Radiotherapy and Oncology 202 (2025) 110609

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Knowledge-based planning for fully automated radiation therapy treatment planning of 10 different cancer sites

Christine V. Chung [a,*], Meena S. Khan [a], Adenike Olanrewaju [a], Mary Pham [a], Quyen T. Nguyen [a], Tina Patel [c], Prajnan Das [b], Michael S. O'Reilly [b], Valerie K. Reed [b], Anuja Jhingran [b], Hannah Simonds [d], Ethan B. Ludmir [b], Karen E. Hoffman [b], Komeela Naidoo [d], Jeannette Parkes [e], Ajay Aggarwal [c], Lauren L. Mayo [b], Shalin J. Shah [b], Chad Tang [b], Beth M. Beadle [f], Julie Wetter [e], Gary Walker [g], Simon Hughes [c], Vinod Mullassery [c], Stephen Skett [c], Christopher Thomas [c], Lifei Zhang [a], Son Nguyen [a], Raymond P. Mumme [a], Raphael J. Douglas [a], Hana Baroudi [a,h], Laurence E. Court [a,b]

[a] Department of Radiation Physics, University of Texas MD Anderson Cancer Center, Houston, TX, USA
[b] Department of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, TX, USA
[c] Department of Oncology, Guy's and St Thomas' NHS Foundation Trust, London, UK
[d] Division of Radiation Oncology, Stellenbosch University, Stellenbosch, South Africa
[e] Division of Radiation Oncology, University of Cape Town, Cape Town, South Africa
[f] Department of Radiation Oncology, Stanford University, Stanford, CA, USA
[g] Department of Radiation Oncology, Banner MD Anderson Cancer Center, Gilbert, AZ, USA
[h] Graduate School of Biomedical Sciences, University of Texas, Houston, TX, USA

ARTICLE INFO

Keywords:
Treatment planning
Artificial intelligence
Knowledge-Based Planning
Automation
VMAT

ABSTRACT

Purpose: Radiation treatment planning is highly complex and can have significant inter- and intra-planner inconsistency, as well as variability in planning time and plan quality. Knowledge-based planning (KBP) is a tool that can be used to efficiently produce high-quality, consistent, clinically acceptable plans, independent of planner skills and experience. In this study, we created and validated multiple clinically acceptable and fully automatable KBP models, with the goal of creating VMAT plans without user intervention.

Methods: Ten KBP models were configured using high quality clinical plans from a single institution. They were then honed to be part of a fully automatable system by incorporating scriptable planning structures, plan creation, and plan optimization. These models were verified and validated using quantitative (model statistics) and qualitative (dose-volume histogram estimation review) analysis. The resulting KBP-generated plans were reviewed by physicians and rated for clinical acceptability.

Results: Autoplanning models were created for anorectal, bladder, breast/chest wall, cervix, esophagus, head and neck, liver, lung/mediastinum, prostate, and prostate with nodes treatment sites. All models were successfully created to be part of a fully automated system without the need for human intervention to create a fully optimized plan. The physician review indicated that, on average, 88% of all KBP-generated plans were "acceptable as is" and 98% were "acceptable after minor edits."

Conclusion: KBP models for multiple treatment sites were used as a basis to generate fully automatable, efficient, consistent, high-quality, and clinically acceptable plans. These plans do not require human intervention, demonstrating the potential this work has to significantly impact treatment planning workflows.

* Corresponding author at: Department of Radiation Physics, Unit 1420, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030, USA.
E-mail address: cchung2@mdanderson.org (C.V. Chung).

https://doi.org/10.1016/j.radonc.2024.110609
Received 1 May 2024; Received in revised form 25 October 2024; Accepted 28 October 2024
Available online 30 October 2024

Introduction

Radiation treatment planning is highly complex, with many aspects of planning dependent on the treatment planner's training, skills, and experience [1–4]. This dependency results in considerable inter- and intra-planner inconsistency and variability in plan quality and planning time [1,2,5,6]. Knowledge-based planning (KBP) has been shown to significantly reduce both planning time and plan variability and increase overall plan quality [2,4,7–9], and is available from commercial vendors, such as RapidPlan from Varian Medical Systems (Palo Alto, CA). KBP models use machine learning to evaluate patient geometry referenced against existing treatment plan doses to generate dose-volume histogram (DVH) estimates [10]. These generated DVH estimates can then be loaded into the optimization process to give patient-specific starting points for optimization of intensity-modulated radiation therapy or volumetric-modulated arc therapy [10]. Inputted data are used to customize models to individual clinic standards. Although generating a model has significant upfront cost, the ability to create a valid KBP model is possible for many cancer centers [8] because, in many cases, only 20 cases are needed, plus a recommended 20 additional cases for validation of the model [11].

