Automated robust planning for IMPT in oropharyngeal cancer patients using machine learning

Bruggen, I.G. van; Huiskes, M.; Vette, S.P.M. de; Holmström, M.; Langendijk, J.A.; Both, S.; ... ; Korevaar, E.W.

INTERNATIONAL JOURNAL OF
RADIATION ONCOLOGY • BIOLOGY • PHYSICS
www.redjournal.org

PHYSICS CONTRIBUTION

Automated Robust Planning for IMPT in Oropharyngeal Cancer Patients Using Machine Learning

Ilse G. van Bruggen, MSc,* Merle Huiskes, MSc,* Suzanne P.M. de Vette, MSc,* Mats Holmström, MSc,† Johannes A. Langendijk, MD, PhD,* Stefan Both, PhD,* Roel G.J. Kierkels, PhD,‡ and Erik W. Korevaar, PhD*

*Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands; †RaySearch Laboratories, Stockholm, Sweden; and ‡Radiotherapiegroep, Deventer, the Netherlands

Received Apr 15, 2022; Accepted for publication Dec 5, 2022

Purpose: The aim of this study was to evaluate an automated treatment planning method for robustly optimized intensity modulated proton therapy (IMPT) plans for oropharyngeal carcinoma patients and to compare the results with manually optimized robust IMPT plans.

Methods and Materials: An atlas regression forest-based machine learning (ML) model for dose prediction was trained on CT scans, contours, and dose distributions of robust IMPT plans of 88 oropharyngeal cancer (OPC) patients. The ML model was combined with a robust voxel and dose volume histogram-based dose mimicking optimization algorithm for 21 perturbed scenarios to generate a machine-deliverable plan from the predicted dose distribution. Machine learning optimization (MLO) configuration was performed using a cross-validation approach with 3 × 8 tuning patients and comprised of adjustments to the mimicking optimization, to generate higher-quality MLO plans. Independent testing of the MLO algorithm was performed with another 25 patients. Plan quality of clinical and MLO plans was evaluated by the clinical target volume (D98% voxel-wise minimum dose >94%), organ at risk (OAR) doses, and the normal tissue complication probability (NTCP) (sum (Σ) of grade-2 and grade-3 dysphagia and xerostomia).

Results: Adequate robust target coverage was achieved in 24/25 clinical plans and in 23/25 MLO plans in the primary clinical target volume (CTV). In the elective CTV, 22/25 clinical plans and 24/25 MLO plans passed the robust target coverage evaluation threshold. The MLO average Σgrade 2 and Σgrade 3 NTCPs were comparable to the clinical plans (Σgrade 2 NTCPs: clinical 47.5% vs MLO 48.4%, Σgrade 3 NTCPs: clinical 11.9% vs MLO 12.3%). Significant increases in OAR average doses in MLO plans were found in the pharynx constrictor muscles (163 cGy, $P = .002$) and cervical esophagus (265 cGy, $P = .002$). The MLO plans were created within 45 minutes.

Conclusion: This study showed that automated MLO planning can generate robustly optimized MLO plans with quality comparable to clinical plans in OPC patients. © 2022 Published by Elsevier Inc.

Corresponding author: Ilse G. van Bruggen, MSc; E-mail: i.g.van.bruggen@umcg.nl

Ilse G. van Bruggen and Merle Huiskes contributed equally to this work.

M. Huiskes is currently at the Department of Radiation Oncology, Leiden University Medical Centre, Leiden, the Netherlands.

Disclosure: The department has a research collaboration with RaySearch Laboratories.

Research data are stored in an institutional repository and will be shared upon request to the corresponding author.

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ijrobp.2022.12.004.

Int J Radiation Oncol Biol Phys, Vol. 000, No. 00, pp. 1–8, 2022
0360-3016/$ - see front matter © 2022 Published by Elsevier Inc.
https://doi.org/10.1016/j.ijrobp.2022.12.004

Introduction

In the last decades, radiation therapy treatment techniques have transitioned from 3-dimensional conformal radiation therapy to intensity modulated radiation therapy with photons and protons. Although intensity modulated proton therapy (IMPT) is becoming more common worldwide, it is still quite limited in availability. Therefore, an efficient and cost-effective decision-support system to select patients for proton therapy is needed.

