



Original Article

Deep learning prediction of proton and photon dose distributions for paediatric abdominal tumours



F. Guerreiro^{a,*}, E. Seravalli^a, G.O. Janssens^{b,c}, J.H. Maduro^{c,d}, A.C. Knopf^d, J.A. Langendijk^d, B.W. Raaymakers^a, C. Kontaxis^a

^a Department of Radiotherapy, University Medical Center Utrecht; ^b Department of Radiation Oncology, University Medical Center Utrecht; ^c Princess Máxima Center for Pediatric Oncology; and ^d Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

ARTICLE INFO

Article history:

Received 6 August 2020

Received in revised form 9 November 2020

Accepted 23 November 2020

Available online 29 November 2020

Keywords:

Deep learning

Dose prediction

Paediatric abdominal tumours

Patient referral

Proton therapy

Photon therapy

ABSTRACT

Objective: Dose prediction using deep learning networks prior to radiotherapy might lead to more efficient modality selections. The study goal was to predict proton and photon dose distributions based on the patient-specific anatomy and to assess their clinical usage for paediatric abdominal tumours.

Material and methods: Data from 80 patients with neuroblastoma or Wilms' tumour was included. Pencil beam scanning (PBS) (5 mm/ 3%) and volumetric-modulated arc therapy (VMAT) plans (5 mm) were robustly optimized on the internal target volume (ITV). Separate 3-dimensional patch-based U-net networks were trained to predict PBS and VMAT dose distributions. Doses, planning-computed tomography images and relevant optimization masks (ITV, vertebra and organs-at-risk) of 60 patients were used for training with a 5-fold cross validation.

The networks' performance was evaluated by computing the relative error between planned and predicted dose-volume histogram (DVH) parameters for 20 inference patients. In addition, the organs-at-risk mean dose difference between modalities was calculated using planned and predicted dose distributions ($\Delta D_{\text{mean}} = D_{\text{VMAT}} - D_{\text{PBS}}$). Two radiation oncologists performed a blind PBS/VMAT modality selection based on either planned or predicted ΔD_{mean} .

Results: Average DVH differences between planned and predicted dose distributions were $\leq |6\%|$ for both modalities. The networks classified the organs-at-risk D_{mean} difference as a gain ($\Delta D_{\text{mean}} > 0$) with 98% precision. An identical modality selection based on planned compared to predicted ΔD_{mean} was made for 18/20 patients.

Conclusion: Deep learning networks for accurate prediction of proton and photon dose distributions for abdominal paediatric tumours were established. These networks allowing fast dose visualisation might aid in identifying the optimal radiotherapy technique when experience and/or resources are unavailable.

© 2020 The Authors. Published by Elsevier B.V. Radiotherapy and Oncology 156 (2021) 36–42 This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Neuroblastoma (NBL) and Wilms' tumour (WT) belong to the most common types of abdominal cancer in children [1,2]. Radiation therapy (RT) to the primary tumour and regional nodes is currently delivered in patients with high-risk NBL and a subset of patients with medium-risk NBL and WT [3,4]. In children, extra caution is required with the delivery of radiation due to the radiosensitive nature of their developing tissue. With more-advanced photon RT techniques, high-doses can be conformably

delivered to complex target volumes. However, low-dose levels still remain widely spread within the normal tissue (NT). Although reported late toxicities from RT seem to be inherent to the target volume selection, the clinical translation of low-dose levels towards the anterior part of the abdomen still remains unclear for this patient category [5–8].

With intensity-modulated proton therapy, there is hope of further reducing the likelihood of RT long-term side effects as a result of the unique proton dose-deposition pattern characterised by the minimal exit dose [9,10]. While the standard use of proton therapy is more established in paediatric central nervous system and skull base tumours [11], its utility in extracranial lesions depends primarily on the patient-specific tumour location in relation to the organs-at-risk (OARs) and expected improvement in long-term side effects. In addition, the high costs involved and the limited availability of proton centres worldwide inhibits the use of protons

* Corresponding author at: Department of Radiotherapy, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, HP Q.02.2.312, The Netherlands.