While the KBP process is typically used to generate initial plan optimization objectives and constraints, additional modification is necessary to create an optimal plan for each individual patient [8]. The treatment planner still needs to continually review the plan and DVH estimates and refine the optimization objectives until an optimal plan is achieved. Olanrewaju et al. [3] and Rhee et al. [1], showed that it is possible, however, to refine the KBP model by the use of additional planning structures and objectives so that no subsequent manual optimization is necessary for the majority of patients. In that case, KBP can form the basis of a fully automated planning process that creates a clinically acceptable plan without user intervention. In this paper, we describe our work in development and validation of KBP-based plan optimization approaches for multiple cancer sites, with the goal of creating clinically acceptable plans without user intervention.

Methods and Materials

Separate KBP models were developed for multiple treatment sites using RapidPlan (Varian Medical Systems, v15.6.06). All procedures were performed in compliance with the Declaration of Helsinki and institutional guidelines, and the use of this work was approved by the University of Texas MD Anderson Institutional Review Board (IRB Number: PA16-0379), including a waiver for informed consent. For all treatment sites, the data collection, data extraction, model training, verification, and validation followed the same procedures, as described below.

Model generation

In the data collection phase, retrospective clinically treated cases were reviewed and selected based on commonalities in clinical variables such as modality, treatment site, prescription dose, and extent of disease. Plans' quality, including organs at risk (OAR) sparing and target coverage, conformality, and homogeneity, were assessed by dosimetrists based on clinical experience; suboptimal cases and those that were considered atypical (e.g., recurrent treatment, unique setups or targets) were filtered out of the cohorts. Each model was configured using anticipated necessary structures, such as targets, OARs, and planning structures (e.g., planning risk volumes, normal tissue, and ring structures; more details can be found in Table 4 per model site in Appendix A). Clinical structures that were deemed relevant but were not included in the original plan (e.g., pelvic bone marrow) were added by the dosimetrist (all manually contoured except for the rectal model, which used deep-learning pelvic OAR contours [1]) during case evaluation; otherwise, all matched Plan Structures were the original clinical contours. Model Structures included were either anatomic structures that could be automated using deep learning models or planning structures (e.g., dummy structures created using Boolean operations) that can be automatically generated using operations in the treatment planning system through the Eclipse application programming interface (API). The goal was to use a minimum of 20 high-quality plans for each model in development and an additional 20 for validation cases (Table 1), although we typically used more, depending on the availability of consistent plans for each site. For models with limited case availability, such as the anorectal short course model, the minimum number of cases that should be used in the DVH estimation was met first, with the remaining cases being used for validation. For models with vast case availability, the minimum number of cases that should be used for validation was met first, with the remaining cases being used for the DVH estimation. As evidenced-based practices evolve with time, case selection prioritized the most recent cases to ensure that results would reflect and include current clinical practices. Once the data is curated for each site, it is inputted to generate a tailored DVH estimation model.

Model analysis

Following model training, ModelAnalytics (MyVarian.com) was used for the initial analysis. Feedback was noted per structure (e.g., highlighted a specific case's OAR that may distort the estimated DVH, suggested addition of more cases to fill data gaps), and edits were made to the model where possible (e.g., removed the specific case's OAR from model, added more cases if available). Further model analysis consisted of reviewing each trained structure's DVH plot, coefficient of determination (R^2 value), and regression plot. DVH plots were used to identify errors such as incorrect prescription entries or structures that did not meet constraints. $R^2 = 0.7$ was used as the targeted lower threshold for each structure, though in some cases this was not achievable if the OAR was variably distanced from the targets. To increase $R^2 < 0.7$ values, regression plots with geometric and dosimetric variables were used to detect outliers that may need to be reviewed or removed from the model to achieve a better fit. This consisted of reviewing outliers to ensure there were no mistakes in contouring (e.g., incorrect laterality associated, incorrect contours, stray pixels). If the contours appeared to be correct, the next step consisted of iteratively trialing removal of outliers and comparing R^2 and regression plots for improvements. Instances of outlier removal that resulted in reduced R^2 value or minimal increase to the R^2 value (<0.05) were added back to preserve variety in the model. The optimal R^2 value per structure was considered achieved when there were no further increases in R^2 values despite changes made to the

Table 1
Number of Treatment Plans Used for Model Training and Validation for Each Treatment Site.

Treatment Site	Model Set, No.	Validation Set, No.
Anorectal*	28	35
Bladder	48	20
Breast/chest wall	100	93
Cervical	120	75
Esophagus	55	20
Head and neck†	–	75
Liver‡	67	20
Lung/mediastinum	78	25
Prostate	95	38
Prostate with nodes	37	20

*Anorectal model was initially generated for supine cases, with the minimum of 20 curated cases for the model. After validation showed feasibility for the supine-only model on prone setups, an additional 8 prone cases were curated and added to the model.
†Head and neck model came with the Eclipse treatment planning system with no model data.
‡Liver model was downloaded from ORBIT-RT.com, then customized and validated.

model.