Following the Dutch model-based approach, normal tissue complication probability (NTCP) values derived from both photon and proton dose distributions and corresponding complication probability thresholds have been introduced to select patients for proton therapy.[1] Multiple studies showed methods to make the selection procedure more efficient.[2-4] Cheng et al described the implementation of a decision-support prototype to automatically select patients for photon or proton therapy.[2] However, their prototype requires the existence of the actual photon and proton treatment plans. Delaney et al investigated a knowledge-based planning tool to automatically generate IMPT treatment plans.[3] Their proton plans were optimized and evaluated on the planning target volume (PTV). However, the PTV concept has limitations, especially in IMPT. The PTV concept assumes that patients' anatomy changes will not affect the dose distribution. However, range errors should be taken into account because, for example, a range undershoot could lead to cold spots in the clinical target volume (CTV). Robust treatment planning strategies have been proposed as a mathematical approach to account for patient setup and proton range uncertainties.[5] An algorithm to automatically achieve a robust optimized proton plan was evaluated by Kierkels et al, their algorithm required a photon plan dose distribution for proton plan optimization.[4] The authors noted that their algorithm does not seek to improve on reference photon dose even though it may be possible with IMPT.

Ideally, the decision-support system should include robustly optimized automated proton therapy planning using the full potential of IMPT. Therefore, the purpose of this study was to develop and evaluate automated robust IMPT planning with treatment plan quality comparable to the currently used clinical IMPT plans for oropharyngeal cancer (OPC) patients. Hence, the automated IMPT planning algorithm was developed for actual proton therapy delivery and embedded with the decision-support system for patient selection. To automatically create robust IMPT plans, we combined a machine learning driven dose prediction algorithm previously introduced by McIntosh et al with a robust mimicking optimization algorithm.[4,6]

Methods and Materials

Study population

We included 113 consecutive OPC patients with varying local tumor extensions and regional lymph node statuses (Table 1). All patients were previously treated with curative intent using volumetric modulated arc therapy (VMAT) or spot scanning IMPT and a prescribed dose of 7000 cGy (relative biological effectiveness [RBE] of 1.1 for proton plans) to the primary tumor (CTVprimary) and 5425 cGy (RBE) to the bilateral elective lymph node regions (CTVelective) in 35 fractions (7 weeks). Each patient had both a VMAT and an IMPT plan, but the actual treatment modality decision was based on the outcome of the model-based selection procedure.[1]

Table 1 Overview of patient characteristics for the training, validation, and test set.

	Training and validation patients (n = 88)	Test patients (n = 25)
Sex (n)		
Female	8	6
Male	80	19
Age (average [SD])	62 (9)	64 (8)
TNM stage (range)		
T	1-4	1-4
N	0-5	0-5
M	0	0
P16 (n)		
Positive	41	10
Negative	47	15
Baseline score xerostomia (n)		
Grade 0	57	23
Grade 1	31	2
Baseline score dysphagia (n)		
Grade 0	34	15
Grade 1	24	5
Grade 2+	30	5
Clinical treatment technique (n)		
IMPT	61	23
VMAT	27	2

Abbreviations: IMPT = intensity modulated proton therapy, VMAT = volumetric modulated arc therapy.

IMPT planning and evaluation

The clinical IMPT plans were robustly optimized (21 scenarios) for the CTV with a 3 mm (n = 75 patients) or 5 mm (n = 39 patients) isotropic setup uncertainty (the clinical

protocol changed from 5-3 mm[7]) and 3% density uncertainty in RayStation 9A (RaySearch Laboratories AB, Stockholm, Sweden).[8] All plans consisted of 6 beams with 4 coplanar beam directions (anterior and posterior oblique). Four beams had a range shifter of 4 cm and, in the 2 posterior oblique directions, beams were added without a range shifter. The computational dose grid was $3 \times 3 \times 3$ mm^3 and the statistical uncertainty of the Monte Carlo dose computation was 1.0%. The IMPT plans were optimized for the IBA ProteusPlus proton gantry (Ion Beam Applications SA, Louvain-la-Neuve, Belgium), with energies between 70 and 230 MeV with corresponding spot sizes in air of 6.5 and 3.0 mm (one sigma), respectively.

Moreover, segmentations of the spinal cord, brain, optic chiasm, brain stem, salivary glands, oral cavity, pharyngeal constrictor muscles (PCMs), thyroid, and buccal mucosa were used for plan optimization and evaluation.[1] Robustness of CTV coverage was assessed from voxel-wise minimum (voxmin) and maximum (voxmax) dose distributions derived from 28 perturbed dose distributions following the procedure described by Korevaar et al.[9] Robust optimization was performed with 21 scenarios as a compromise between speed and accuracy. The robustness evaluation only needed to be performed once; therefore, we used a higher number of scenarios in the evaluation than in the optimization. Robust target coverage was achieved when CTV D98%≥94% in the voxmin dose distribution. This target coverage criterion were used because the study of Korevaar et al showed strong correlation with the D98%>95% in PTV planning.