E-mail addresses: F.Guerreiro@umcutrecht.nl (F. Guerreiro), E.Seravalli@umcutrecht.nl (E. Seravalli), G.O.Janssens@umcutrecht.nl (G.O. Janssens), j.h.maduro@umcg.nl (J.H. Maduro), a.c.knopf@umcg.nl (A.C. Knopf), j.a.langendijk@umcg.nl (J.A. Langendijk), b.w.raaymakers@umcutrecht.nl (B.W. Raaymakers), c.kontaxis@umcutrecht.nl (C. Kontaxis).

as standard technique to treat the majority of the paediatric indications.

To assess the efficacy of proton against photon therapy in adults, Langendijk et al. have proposed the model-based approach [12,13]. The main goal is to select individual patients that show a significant NT dose reduction in favour of protons in a treatment planning comparison study. Patients will be selected for proton therapy if the NT sparing translates in a reduction of the treatment-related NT complications probability (NTCP). Although promising for adult patients, NTCP models are currently not available for children. Thus, patient referrals in paediatric cancer are nowadays done depending on the tumour category, the fast access to a proton facility and occasionally based on a proton-photon dose distribution comparison. Performing treatment plan comparisons is difficult to achieve in practice (1) due to the potential delay introduced in the start of treatment [14], (2) because it's a resource intensive approach as both proton and photon dose distributions need to be computed and (3) because hospitals without proton capabilities lack in either the availability of a treatment planning system (TPS) and/or the clinical experience to calculate optimal proton dose distributions.

The goal of this study was to develop a fast and fully-automated tool using deep learning networks to predict both proton and photon dose distributions based solely on the patient-specific anatomy for children with abdominal tumours (NBL and WT). Moreover, to demonstrate the potential clinical application of such networks, both planned and predicted dose distributions were used to assist in the treatment modality selection process.

Methods and material

Patient cohort and imaging details

After institutional review board approval (WAG/mb/17/008865), data from 80 patients, 46 NBL and 34 WT patients (mean: 3, range: 1–10 years), undergoing treatment at the radiotherapy department of the University Medical Centre Utrecht (UMCU) between January 2015 and January 2020 was included in this study. The female/male ratio was 32/48 and the left/right tumour location ratio was 49/31.

During treatment preparation, a 4-dimensional computed tomography (4D-CT) was acquired for every patient fixated in a vacuum mattress (Bluebag, Elekta, Stockholm, Sweden) using a 16-, 40- or 64-channel detector scanner (Brilliance, Philips Medical Systems, Best, The Netherlands). 4D-CT images were obtained as a series of ten breathing phases which were averaged to obtain the planning-CT.

Planning details

The clinical target and OARs (ipsilateral and contralateral kidneys, liver, spleen and pancreas) contours delineated by a paediatric radiation oncologist were used in this study. The gross tumour volume (GTV) consisted of the pre-operative tumour extension, including the pathologic lymph nodes, residual disease if present and correction for post-operative organ shift. The clinical target volume was created using an expansion of the GTV by 5 mm for NBL and by 5 to 10 mm for WT patients. To compensate for the breathing motion during planning, a patient-specific internal target volume (ITV) was delineated by measuring surgical clips displacements between the maximum and minimum 4D-CT phases [15].

For the purpose of this study, proton and photon planning were performed using RayStation (RaySearch Laboratories, Stockholm, Sweden). Intensity-modulated pencil beam scanning (PBS) plans were optimized using 2–3 posterior-oblique irradiation fields to

exploit the dorsal location of the target. The number and orientation of the proton beams (range [90°; 240°]) were patient-specific, depending on the plan robustness achievement in addition to the shape and location of the target. Volumetric-modulated arc therapy (VMAT) plans consisted of a 6 MV 360°-arc for all patients. The prescribed dose ranged from 10.8 to 36 Gy and was delivered in 6 to 20 fractions, depending on the histology and the presence of residual disease.