Model validation

The model validation process had two steps. The first step was to validate the model with clinical data by comparing clinical dose to the DVH estimation bands to determine a good fit, as the *Generated* and *Line* model objectives are produced from the lowest values along the DVH estimation bands. This step was completed on at least 10 patients from the validation set per site to ensure a workable potential for a quality model before moving onto the second step of validation. The validation set comprised additional cases of similar quality and diversity as the model training set (see Table 1 for total number of validation cases per site). Case selection was intended to be robust, meaning that it consisted of variations of plans that the model would be used for. If specific structures were consistently a poor fit for ideal scenario cases, the dosimetrist would iteratively go back and forth between this first step of validation and the model analysis steps to achieve a good fit.

The second step of validation involved creating universal plan setup and planning objectives to be used in every case per model, focusing on *Generated* and *Line* objectives where applicable, though a combination of both automatically generated and manual values proved to be beneficial. Details can be found in the Model Setup section of each site in Appendix A. An iterative process was used to determine the best solution for each treatment site, similar to manual treatment planning. The dosimetrists used their clinical experience to test and introduce fields and planning structures (e.g., normal tissue rings, OAR avoidance structures) that could be automated through the Eclipse API into each model by using combinations of simple operations, such as Boolean, Crop Structure, Expansion, etc., for a potential fully automated optimized plan (Tables 3-4 per site in Appendix A). The final validation was performed methodically identically on the entire validation set of each model (Table 1).

In addition to the new models described above, various additional external models sourced from Eclipse (Varian Medical Systems) and ORBIT-RT (orbit-rt.com) databases were also validated using the model validation methods detailed above.

Physician review

All cases in the validation set were fully anonymized and reviewed for clinical acceptability by radiation oncologists sub-specialized in the respective treatment site (see Table 4 for number of reviewing physicians per site). The physicians were made aware that all cases were KBP generated and were provided with a 5-point rubric (Table 2) for their ratings. Any feedback was noted. If feedback could be addressed through automatable means (i.e. feedback that could only be ameliorated using human discernment was not addressed), the process was restarted to mitigate their concerns with the model.

Table 2
Rubric Provided to Radiation Oncologists During Review to Rate Validation Plans.

Rating	Likert Scale	Description
5	Strongly agree	Use-as-is (i.e., clinically acceptable, and could be used for treatment without change).
4	Agree	Minor edits that are not necessary. Stylistic differences, but not clinically important. The current contours/plan are acceptable.
3	Neither agree or disagree	Minor edits that are necessary. Minor edits are those that can be made in less time than starting from scratch or are expected to have minimal effect on treatment outcome (e.g., beam weighting adjustment).
2	Disagree	Major edits. This category indicates that the necessary edits are required to ensure appropriate treatment and sufficiently significant that the user would prefer to start from scratch.
1	Strongly disagree	Unusable. This category indicates that the quality of the automatically generated contours or plan are so bad that they are unusable.

Results

Using R^2 to assess model quality, overall, 77 % of trained Model Structures had good fit ($R^2 \geq 0.7$) or modest fit ($0.7 > R^2 \geq 0.5$). Prostate (43 %) and rectum (50 %) had the lowest percentage of good or modest fit Model Structures, while the liver, lung/mediastinum, and bladder models had the highest with 100 %. More details are found in Table 1 of each model section in Appendix A. Through qualitative analysis in step one of validation, DVH estimation bands showed overall reasonable estimation band widths and similar line estimations comparable to their respective clinical doses, often despite poor fit ($R^2 < 5$, Fig. 1). Poor fits were primarily due to structures being variably distanced from targets. For example, the esophagus model included an array of tumors with varying distances from the heart ($R^2 = 0.38$). All seven cases that did not meet constraints for the heart (Appendix A, page 12, Table E2) were inside threshold values for the model, however only two clinical plans were able to meet constraints when the KBP plans were not (specifically, heart mean < 20 Gy). Of those two, only one showed that the DVH estimation itself was insufficient to meet the clinically achieved doses (actual KBP heart mean achieved $= 21.6$ Gy, clinical heart mean $= 19.62$ Gy; Fig. 1B).

The head and neck model from Washington University (included in Eclipse, Varian Medical Systems) and the liver SBRT model from ORBIT-RT DVH estimates were comparable to clinical doses in the first step of validation, and the models were successfully configured to meet local clinical standards through the methods of the second step of validation (Appendix A, pages 15–21). Review of the DVH estimation bands for the anorectal, esophagus, and lung/mediastinum models from ORBT-RT did not overall produce line estimates compatible with local clinical doses in the first step of validation. As a result, new models for these sites were developed with local clinical data.

All models were successfully designed to work with the Eclipse API to complete a fully optimized plan without any manual intervention or need for human discernment. All except for the bladder model required planning structures that were created solely using simple functions that can be automated through the programming interface (see Table 4 per model site in Appendix A for details). The esophagus and lung/mediastinum models required adding and training planning Model Structures from the most optimal cases in the model, such as normal tissue or avoidance structures (e.g., contralateral lung minus the planning target volume with margin). For example, the lung/mediastinum model consistently did not meet V5 and V10 (i.e., the percentage of the structure's normal tissue receiving ≥ 5 Gy or ≥ 10 Gy, respectively) constraints; however, after contouring a lung avoidance in 30 cases that most optimally met V5 and V10 and training it in the model, the plans were able to meet these constraints without compromising coverage (Fig. 2).