Machine learning optimization

The machine learning optimization (MLO) method combined a per-voxel dose prediction model with a worst-case robust mimicking optimization algorithm to automatically create treatment plans for novel patients. An atlas regression forest model was trained with radiomic features of patient computed tomography (CT) images, features of delineated targets and organs, and the corresponding clinical proton dose distribution. The dose prediction of the MLO method is described in detail by McIntosh et al.[6,10,11] In this study we extended MLO with robust dose mimicking optimization for IMPT plans and extensive model configuration.

The mimic optimization problem was formulated as a voxel-wise least squared error (LSE) minimization problem. Voxel weights were determined in an automatic fashion based on the geometric position of the voxel in relation to the regions of interest (ROIs), as well as by the assigned dose value of the voxel in predicted dose. The LSE objective function was divided into 2 one-sided terms (min and max objective), enabling different weights depending on whether the optimized dose was greater or less than the predicted dose. The optimization strove to obtain the nominal dose volume histogram (DVH) of robust ROIs (eg, CTVs) for each robust scenario.

The dose mimicking optimization uses the same beam arrangement, robustness settings, and perturbed scenarios as in the clinical plans. The CTVs were optimized robustly and a total of 275 optimization iterations for each plan were used over 5 runs (60, 60, 60, 45, and 45 iterations, respectively). After each run, the mimicked dose was compared with the predicted dose and objective weights were adjusted accordingly.

All clinical and MLO plans were generated using an Intel Xeon CPU E5-2698 v4 @ 2.20GHz processor. This method was implemented in a nonclinical version of RayStation (version 10B Development).

MLO configuration

Eighty-eight treatment plans out of 113 OPC plans were randomly selected for training and MLO configuration. The training and configuration were performed following a 3-fold leave-8-out cross validation approach; the models were trained with 80 plans, excluding different sets of 8 patients in each model. With this approach the model configuration could be done on 24 plans. The MLO configuration was performed iteratively until an acceptable strategy for MLO planning was achieved on the 24 validation plans. The remaining 25 patients were used as an independent test set.

The MLO configuration was done by so-called "predict and mimic settings." The predict settings are a set of ROI-specific functions that were configured to a set of DVH-specific dose constraints. These constraints were then used to update the spatial dose distribution (when necessary) to ensure that the predicted dose better resembles the desired clinical dose. In the mimic settings, a weights list, ROI-goals (optional), and number of iterations and optimization runs was set. In the list of weights the initial ROI-specific weights were set. Here, the ROI's robustness status was also configured. By using the ROI goals the voxel weights were modified dynamically, between iterations, based on different criteria that are stipulated as functions of the predicted and current optimized dose. A list of MLO configuration settings can be found in the supplementary material (Table A).

Plan evaluation

Target coverage robustness of the MLO plans was assessed using the multiscenario robustness evaluation as was used for the clinical IMPT plans, with the same scenarios. In addition, for each plan, target robustness was assessed on a repeat CT (rCT) to evaluate robustness of clinical and MLO plans during the treatment course. The rCT of the last week of treatment (potentially largest geometric variation compared with the planning CT) was selected for patients without plan adaptations during their actual treatment. For patients with a plan adaptation during the fractionated treatment course, the corresponding rCT was selected for evaluation. The setup uncertainty for the rCT robustness evaluation was set to 1 mm isotropically, accounting for

imaging versus treatment isocenter accuracy, in accordance with the clinical protocol.[7]

Additionally, the target conformity index (CI) was compared between MLO and clinical plans. The CI is defined as $CTV_{94\%voxmin}/V_{94\%voxmin}$, where $CTV_{94\%voxmin}$ is the volume of the target in the voxmin dose distribution that receives 94% of the prescribed dose, and $V_{94\%voxmin}$ is the total volume in the voxmin dose distribution that receives 94% of the prescribed dose.

The OARs were dosimetrically evaluated and the NTCP values for the sum(Σ) of grade 2 and grade 3 xerostomia and dysphagia 6 months postradiation therapy were estimated based on previously published NTCP models.[1] Finally, the model-based plan comparison was performed and evaluated for the MLO and clinical plans using the clinical VMAT plan.

Statistical analysis

A Bonferroni corrected Wilcoxon signed rank test was performed in SPSS Statistics 23 (IBM SPSS, Inc, Armonk, New York) to assess whether the MLO IMPT dosimetric parameters differed statistically significantly from the clinically IMPT plans with $P = .007$ ($\alpha = 0.05/7$ parameters) for target coverage, $P = .006$ ($\alpha = 0.05/9$ structures) for OARs and $P = .025$ ($\alpha = 0.05/2$) for grade 2 and grade 3 NTCPs (Table 2).