3D plan optimization was performed using the planning-CT and a uniform dose grid of 3 mm for both modalities. To ensure planning consistency between modalities, all plans were optimized by one experienced user. A pencil beam algorithm and a collapse cone engine were used for PBS and VMAT plans optimization, respectively. Both dose distributions were ITV-based robustly optimized using a minimax optimization method [16] and accounting for a 5 mm patient set-up uncertainty and a 3% range uncertainty in case of PBS. The robustness of the planned dose distributions was evaluated using a voxel-wise minimum dose map (V_{wmin}) [17]. During robustness evaluation, PBS and VMAT dose distributions were computed for different scenarios using the same 5 mm set-up uncertainties (in 14 fixed directions) and range uncertainties ($\pm 3\%$) in case of PBS. In total, 28 and 14 evaluation scenarios were calculated for the PBS and VMAT dose distributions, respectively. V_{wmin} was computed by extracting, for each voxel, the minimum dose across all evaluation scenarios. Plans were considered clinically robust if in the V_{wmin} 98% of the ITV received at least 95% of the prescribed dose [17]. For all OARs, planning was performed based on the as-low-as-reasonably-achievable (ALARA) principle for both modalities.

In addition, to reduce the risk of asymmetric bone growth, a homogeneous dose with left–right dose gradients lower than 3 to 5 Gy was aimed for the vertebra primary ossification centres adjacent to the ITV [18].

Deep learning network

Two separate 3D patch-based U-net networks were used in this study to predict PBS and VMAT dose distributions. The U-net architecture was previously published in [19] and includes a decoding and encoding path.

Data

Each network used a set of 3D volumes as input and predicted a single 3D dose distribution with the same dimensions. The experiments, including all input and output data, were performed at a uniform 3 mm grid spacing which was the resolution used during treatment planning (See Planning Details).

In total, 10 channels were selected for the input data including the planning-CT, ITV, OARs and vertebra contours. No additional information regarding the PBS/VMAT beam arrangement and planning constraints/objectives was given during the networks training. The Hounsfield units of the planning-CT were used while each optimization contour was set as a separate binary mask in its own channel. In case the patient had missing one of the OARs, the corresponding channel was put to zero for that input. The ITV and vertebra binary masks were set to the patient-specific prescribed dose allowing for the networks to discriminate between different clinical dose prescriptions. In addition, all planned and predicted dose distributions were normalized such that the ITV mean dose (D_{mean}) was equal to the patient prescribed dose.

Network training

To assess the overall performance of the networks, 60 patients were included in a 5-fold cross validation procedure while the remaining 20 patients were used for inference. In the training phase, each fold divided the 60 patients into 48 training and 12 val-

ication patients. This led to five separate U-net networks per modality, which were trained and validated including a unique patient data combination.

Each network/fold was trained using a patch-based method, processing random 64x64x64 sub-volumes from the full anatomy using a batch size of four per iteration. The ADAM optimizer was used with the learning rate, beta1, beta2 and epsilon parameters set to 10^{-4} , 0.9, 0.999 and 10^{-8} respectively. The root mean squared error (RMSE) was selected as the loss function. All networks/folds were trained using a single NVIDIA® GTX Titan card.

Evaluation

Predicted PBS and VMAT dose distributions were calculated for the 20 inference patients by averaging the dose distributions predicted by each fold. For these patients, the NBL/WT ratio was 8/12, the female/male ratio was 11/9 and the left/right tumour location ratio was 10/10.