For models that did not routinely produce acceptable plan quality after the first KBP-based optimization, an approach was developed to add additional structures. These structures were then used for a second optimization, which was scored for acceptability. The new structures were specifically designed so that their generation could be automated using the Eclipse API, generally based on isodoses of the first optimization, so that the second optimization can be run without the need for user editing/intervention. Optimization objectives, including additional structures for subsequent optimizations, can be found in Table 5 per model in Appendix A.

Throughout the validation process, a wide variety of cases were validated with the model to assess which scenarios could work with each model. This created exclusions and inclusions per model (specifics can be found in the Overview section of each model in Appendix A). For an example of found exclusions: the lung/mediastinum model was created

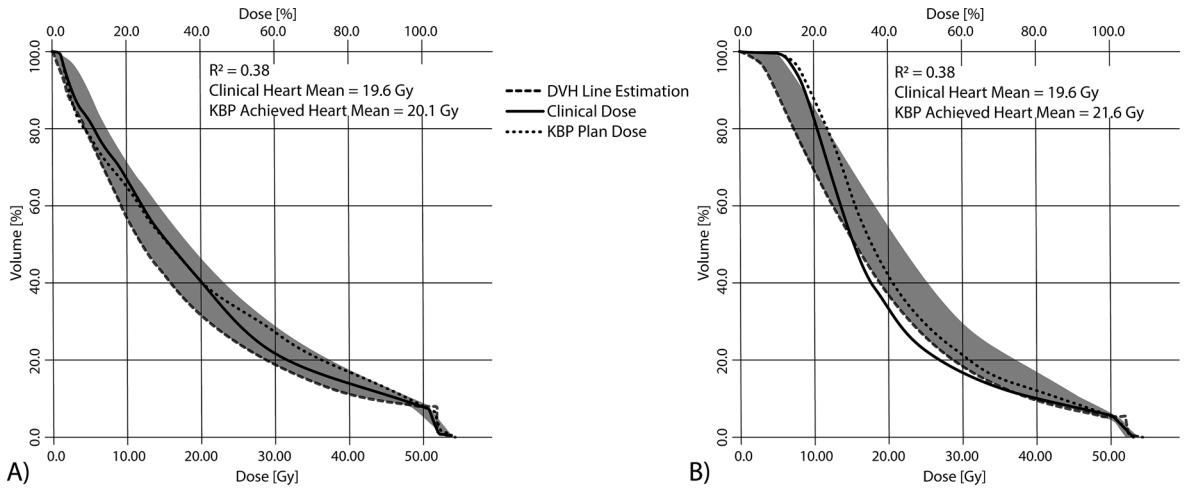

Fig. 1. A. A knowledge-based planning (KBP)-generated dose-volume histogram (DVH) estimation of heart band width (shaded area) and line estimation (dashed line) with reasonable comparison to the dose achieved in the clinic (solid line) despite a poor fit of $R^2 = 0.38$. **B.** A case in which the DVH estimation (dashed line and shaded area) would be insufficient in generating objectives to meet clinically achieved heart sparing between V15Gy and V35Gy.

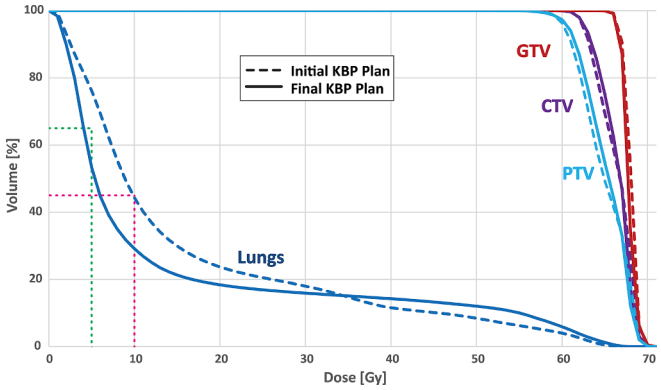

Fig. 2. Dose-volume histogram comparison of a knowledge-based planning (KBP) plan before and after using trained planning structures in the model. Target coverage is similar between the initial and final plans, but the dose to the lungs better meets V5 < 65 % (dotted green line) and V10 < 45 % constraints (dotted magenta line) in the final model. CTV, clinical target volume; GTV, gross tumor volume; PTV, planning target volume. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for multiple prescriptions, however, it was found during validation that the model would have to be more specific, and hypofractionated cases were excluded. For an example of found inclusions: the anorectal model was initially designed for supine setups, but upon further validation, it was proven to work for prone setups as well.

