quality is mainly due to the higher beam energy (average ~2 MV for linac; 1.25 MV for ^{60}Co) and the improved multi-leaf collimator design in the linac. Online adapted MRI-linac plans were also shown to be roughly comparable in terms of plan quality while offering improved OAR dose metrics relative to original, unadapted plans [13],. Our results were in line with previously established key conclusions about MRgRT but also expanded upon them by analyzing each plan specifically and including intrinsic plan difficulty.

A limitation of this study was that other plan quality metrics such as dose conformality and OAR dose sparing were not easy to compare because our models could only predict GTV dose. Future work will include using more advanced models to expand 3D dose predictions beyond the GTV to explore a more complete picture of plan comparisons. Another limitation was that the models developed and plans analyzed were only from a single institution. MRgRT workflows and

Table 1
Summary of plan comparisons based on V95 predictions for adapted plans from offline plan models.

	^{60}Co Adapted (n = 516)	Linac Adapted (n = 265)	^{60}Co Adapted in Linac (n = 516)
ΔV95 (%): Mean ± σ	−4.0 ± 5.9	0.4 ± 2.4	−9.4 ± 6.5
Acceptable	348 (67%)	241 (91%)	109 (21%)
(mean ± σ)	−1.3 ± 2.5	0.2 ± 1.5	−1.6 ± 1.4
Inferior	157 (30%)	6 (2%)	405 (78%)
(mean ± σ)	−10.9 ± 4.9	−7.7 ± 3.2	−11.5 ± 5.6
Superior	11 (2%)	18 (7%)	2 (0%)
(mean ± σ)	8.0 ± 1.8	5.3 ± 1.9	4.1 ± 0.8
ΔV95 < -10%	80 (16%)	2 (1%)	203 (39%)
ΔV95 < -20%	7 (1%)	0	34 (7%)

online adaptive planning strategies differ across various institutions. We are hopeful that the results presented here are deemed useful for a better understanding of the difficulties and capabilities of online adaptive MRgRT as a rapidly growing application for improved treatment of cancer with RT worldwide.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was partially supported by the Agency for Healthcare Research and Quality (AHRQ) grant number R01-HS022888, National Institute of Biomedical Imaging and Bioengineering (NIBIB) grant R03-EB028427 and National Heart, Lung, and Blood Institute (NHLBI) grant R01-HL148210.

This paper is part of a special issue that contains contributions originally submitted to the scientific meeting MR in RT, which was planned to take place 05/2020, organized by the German Research Center (DKFZ) in Heidelberg. We acknowledge funding by DKFZ for the publication costs of this special issue.

References

[1] Henke L, Kashani R, Robinson C, Curcuru A, DeWees T, Bradley J, et al. Phase I trial of stereotactic MR-guided online adaptive radiation therapy (SMART) for the treatment of oligometastatic or unresectable primary malignancies of the abdomen. Radiother Oncol 2018;126:519–26. https://doi.org/10.1016/j.radonc.2017.11.032.

[2] Henke LE, Olsen JR, Contreras JA, Curcuru A, DeWees TA, Green OL, et al. Stereotactic MR-guided online adaptive radiation therapy (SMART) for ultracentral thorax malignancies: results of a phase 1 trial. Adv Radiat Oncol 2018;4:201–9. https://doi.org/10.1016/j.adro.2018.10.003.

[3] Rudra S, Jiang N, Rosenberg SA, Olsen JR, Roach MC, Wan L, et al. Using adaptive magnetic resonance image-guided radiation therapy for treatment of inoperable pancreatic cancer. Cancer Med 2019;8:2123–32. https://doi.org/10.1002/cam4.2100.

[4] Bohoudi O, Bruynzeel AME, Senan S, Cuijpers JP, Slotman BJ, Lagerwaard FJ, et al. Fast and robust online adaptive planning in stereotactic MR-guided adaptive radiation therapy (SMART) for pancreatic cancer. Radiother Oncol 2017;125:439–44. https://doi.org/10.1016/j.radonc.2017.07.028.

[5] El-Bared N, Portelance L, Spieler BO, Kwon D, Padgett KR, Brown KM, et al. Benefits and practical pitfalls of daily online adaptive MRI-guided stereotactic radiation therapy for pancreatic cancer. Pract Radiat Oncol 2019;9:e46–54. https://doi.org/10.1016/j.prro.2018.08.010.

[6] Tyran M, Jiang N, Cao M, Raldow A, Lamb JM, Low D, et al. Retrospective evaluation of decision-making for pancreatic stereotactic MR-guided adaptive radiotherapy. Radiother Oncol 2018;129:319–25. https://doi.org/10.1016/j.radonc.2018.08.009.

[7] Olberg S, Green O, Cai B, Yang D, Rodriguez V, Zhang H, et al. Optimization of treatment planning workflow and tumor coverage during daily adaptive magnetic resonance imaging guided radiation therapy of pancreatic cancer. Radiat Oncol 2018;13:51. https://doi.org/10.1186/s13014-018-1000-7.