Network performance

To evaluate the networks performance, the prediction accuracy was determined by calculating: (1) relative errors between the planned (ground truth) and the predicted dose-volume histogram (DVH) metrics (2) mean relative errors between the planned (ground truth) and the predicted dose distributions within the body volume ($\text{Diff}_{\text{mean}}$) and (3) dice similarity coefficient (DSC) for selected isodose lines. For (1), DVH parameters including the $D_{98\%}$, $D_{50\%}$ and $D_{2\%}$ for the ITV, D_{mean} and $D_{2\%}$ for the OARs and D_{mean} for the vertebra were computed. Moreover, the p -value of a two-tailed student's t -test at a 0.05 significance level and the Pearson's correlation coefficient (r) were calculated to assess the differences between planned and predicted DVH metrics for each modality. For (2) and (3), $\text{Diff}_{\text{mean}}$ and DSC metrics were calculated using the following dose thresholds: 10%, 30%, 50%, 70% and 90% of the prescribed dose.

Clinical usage

The application of these networks, allowing a fast proton-photon dose distribution comparison, might be recurrent in clinic when needing to select the optimal treatment modality for a specific patient case. Thus, the potential clinical use of the predicted dose distributions was quantitatively and qualitatively evaluated.

For the quantitative evaluation, the absolute differences on the OARs D_{mean} between both modalities were calculated when using (1) planned and (2) predicted dose distributions ($\Delta D_{\text{mean}} = D_{\text{VMAT}} - D_{\text{PBS}}$, in Gy). In addition, the p -value of a two-tailed student's t -test at a 0.05 significance level and the Pearson's correlation coefficient (r) were calculated to assess the differences between planned and predicted ΔD_{mean} for each modality.

For the qualitative evaluation, two experienced paediatric radiation oncologists performed a blind PBS/VMAT modality selection for the 20 inference patients. For each patient, one case of planned and one case of predicted PBS and VMAT dose distributions were available. In total, 40 cases including the computed ΔD_{mean} and DVHs for selected structures (ITV, vertebra and OARs) were presented in random order to the two radiation oncologists. For each case, the radiation oncologists were instructed to objectively select the preferred modality. Decisions were based on the ALARA principle, on a prioritisation between the OARs and on specific OARs constraints. The ipsilateral kidney ($D_{\text{mean}} < 12\text{--}15$ Gy) and pancreas ($D_{\text{mean}} < 10$ Gy) were classified as very important, the contralateral kidney ($D_{\text{mean}} < 5\text{--}8$ Gy) and spleen ($D_{\text{mean}} < 10$ Gy) as important and the liver as a less important OAR. Preferences between planned and predicted cases for each patient were matched and the number of patients for which the modality selection was identical was reported.

Results

The PBS and VMAT networks training took approximately 3.6 h per fold. All folds were trained for a total of 6400 iterations, at which the relationship between the validation and training losses had stabilized (Fig. 1). The average training/validation RMSE among all folds were 0.06/0.1 Gy for PBS and 0.07/0.1 Gy for VMAT. The inference of a single 3D predicted dose distribution took approximately 0.5 sec per patient and modality.

At the target location, a very good similarity between the planned and predicted dose distributions was observed for both modalities (Fig. 2). The ITV coverage ($D_{98\%}$, $D_{50\%}$) was predicted on average with less than 1% difference from the planned dose distribution for both modalities (Table 1). Moreover, the required dose homogeneity in the vertebra was well reproduced in the predicted doses for all inference patients and modalities (Fig. 2). For this structure, the average D_{mean} relative error was below 1% (range $[-5.3; 5.2]$ %) for both PBS and VMAT dose distributions. Larger differences between planned and predicted dose distributions were denoted for the low-dose regions for both modalities (Fig. 2). For all OARs, the average relative error between the planned and the predicted D_{mean} was $-0.3 \pm 2.9\%$ (range $[-8.0; 11.1]$ %) for PBS and $-1.5 \pm 6.8\%$ (range $[-26.8; 19.2]$ %) for VMAT. Similarly, the average relative error for the $D_{2\%}$ was $-2.3 \pm 7.5\%$ (range $[-26.3; 14.8]$ %) for PBS and $-2.6 \pm 8.7\%$ (range $[-32.7; 27.4]$ %) for VMAT. Nevertheless, a high correlation ($r > 0.9$) and no significant differences for the majority of the DVH parameters were found between planned and predicted dose distributions for both modalities (Fig. 3, Table 1).