Dosimetric analysis showed that on average across all models, 86 % of KBP-generated plans met 90 % of OAR constraints (Table 3). Physician review evidenced that most often that cases did not meet OAR constraints were due to overlap or proximity of contours with targets and were still deemed clinically acceptable with this in mind. For example, only 38 % of anorectal cases passed bladder constraints, but 94 % of cases were deemed clinically acceptable as is (Table 3), with physician feedback noting the overlap. In addition, on average, 87 % ± 17 % of cases met all target objectives and constraints, and the volume of the prescription dose or all planning target volumes (PTVs) per model were met at 98 % ± 1 % (Table 3). This high average suggests that cases that did not meet their PTV objectives were likely borderline (six treatment sites had V100% > 95 % as their PTV objective). For example, the anorectal short course model had 71 % of cases meet the V100% > 95 % PTV objective; however, the average dose was 96 % ± 1 % (Table 3). Further details on planning objectives and validation results are found in Table 2 of each model section in Appendix A.

A model was determined to be robust if it had at least 90 % "acceptable as is" ratings by physicians, which were considered a score 4 or 5 on the Likert scale. The anorectal, bladder, cervical, lung/

Table 3
Dose-Volume Histogram Evaluation of the Knowledge-Based Planning–Generated Plans.

Model Site	No. of Cases	Cases that Met 90 % of OAR Constraints, %	Average Target Objective Pass Rate, %	Average Dose for PTVs, %	Average Dose for Plan D_{max}, %
Anorectal	34	94	87	95.5 ± 1.1	107.1 ± 0.5
Bladder	20	95	100	99.2 ± 0.4	105.7 ± 1
Lt/Rt breast and Rt chest wall	68	76	87	99.1 ± 0.9	111.2 ± 2
Lt chest wall*	25	44	94	98.8 ± 0.9	113.6 ± 2.7
Cervical†	40	85	59	99.9 ± 0.3	112.4 ± 4.5
Esophagus	20	90	100	97.5 ± 1.2	108.7 ± 1.3
Head and neck†	57	75	99	92.9 ± 3.7	107.3 ± 1.1
Liver SBRT	20	100	95	97.8 ± 1.7	–
Lung/mediastinum	25	92	97	97.7 ± 2.3	106.9 ± 1.2
Prostate	20	95	46	96.9 ± 0.6	105.7 ± 0.4
Prostate with nodes	20	95	94	98.6 ± 0.8	106.8 ± 1.4

*Separated left chest wall data from left and right breast and right chest wall cases due to significant difference in results.
†Data from international multi-institutional study [3,12].
D_{max}, maximal dose; Lt, left; OAR, organs at risk; PTV, planning target volume; Rt, right.

Table 4
Clinical Acceptability of Knowledge-Based Planning–Generated Plans Based on Physician Review.

Model Site	Number of Cases	Number of Physicians	Acceptable as is (4–5)*, %	Acceptable after Minor Edits (3 +)*
Anorectal	35	2	96	100
Bladder	20	2	100	100
Lt/Rt breast and Rt chest wall	68	1	88	97
Lt chest wall†	25	1	68	88
Cervical‡	75	3	99	100
Esophagus	20	1	75	100
Head and neck‡	75	3	87	96
Liver SBRT	20	1	80	100
Lung/ mediastinum	25	1	96	96
Prostate	38	3	83	100
Prostate with nodes	20	2	98	100

*Likert scale.
†Separated left chest wall data from left and right breast and right chest wall cases due to significant difference in results.
‡Data from international multi-institutional study [3,12].
Lt, left; Rt, right.

mediastinum, and prostate with nodes models were able to achieve this goal (Table 4). Left/right breast and right chest wall cases validated with the breast/chest wall model were able to achieve the goal of 95 % "acceptable after minor edits" or better ratings (3, 4, or 5 on the Likert scale). Several left chest wall cases validated with the breast/chest wall model faced a significant challenge in meeting the left ventricle V5 dose constraint, which led to low scores. This was likely due to the proximity of the chest wall to the left ventricle and lower volume of tissue compared to the left breast. The liver SBRT model validated several plans with significant chest wall involvement that scored below 4 due to proximity or overlap of chest wall structures; the model heavily prioritizes target coverage, which weighted the model from achieving the chest wall goals. Despite this, 100 % of the cases validated with the SBRT model were rated at least a 3 (acceptable after minor edits), with physician feedback requesting chest wall dose to be lowered if possible (Table 4). Multiple plans received varying feedback from reviewing physicians on whether they would accept a plan that went over constraints as tradeoff between target coverage and OAR sparing; these were determined to require physician discernment on a case-by-case basis.

Discussion

This study showed that it is possible to generate fully automatable plans that are efficient, consistent, high-quality, and clinically acceptable using Knowledge Based Planning (KBP) models for multiple treatment sites (Fig. 3). These plans do not require adjustment by the treatment planner and are thus well suited for inclusion in a fully automated treatment planning process. Plan review by a physician is still necessary, both due to medicolegal and ethical obligations as well as the fact that even the best models did not produce clinically acceptable plans for all patients. Importantly, there were instances of cases that did not meet OAR constraints but were still deemed clinically acceptable by physicians due to patient anatomy; conversely, there were cases that met constraints but were rated unacceptable or edits required due to physician preference of dose distribution. Valuable work on risk in automated radiotherapy has also highlighted the importance of human review of the output of automated processes in radiotherapy [12].