Results

The MLO automatically generated deliverable IMPT plans within 45 minutes per patient. An example of a dose

Table 2 Comparison of the target parameters, OAR doses, and NTCP values for plans of the 25 test patients

	Clinical plans (avg (min-max))	MLO plans (avg (min-max))	Difference (Clinical minus MLO)	*P* values
CTVprimary				
D98%>65.8 voxmin (Gy RBE)	66.2 (61.6−67.4)	66.2 (65.7−66.7)	−0.1	.201
CI voxmin (%)	66.9 (39.7−80.7)	67.3 (49.0−81.1)	0	.897
D0.1cc <79.8 voxmax (Gy RBE)	77.1 (74.9−79.2)	77.4 (75.4−80.4)	−0.4	.135
D98%>65.8 voxmin rCT (Gy RBE)*	65.2 (55.4−67.7)	64.9 (53.2−67.2)	0.5	.070
CTVelective				
D98%>51.0 voxmin (Gy RBE)	**50.9 (41.2−51.8)**	**51.8 (50.9−52.6)**	**−1.6**	**<.001**†
CI voxmin (%)	**61.4 (48.6−72.6)**	**57.1 (47.5−67.2)**	**4.3**	**<.001**†
D98%>51.0 voxmin rCT (Gy RBE)†	**51.0 (42.2−52.7)**	**51.4 (45.5−53.0)**	**−0.8**	**.001**†
OARs				
D0.1 brain (Gy RBE)	16.4 (0−38.3)	18.0 (0.2−34.9)	−1.6	.619
D0.1 brain stem (Gy RBE)	14.9 (2.1−39.7)	15.6 (0.2−37.0)	−0.7	.231
D0.1 spinal cord (Gy RBE)	33.4 (19.9−43.6)	33.6 (18.9−40.7)	−0.2	.968
Dmean parotid glands (Gy RBE)	16.7 (1.7−37.7)	16.7 (2.2−37.8)	0	.989
Dmean submandibular glands (Gy RBE)	49.6 (18.6−70.0)	49.4 (15.0−69.3)	0.2	.638
Dmean oral cavity (Gy RBE)	28.9 (9.5−49.1)	29.1 (9.2−49.2)	−0.2	.382
Dmean PCMs (Gy RBE)	**35.5 (5.2−71.4)**	**37.1 (5.4−70.0)**	**−1.6**	**.002**†
Dmean esophagus (Gy RBE)	**14.7 (4.6−52.0)**	**17.3 (4.0−55.0)**	**−2.7**	**.002**†
Dmean thyroid (Gy RBE)	47.5 (39.1−55.4)	48.7 (41.0−55.2)	−1.2	.183
NTCPs				
Σ xerostomia + dysphagia (grade ≥2) (%)	47.3 (31.1−70.8)	48.2 (31.4, 73.8)	−0.9	.116
Σ xerostomia + dysphagia (grade ≥3) (%)	11.8 (6.43−23.7)	12.1 (6.45, 22.7)	−0.3	.059

Min: minimum and max: maximum values are notes between brackets.

Abbreviations: CI = conformity index, CTV = clinical target volume, NTCP = normal tissue complication probability, OARs = organs at risk, PCMs = pharyngeal constrictor muscles, RBE = relative biological effectiveness, rCT = repeat computed tomography, Voxmax = voxel-wise maximum dose distribution, Voxmin = voxel-wise minimum dose distribution.

* All results were measured on the planning computed tomography (CT) scan, except for the D98% in the voxmin plan of the repeat CT.

† Significant differences (marked bold) with $P < .007$ for target coverage, $P < .025$ for NTCP and $P < .006$ for OARs.

distribution is shown in Figure 1. The MLO plans were comparable to the clinical IMPT plans in terms of target coverage robustness and OAR dose (Table 2) and similar DVHs were observed (Fig. 2).

Target coverage robustness for CTVprimary was achieved in 24/25 and 23/25 clinical and MLO plans, and the average D98% was comparable (Clinical: 66.2 Gy RBE, MLO: 66.2 Gy RBE, $P = .201$; Table 2). For CTVelective, 22/25 clinical plans and 24/25 MLO plans passed the target coverage robustness evaluation threshold. The average D98% was significantly higher in MLO plans (Clinical: 50.9 Gy RBE, MLO: 51.8 Gy RBE, $P < .001$). Furthermore, the CI for CTVelective was significantly lower in MLO plans (5 percentage points, $P < .001$).