Table 2 shows the average $\text{Diff}_{\text{mean}}$ and DSC computed for selected isodose lines among all patients. For all dose thresholds, average $\text{Diff}_{\text{mean}}$ errors were below 2% for both PBS (range $[-8.2; 7.2]$ %) and VMAT (range $[-5.8; 6.6]$ %) dose distributions. In addition for both modalities, high average DSC values (≥ 0.9 , range $[0.81; 0.98]$) were computed for all isodose lines.

Fig. 4 denotes the planned versus predicted ΔD_{mean} for all OARs in the inference cohort. A high correlation ($r = 0.88$) between planned and predicted ΔD_{mean} was found. The networks classified the D_{mean} difference between VMAT and PBS dose distributions on the OARs as a gain ($\Delta D_{\text{mean}} > 0$) with 98% precision. This result was supported by the qualitative evaluation of the planned/predicted dose distributions, in which an identical modality selection was achieved for 18 (out of 20) patients for both radiation oncologists. The modality choice differed between the planned and predicted dose distributions for two patients as a result of (1) crossing PBS and VMAT DVHs in one of the planned or predicted cases (e.g ipsilateral kidney) and/or (2) large differences between the planned and predicted ΔD_{mean} (e.g pancreas and spleen).

Discussion

In this study, deep learning networks for accurate prediction of proton and photon dose distributions based solely on the patient-specific anatomy were presented for paediatric abdominal tumours. For both modalities, relative errors in the predicted dose distributions were minimal in the target dose region. Moreover, average dose differences between planned and predicted dose distributions were below 2% within the body volume. The clinical evaluation showed that both networks accurately predicted the difference between photon and proton dose distributions.

Machine learning and more recently deep learning methods have been applied to the field of radiotherapy treatment planning in the last few years. The capability to predict accurate proton or photon dose distributions has been demonstrated for several treatment sites including prostate, head and neck and lungs in adult