As KBP relies on geometry, this study demonstrated that a set of patient data in one setup can be used in multiple setups of similar geometric layout. For example, the supine anorectal model generated equal quality plans and clinical acceptability scores for prone setup patients, and similarly, the left breast model for right breast cases. There was no direct correlation of the number of patient cases in each model with plan quality or clinical acceptability, so it is possible to generate a quality model with a limited patient pool if the plans are of consistent quality and planning structures are used.

Cases that were rated not clinically acceptable often received such rating due to not meeting standard OAR constraints and differences in the reviewing physician's goals, where the reviewing physician would prefer a tradeoff of target coverage (highly prioritized in the model) to meet an OAR constraint. Consistent with findings from other studies [6,13], there was notable inter-physician variability in ratings. Inter-user variability in clinical acceptability [6] mean that different users may have different opinions about the same plan. Our 5-point scale attempts to include this, with a score of 5 meaning acceptable, and 4 meaning that it is acceptable, but some refinement might be helpful if time or resources allow. In general, a 4 or 5 denote an acceptable plan. This will be further clarified in future prospective use of these tools. Whether a planner will/should further adjust the plan is a complicated question that is affected by institutional planning practice as well as the time and effort needed to make any changes. For a few examples experienced in this study, the prostate model was reviewed by three physicians, and while there were some variability in ratings between all three physicians for each case, the largest difference came from one physician who rated six plans as "acceptable with minor edits" that the others rated "acceptable as is." Feedback noted that clinical preference differed, as for one, the prescription isodose line is preferred to be within the target boundaries and they requested that the plans be normalized to reflect this, and for the others, clinical preference is for the target to be covered by the prescription isodose line with margin. For the anorectal model, one physician rated 100 % of cases as "acceptable as is," and the other rated 9 % as needing minor edits due to differences in preference for dose distribution and target dose conformality.

There were also instances of cases that failed due to unclear reasons, such as one lung/mediastinum case rated unacceptable (2) that did not meet half the planning objectives despite having similar aspects of other plans that performed well; additions to the model of cases with comparable characteristics did not increase the model's performance for that plan. With these exceptions in mind, on average, 88 % of all model cases were rated "acceptable as is" and 98 % "acceptable after minor edits." The remaining percentage of cases will likely require manual planning, which is consistent with previous studies' findings [3]. This provides a potential to significantly reduce treatment planning workload by offloading more straightforward cases, allowing for more time and efforts to focus on the difficult cases.

McIntosh et al. [9] and Olanrewaju et al. [3] showed comparable or higher acceptability ratings during blind physician review between clinical and auto-planned cases in concept. However, McIntosh et al. went on to show that, while clinical acceptability ratings remained similar following actual deployment, the number of auto-planned cases actually accepted for treatment was significantly lower. This finding shows that these acceptance ratings may not be fully reflective of actual plan acceptability when used for clinical treatment of patients.

The head and neck and cervical models are currently integrated into the Radiation Planning Assistant (RPA.MDAnderson.org), a fully automated contouring and planning system that is FDA 510(k) approved, not marketed in the USA, but is clinically deployed in cancer centers in South Africa. The remaining models are currently being integrated into the next version of the system, and are part of an ARCHERY protocol to evaluate AI-based contouring and planning [14].

Our study had a few limitations to note. With the exception of the head and neck, cervical, and prostate models, all reviews were from a single institution—and often from a single physician—and therefore do not reflect the breadth and diversity of clinical practices. In addition, except for the head and neck and cervical models, all model and