The max doses in the brain, brain stem, and spinal cord were comparable between clinical and MLO plans (Table 2). The average dose to the superior PCM and the esophagus was significantly lower in the clinical plans (PCM superior: −1.6 Gy RBE $P = .002$, esophagus: −2.7 Gy RBE $P = .002$).

On average, the NTCPs were comparable between the clinical and MLO plans (Σgrade 2 NTCPs: clinical 47.5% vs MLO 48.4%, Σgrade 3 NTCPs: clinical 11.9% vs MLO 12.3%). Of 25 patients, 22 clinical plans and 20 MLO plans qualified for proton therapy.

The CTVprimary was adequately robust on the rCT in 14/25 clinical plans and 12/25 MLO plans. The evaluation on the rCT also showed that the average CTVprimary D98%voxmin was comparable between clinical plans and MLO plans (Clinical: 65.2 Gy RBE, MLO: 64.9 Gy RBE, $P = .070$; Fig. 3). For CTVelective, 18/25 clinical plans and 20/25 MLO plans passed the robustness evaluation threshold. On average, D98%voxmin was higher in MLO plans for CTVelective (Clinical: 51.0 Gy RBE, MLO: 51.4 Gy RBE, $P = .001$; Fig. 4).

Discussion

We evaluated an automated robust IMPT planning MLO algorithm for OPC patients. This MLO automatically generated IMPT plans with on average similar plan quality in terms of target coverage and NTCP values as the corresponding clinical plans. The MLO plans can be further optimized by tuning toward lower D98% in the voxmin of CTVelective, higher CI in CTVelective, and lower OAR doses. The MLO planning shows potential to generate robustly optimized proton plans within 45 minutes.

Automated robust IMPT planning has been presented previously.[3,4,12] Delaney et al performed a preliminary test to generate automated robustly optimized IMPT plans for 3 cases.[3] All 3 cases met their robustness criteria and plan quality was largely comparable to that of the clinical plans. Taasti et al showed that their fully automated treatment planning framework generated robust proton treatment plans that fulfilled all the hard maximum/mean/dose-

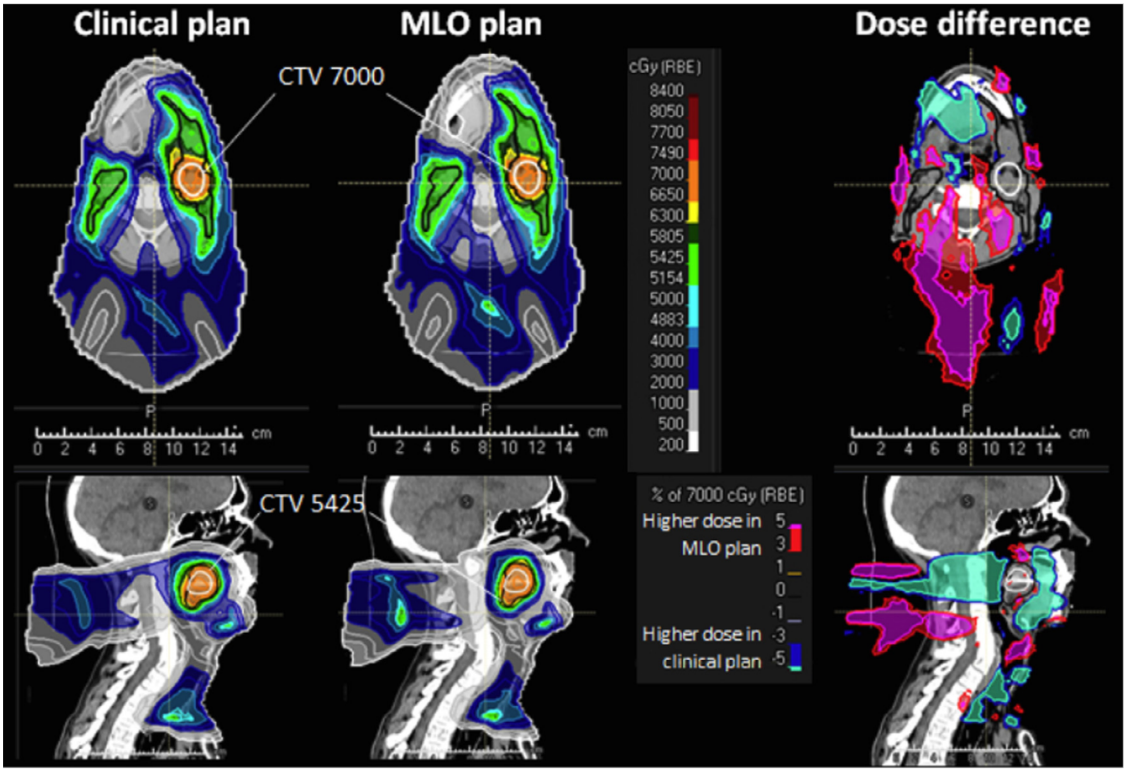

Fig. 1. A transversal (top row) and sagittal (bottom row) cross-section of a computed tomography scan of a representative case with the nominal dose distribution of the clinical intensity modulated proton therapy plan (left), the machine learning optimized plan (middle), and the dose difference (clinical minus machine learning optimized) (right).