[8] Chen I, Mittauer KE, Henke LE, Acharya S, Lu Y, Chen, et al. Quantification of interfractional gastrointestinal tract motion for pancreatic cancer radiation therapy. Int J Radiat Oncol Biol Phys 2016;96:E144. https://doi.org/10.1016/j.ijrobp.2016.06.954.

[9] Abbas H, Chang B, Chen Z. Motion management in gastrointestinal cancers. Int J Radiat Oncol Biol Phys 2014;5:223–35. https://doi.org/10.3978/j.issn.2078-6891.2014.028.

[10] Liu F, Erickson B, Peng C, Li XA. Characterization and management of interfractional anatomic changes for pancreatic cancer radiotherapy. J Med Imaging Radiat Sci 2012;83:e423–9. https://doi.org/10.1016/j.ijrobp.2011.12.073.

[11] Fischer-Valuck BW, Henke L, Green O, Kashani R, Acharya S, Bradley JD, et al. Two-and-a-half year clinical experience with the world's first magnetic resonance image guided radiation therapy system. Adv Radiat Oncol 2017;2:485–93. https://doi.org/10.1016/j.adro.2017.05.006.

[12] Ramey SJ, Padgett KR, Lamichhane N, Neboori HJ, Kwon D, Mellon EA, et al. Dosimetric analysis of stereotactic body radiation therapy for pancreatic cancer using MR-guided Tri- 60Co unit, MR-guided LINAC, and conventional LINAC-based plans. Pract Radiat Oncol 2018;8:e312–21. https://doi.org/10.1016/j.prro.2018.02.010.

[13] van Timmeren JE, Chamberlain M, Krayenbuehl J, Wilke L, Ehrbar S, Bogowicz M, et al. Treatment plan quality during online adaptive re-planning. Radiat Oncol 2020;15:203. https://doi.org/10.1186/s13014-020-01641-0.

[14] Thomas MA, Fu Y, Yang D. Development and evaluation of machine learning models for voxel dose predictions in online adaptive magnetic resonance guided radiation therapy. J Appl Clin Med Phys 2020;21:60–9. https://doi.org/10.1002/acm2.12884.

[15] Li HH, Rodriguez VL, Green OL, Hu Y, Kashani R, Wooten HO, et al. Patient-specific quality assurance for the delivery of Co-60 intensity modulated radiation therapy subject to a 0.35-T lateral magnetic field. Int J Radiat Oncol Biol Phys 2015;91:65–72. https://doi.org/10.1016/j.ijrobp.2014.09.008.

[16] Green OL, Henke LE, Hugo GD. Practical clinical workflows for online and offline adaptive radiation therapy. Semin Radiat Oncol 2019;29:219–27. https://doi.org/10.1016/j.semradonc.2019.02.004.

[17] Wang Y, Mazur TR, Green O, Hu Y, Li H, Rodriguez V, et al. A GPU-accelerated Monte Carlo dose calculation platform and its application toward validating an MRI-guided radiation therapy beam model. Med Phys 2016;43:4040–52. https://doi.org/10.1118/1.4953198.

[18] Ma M, Kovalchuk N, Buyyounouski MK, Xing L, Yang Y. Dosimetric features-driven machine learning model for DVH prediction in VMAT treatment planning. Med Phys 2019;46:857–67. https://doi.org/10.1002/mp.13334.

[19] Shiraishi S, Tan J, Olsen LA, Moore KL. Knowledge-based prediction of plan quality metrics in intracranial stereotactic radiosurgery. Med Phys 2015;42:908. https://doi.org/10.1118/1.4906183.

[20] Shiraishi S, Moore KL. Knowledge-based prediction of three-dimensional dose distributions for external beam radiotherapy. Med Phys 2016;43:378–87. https://doi.org/10.1118/1.4938583.

[21] Campbell WG, Miften M, Olsen L, Stumpf P, Schefter T, Goodman KA, et al. Neural network dose models for knowledge-based planning in pancreatic SBRT. Med Phys 2017;44:6148–58. https://doi.org/10.1002/mp.12621.

[22] Wu B, Ricchetti F, Sanguineti G, Kazdan M, Simari P, Jacques R, et al. Data-driven approach to generating achievable dose-volume histogram objectives in intensity-modulated radiotherapy planning. Int J Radiat Oncol Biol Phys 2011;79:1241–7. https://doi.org/10.1016/j.ijrobp.2010.05.026.

[23] Zhu X, Ge Y, Li T, Thongphiew D, Yin FF, Wu QJ. A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. Med Phys 2011;38:719–26. https://doi.org/10.1118/1.3539749.

[24] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;327:307–10.

[25] Kluter S. Technical design and concept of a 0.35 T MR-Linac. Clin. Transl. Radiat. Oncol. 2019:98–101. https://doi.org/10.1016/j.ctro.2019.04.007.