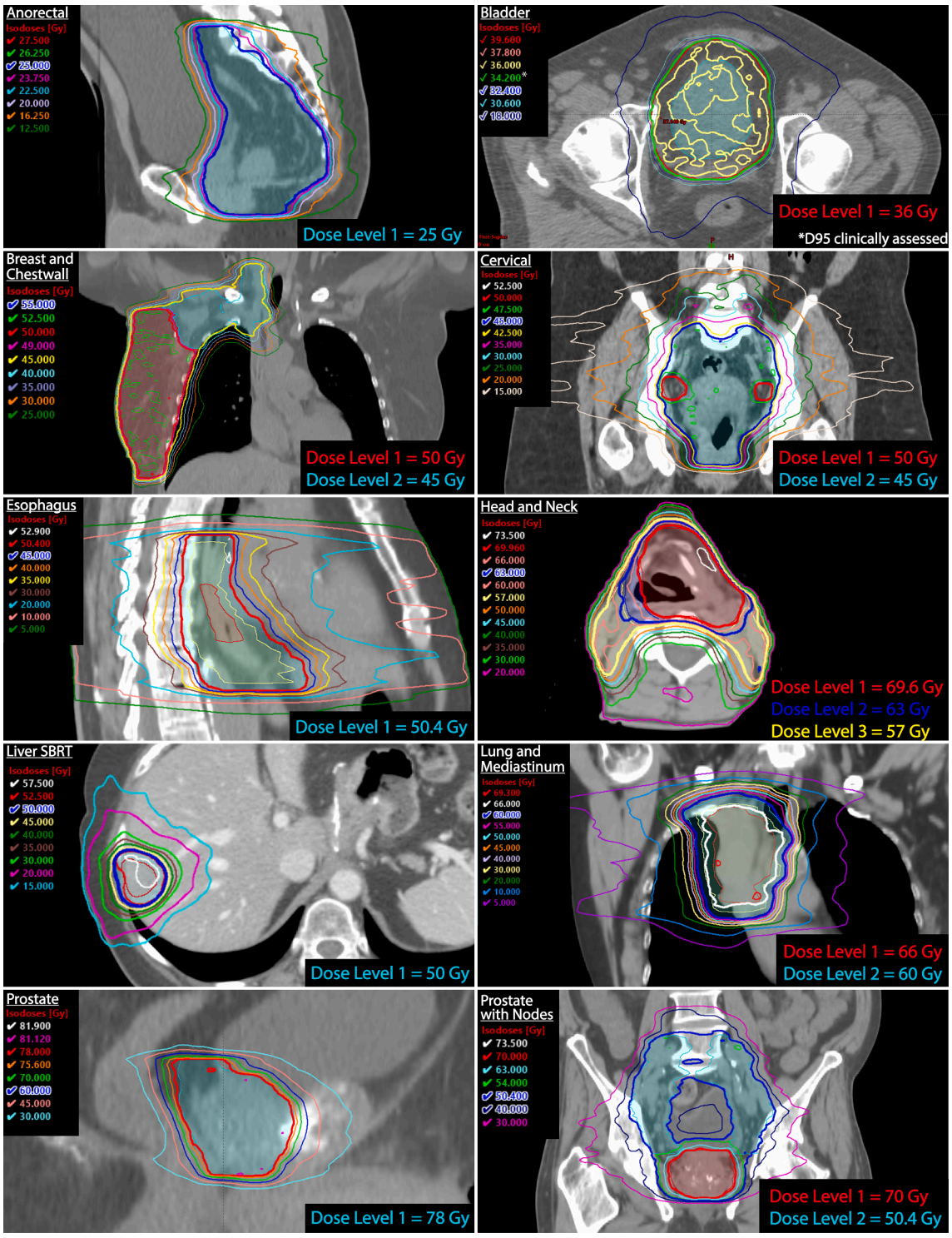

Fig. 3. Example plan of each of the ten fully automatable and clinically acceptable knowledge-based planning models – CT scans with corresponding dose distributions.

validation set cases were collected from a single institution, and results may not be consistent for other populations and clinics.

Conclusion

This study showed that it is possible to generate fully automatable plans using Knowledge Based Planning models as the basis for multiple treatment sites, with 88 % of all model-generated cases being rated as "acceptable as is" on physician review. As these plans do not require human intervention, this work has the potential to significantly impact treatment planning workflows.

CRediT authorship contribution statement

Christine V. Chung: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation. **Meena S. Khan:** Writing – review & editing, Validation, Methodology, Investigation, Data curation.

Adenike Olanrewaju: Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Mary Pham:** Writing – review & editing, Validation, Methodology, Investigation, Data curation. **Quyen T. Nguyen:** Writing – review & editing, Validation, Methodology, Investigation, Data curation. **Tina Patel:** Writing – review & editing, Validation, Methodology, Investigation, Data curation. **Prajnan Das:** Writing – review & editing, Validation. **Michael S. O'Reilly:** Writing – review & editing, Validation. **Valerie K. Reed:** Writing – review & editing, Validation. **Anuja Jhingran:** Writing – review & editing, Validation. **Hannah Simonds:** Writing – review & editing, Validation, Resources. **Ethan B. Ludmir:** Writing – review & editing, Validation. **Karen E. Hoffman:** Writing – review & editing, Validation. **Komeela Naidoo:** Writing – review & editing, Validation, Resources. **Jeannette Parkes:** Writing – review & editing, Validation, Resources. **Ajay Aggarwal:** Writing – review & editing, Validation, Resources. **Lauren L. Mayo:** Writing – review & editing, Validation. **Shalin J. Shah:** Writing – review & editing, Validation. **Chad Tang:** Writing – review & editing, Validation. **Beth M. Beadle:** Writing – review & editing, Validation, Resources. **Julie Wetter:** Writing – review & editing, Validation. **Gary Walker:** Writing – review & editing, Validation. **Simon Hughes:** Writing – review & editing, Validation. **Vinod Mullassery:** Writing – review & editing, Validation. **Stephen Skett:** Writing – review & editing, Software, Resources, Data curation. **Christopher Thomas:** Writing – review & editing, Resources, Formal analysis. **Lifei Zhang:** Writing – review & editing, Software, Resources. **Son Nguyen:** Writing – review & editing, Resources. **Raymond P. Mumme:** Writing – review & editing, Software, Resources, Data curation. **Raphael J. Douglas:** Writing – review & editing, Software, Resources, Data curation. **Hana Baroudi:** Writing – review & editing, Data curation. **Laurence E. Court:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