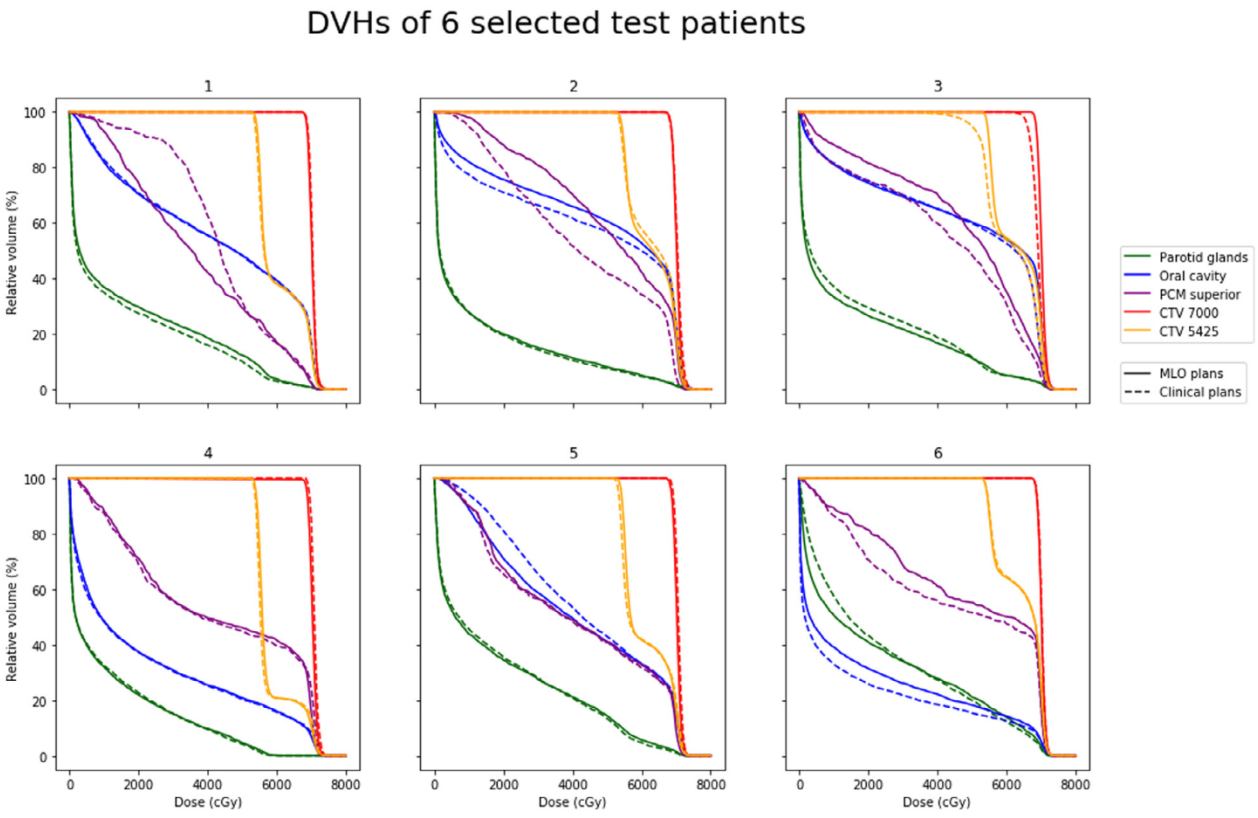

Fig. 2. Dose volume histograms (DVHs) of 6 selected test patients for the nominal dose distribution in the clinical plan and the machine learning optimized (MLO) plan. The DVHs of plot 1 correspond with the patient from Figure 1. The DVHs of plot 4 shows comparable DVHs between the clinical and the MLO plan. Similar DVHs between clinical and MLO plans were also seen in the remaining 19 test patients. Plots 2, 3, 5, and 6 show the largest differences between the clinical and MLO plans. Note that the ipsilateral and contralateral parotid glands were merged.