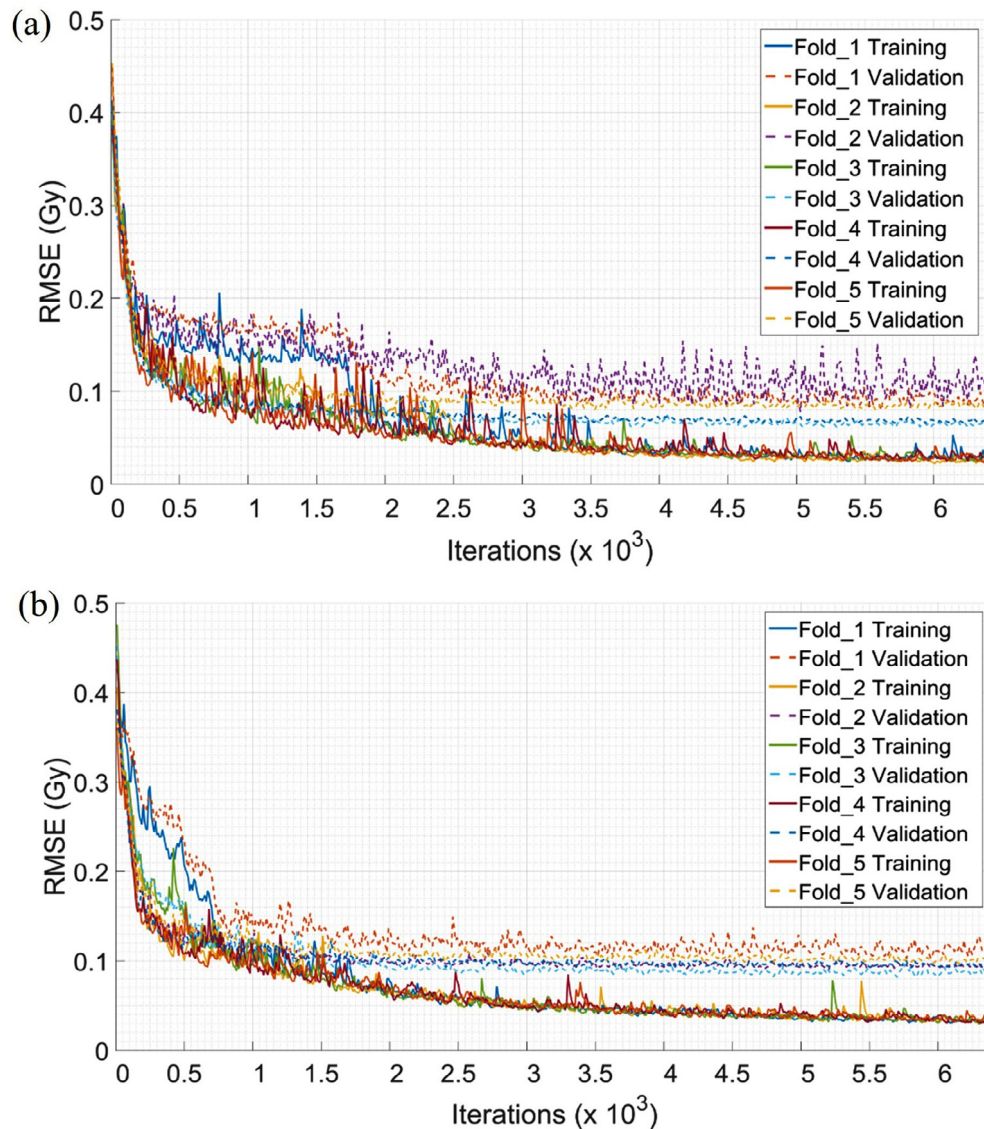


Fig. 1. Training and validation RMSE (Gy) for PBS (a) and VMAT (b) networks for all 5 folds.

patients [20–23]. To the authors' knowledge, this is the first work to predict both proton and photon dose distributions in a paediatric cohort.

Fast and accurate proton and photon dose distribution predictions were achieved in this study. Although each network was trained for approximately 18 h for all folds, the proton/photon dose prediction for a new patient could be achieved in the order of a few seconds after finishing the contouring process. In the target dose region, the average relative error between planned and predicted dose distributions was small ($< 1\%$) for both modalities. In addition, the mandatory dose homogeneity in the vertebra to prevent the risk of asymmetric bone growth was well reproduced in both predicted PBS and VMAT dose distributions. The average relative errors for the OARs DVH parameters were within 6%, while some larger individual differences were observed mostly for the $D_{2\%}$ points. On a voxel-by-voxel basis, $D_{2\%}$ points can be more sensitive to deviations in the dose distributions as a result of the large variability in the patient population. In this study, networks were trained and evaluated with data of 80 paediatric patients presenting a wide range of different target shapes, sizes and locations. Facing the low number of NBL and WT patients treated per year,

imaging and contour datasets for over 80 patients in a single institute are rare. Despite the use of a relatively small and inhomogeneous training dataset, the proposed deep learning networks were able to create models which coped with the morphological variability of the paediatric population. A strong correlation ($r \geq 0.9$) between planned and predicted DVH parameters and large DSC values (≥ 0.9) between planned and predicted isodose lines for both modalities were denoted.

In addition to the patient-specific anatomy, the beam configuration chosen, the plan-specific constraints and manual sub-optimal treatment planning can potentially affect the dose prediction accuracy. In this study, no additional information about the patient-specific number of beams, gantry angles and the physician priority list of target/OARs planning trade-offs was given during the networks training. Nevertheless, beam arrangement was kept consistent for both modalities (posterior irradiation fields vs. full-arc) and dose distributions were optimized by one experienced planner to avoid planning inconsistency between patients. Future work should focus on including this information in the networks' input to allow for a higher level of personalised and accurate dose prediction for each patient.