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

Editorial support was provided by Bryan Tutt, Scientific Editor, Research Medical Library at MD Anderson Cancer Center. Data collection and processing support was provided by Jean R. Guma-De La Vega, Research Assistant I, Radiation Physics – Research at MD Anderson Cancer Center. This process automation was developed with support from the NCI, Varian, Wellcome Trust, and MD Anderson Cancer Center.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2024.110609.

References

[1] Rhee DJ, Jhingran A, Kisling K, Cardenas C, Simonds H, Court L. Automated radiation treatment planning for cervical cancer. Seminars in Radiation Oncology 2020;30:340–7. https://doi.org/10.1016/j.semradonc.2020.05.006.
[2] Scaggion A, Fusella M, Roggio A, Bacco S, Pivato N, Rossato MA, et al. Reducing inter- and intra-planner variability in radiotherapy plan output with a commercial knowledge-based planning solution. Physica Medica 2018;53:86–93. https://doi.org/10.1016/j.ejmp.2018.08.016.
[3] Olanrewaju A, Court LE, Zhang L, Naidoo K, Burger H, Dalvie S, et al. Clinical Acceptability of Automated Radiation Treatment Planning for Head and Neck Cancer Using the Radiation Planning Assistant, Pract. Radiation Oncology 2021;11:177–84. https://doi.org/10.1016/j.prro.2020.12.003.
[4] Fogliata A, Reggiori G, Stravato A, Lobefalo F, Franzese C, Franceschini D, et al. RapidPlan head and neck model: the objectives and possible clinical benefit. Radiation Oncology 2017;12:73. https://doi.org/10.1186/s13014-017-0808-x.
[5] Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems, Pract. Radiation Oncology 2012;2:296–305. https://doi.org/10.1016/j.prro.2011.11.012.
[6] Baroudi H, Brock KK, Cao W, Chen X, Chung C, Court LE, et al. Automated Contouring and Planning in Radiation Therapy: What Is 'Clinically Acceptable'? Diagnostics 2023;13:667. https://doi.org/10.3390/diagnostics13040667.
[7] Schubert C, Waletzko O, Weiss C, Voelzke D, Toperim S, Roeser A, et al. Intercenter validation of a knowledge based model for automated planning of volumetric modulated arc therapy for prostate cancer. The experience of the German RapidPlan Consortium, PLOS ONE 2017;12:e0178034.
[8] Li N, Carmona R, Sirak I, Kasaova L, Followill D, Michalski J, et al. Highly Efficient Training, Refinement, and Validation of a Knowledge-based Planning Quality-Control System for Radiation Therapy Clinical Trials. Int J Radiat Oncol 2017;97:164–72. https://doi.org/10.1016/j.ijrobp.2016.10.005.
[9] McIntosh C, Conroy L, Tjong MC, Craig T, Bayley A, Catton C, et al. Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. Nature Medicine 2021;27:999–1005. https://doi.org/10.1038/s41591-021-01359-w.
[10] Ziemer BP, Shiraishi S, Hattangadi-Gluth JA, Sanghvi P, Moore KL. Fully automated, comprehensive knowledge-based planning for stereotactic radiosurgery: Preclinical validation through blinded physician review, Pract. Radiation Oncology 2017;7:e569–78. https://doi.org/10.1016/j.prro.2017.04.011.
[11] Varian Medical Systems, RapidPlan 15.5 Implementation, (2018).
[12] Nealon KA, Balter PA, Douglas RJ, Fullen DK, Nitsch PL, Olanrewaju AM, et al. Using Failure Mode and Effects Analysis to Evaluate Risk in the Clinical Adoption of Automated Contouring and Treatment Planning Tools, Pract. Radiation Oncology 2022;12:e344–53. https://doi.org/10.1016/j.prro.2022.01.003.
[13] Gronberg MP, Beadle BM, Garden AS, Skinner H, Gay S, Netherton T, et al. Deep Learning-Based Dose Prediction for Automated, Individualized Quality Assurance of Head and Neck Radiation Therapy Plans, Pract. Radiation Oncology 2023;13:e282–91. https://doi.org/10.1016/j.prro.2022.12.003.
[14] Aggarwal A, Court LE, Hoskin P, Jacques I, Kroiss M, Laskar S, et al. ARCHERY: a prospective observational study of artificial intelligence-based radiotherapy treatment planning for cervical, head and neck and prostate cancer – study protocol. BMJ Open 2023;13:e077253.