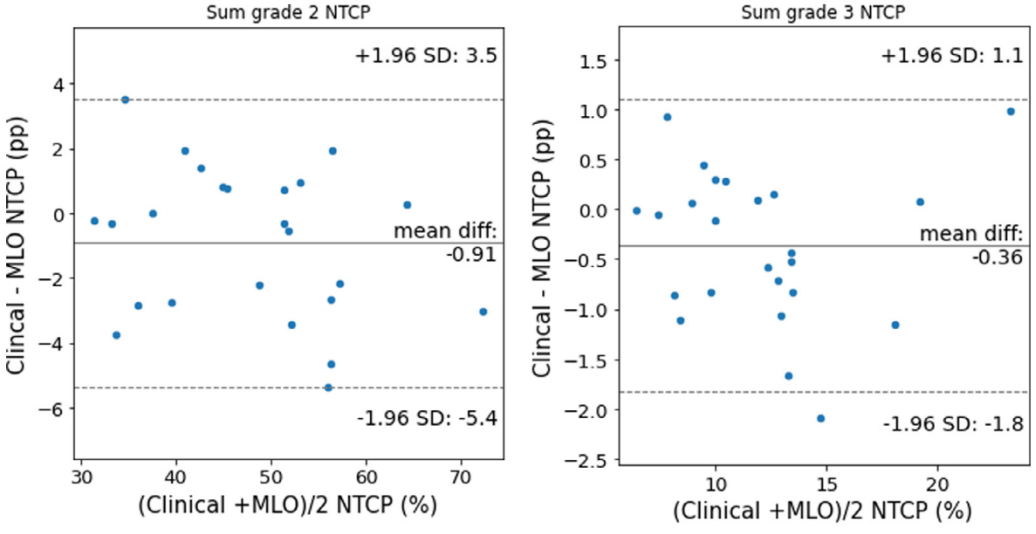

Fig. 3. Bland Altman plots of the sum grade 2 and 3 normal tissue complication probability (NTCP) values of 24 patients. (One patient had the submandibular glands removed; therefore, the xerostomia NTCP was not calculated). The x-axis represents the average sum NTCP of the clinical and machine learning optimized (MLO) plan. The y-axis represents the difference between clinical and MLO plan for the sum NTCPs. The solid lines indicate the mean difference of the NTCP of the clinical plan and the MLO plan. The dashed lines indicate the mean $\pm 1.96^{*}$ SD.

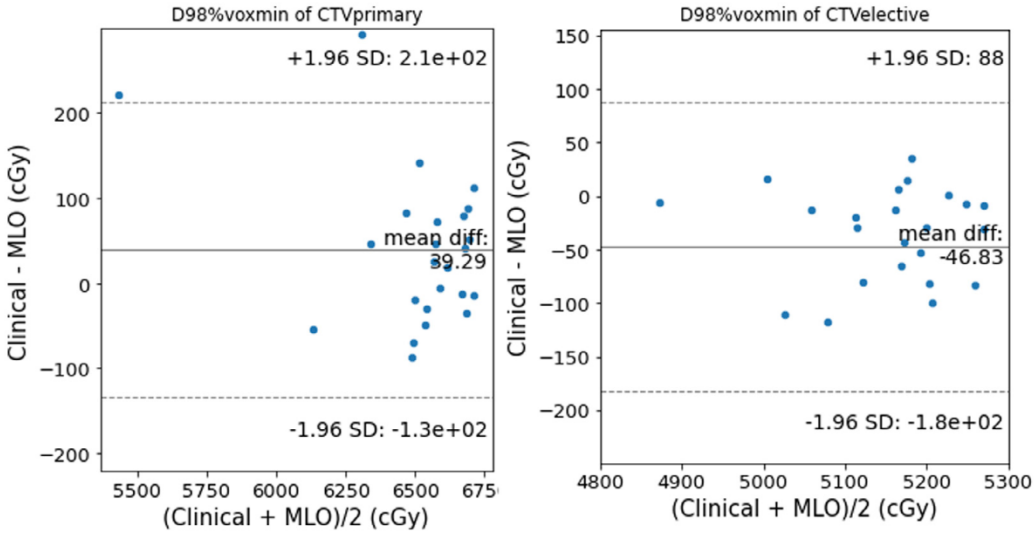

Fig. 4. Bland Altman plot of D98% of the voxel-wise minimum (D98%vm) dose distribution evaluated on the repeat computed tomography scans.

volume constraints.[12] In our study, we achieved similar results in a test set of 25 patients. A larger patient set was also seen in the study of Kierkels et al.[4] In their study, a photon dose distribution was robustly mimicked into an IMPT dose distribution for 40 patients. Our study used a similar robust mimicking optimization approach on predicted proton dose distributions to further use proton beam advantages.