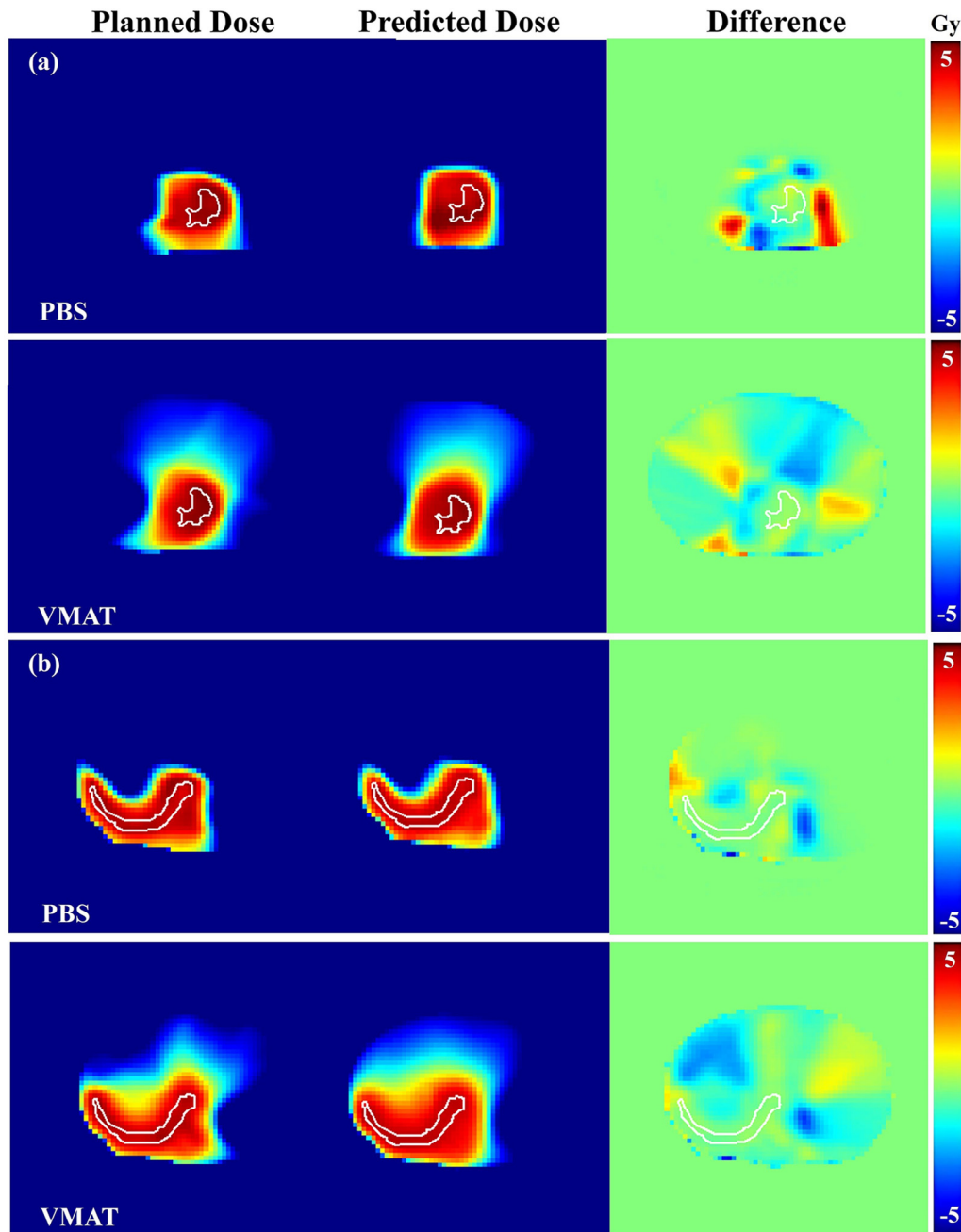


Fig. 2. Comparison between PBS and VMAT planned and predicted transversal dose maps for a NBL (a) and a WT (b) inference patients (prescribed dose = 21.6 Gy and 14.4 Gy, respectively). Planned – predicted dose difference maps are shown on the right using a $[-5; 5]$ Gy dose scale for both patients. The ITV is shown in white.