A general question that arises in implementing automated planning is how plan quality can be guaranteed when only MLO plans will be made and clinical plans are not available for comparison. Especially for model-based selection this is a concern, because OAR doses may vary widely between patients and are therefore difficult to evaluate. First, a thorough plan evaluation will be performed for every patient to check whether constraints are fulfilled, as we do now in manual planning. Second, to ensure the lowest possible OAR doses for all patients, an independent dose prediction model may help to identify possible suboptimal MLO plans.[13] This independent dose prediction model predicts OAR doses based on overlap structures with the target. A manual plan can then be made for patients for whom MLO OAR doses are higher than predicted by the dose-prediction model. Interesting to note is the possibility of further improving MLO plan quality by postprocessing the plan in RayStation by manually adding objectives. It is recommended to perform a blinded plan comparison with qualitative analysis by a clinical expert before using an MLO-plan-only workflow. However, postprocessing and blind comparison were outside the scope of the present study.

A limitation of the study is the relatively small set of 25 patients and only 3 patients who did not qualify for IMPT. One should be careful in extrapolating results of automation studies because good results for the (n+1)th patient are not guaranteed, for example when the patient is outside the scope of the training set. Detailed plan evaluation by physicians and physicists remains an important step in the automated planning workflow.

One of the main advantages of MLO planning is the possibility of sharing models with other institutions. The MLO model can be configured toward the institutions' specific clinical goals. This will especially be useful for smaller institutions with limited resources or patients to train their own MLO models.

Another advantage of MLO planning may be in the application of adaptive planning. Currently, adaptive plans usually are finished multiple days after the rCT scan has been acquired. With MLO planning, because the MLO plan can be generated within 45 minutes instead of 2 days, the adaptive plan can be applied earlier, which may increase treatment effectiveness and decrease adverse effects.

Conclusions

This study showed that automated MLO planning can generate robustly optimized MLO plans with quality comparable to clinical plans.

References

1. Langendijk JA, Hoogeman MS, Monshouwer R, Verheij M. Landelijk Indicatie Protocol Protonen Therapie (versie 2.2) HOOFD-HALSTUMOREN, 2019. Available at: https://nvro.nl/publicaties/rapporten. Accessed January 5, 2023.
2. Cheng Q, Roelofs E, Ramaekers BLT, et al. Development and evaluation of an online three-level proton vs photon decision support prototype for head and neck cancer—Comparison of dose, toxicity and cost-effectiveness. *Radiother Oncol* 2016;118:281-285.
3. Delaney AR, Verbakel WF, Lindberg J, Koponen TK, Slotman BJ, Dahele M. Evaluation of an automated proton planning solution. *Cureus* 2018;10:e3696.

4. Kierkels RGJ, Fredriksson A, Both S, Langendijk JA, Scandurra D, Korevaar EW. Automated robust proton planning using dose-volume histogram-based mimicking of the photon reference dose and reducing organ at risk dose optimization. *Int J Radiat Oncol Biol Phys* 2019;103:251-258.
5. Unkelbach J, Alber M, Bangert M, et al. Robust radiotherapy planning. *Phys Med Biol* 2018;63 22TR02.
6. McIntosh C, Welch M, McNiven A, Jaffray DA, Purdie TG. Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method. *Phys Med Biol* 2017;62:5926-5944.
7. Wagenaar D, Kierkels RGJ, van der Schaaf A, et al. Head and neck IMPT probabilistic dose accumulation: Feasibility of a 2 mm setup uncertainty setting. *Radiother Oncol* 2021;154:45-52.
8. Fredriksson A, Forsgren A, Hårdemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. *Med Phys* 2011;38:1672-1684.
9. Korevaar EW, Habraken SJM, Scandurra D, et al. Practical robustness evaluation in radiotherapy—A photon and proton-proof alternative to PTV-based plan evaluation. *Radiother Oncol* 2019;141:267-274.
10. McIntosh C, Purdie TG. Contextual atlas regression forests: Multiple-atlas-based automated dose prediction in radiation therapy. *IEEE Trans Med Imaging* 2016;35:1000-1012.
11. McIntosh C, Purdie TG. Voxel-based dose prediction with multi-patient atlas selection for automated radiotherapy treatment planning. *Phys Med Biol* 2016;62:415-431.
12. Taasti VT, Hong L, Shim JSA, Deasy JO, Zarepisheh M. Automating proton treatment planning with beam angle selection using Bayesian optimization. *Med Phys* 2020;47:3286-3296.
13. Tambas M, van der van der Laan HP, van der van der Schaaf A, Steenbakkers RJHM, Langendijk JA. A decision support tool to optimize selection of head and neck cancer patients for proton therapy. *Cancers (Basel)* 2022;14:681.