Although the efficacy of proton therapy is well established for children with intracranial tumours [11], the use of protons in abdominal disease depends primarily on the balance between the patient-specific tumour location and the expected radiotherapy-related long-term effects. In addition, the costs and treatment logistics often play an important role. Consequently, patient referrals in favour of proton therapy are commonly done for selected tumour categories and occasionally performed after proton-photon planning comparison, which is a labour intensive approach. Facing these limitations, dose prediction networks might aid clinicians to promptly identify the optimal radiotherapy treatment modality when experience, time and/or required resources are limited. The clinical evaluation performed in this study showed that planned and predicted OARs ΔD_{mean} were strongly correlated. A

dosimetric gain of PBS in comparison to VMAT was predicted with 98% precision. Additionally, identical patient referral based on planned compared to predicted dose distributions was made by two radiation oncologists for the majority of the patients (18 out of 20). These results show that the proposed networks can accurately predict the treatment planning outcome and might be used as a clinical guidance tool for treatment modality selection. Nevertheless, the final decision for referral belongs still to the radiation oncologist's responsibility.

In this work, networks were trained and validated to fast and accurately predict proton and photon dose distributions based solely on the patient-specific anatomy for children with abdominal tumours. The results presented in this study are a promising step for the future clinical implementation of deep learning dose pre-

Table 1

Average relative errors (\pm standard deviation) between planned and predicted dose-volume histogram (DVH) parameters for both modalities. The p -value of a two-tailed student's t -test and the Pearson's correlation coefficient (r) are also shown. Significant differences ($p < 0.05$) are shown in bold.

Structure	Parameter	PBS			VMAT		
		Error (%)	p	r	Error (%)	p	r
ITV	D _{98%}	-0.4 ± 0.9	0.12	1.0	-0.8 ± 0.5	2.6E-06	1.0
	D _{50%}	0.0 ± 0.3	0.71	1.0	0.0 ± 0.1	0.43	1.0
	D _{2%}	0.5 ± 2.2	0.51	0.99	1.1 ± 0.8	2.1E-06	1.0
Vertebra	D _{mean}	0.5 ± 2.7	0.69	0.99	0.1 ± 2.1	0.70	0.99
Ipsilateral Kidney	D _{mean}	0.5 ± 5.2	0.95	0.99	-3.1 ± 4.1	0.01	0.99
	D _{2%}	0.1 ± 4.4	0.29	0.99	-0.3 ± 2.0	0.85	1.0
Contralateral Kidney	D _{mean}	-1.1 ± 2.7	0.09	0.99	-2.7 ± 8.2	0.07	0.93
	D _{2%}	-4.7 ± 6.5	6.6E-04	0.99	-2.2 ± 11.6	0.02	0.91
Liver	D _{mean}	0.2 ± 1.3	0.46	0.98	0.1 ± 4.9	0.96	0.90
	D _{2%}	1.8 ± 6.5	0.97	0.98	0.2 ± 3.7	0.97	0.99
Spleen	D _{mean}	0.4 ± 1.9	0.61	1.0	0.6 ± 6.6	0.95	0.98
	D _{2%}	0.8 ± 3.9	0.67	1.0	-1.9 ± 11.3	0.26	0.97
Pancreas	D _{mean}	-1.1 ± 3.7	0.20	1.0	-3.3 ± 7.6	0.05	0.98
	D _{2%}	-5.9 ± 10.0	0.02	0.97	-3.7 ± 6.7	0.02	0.97

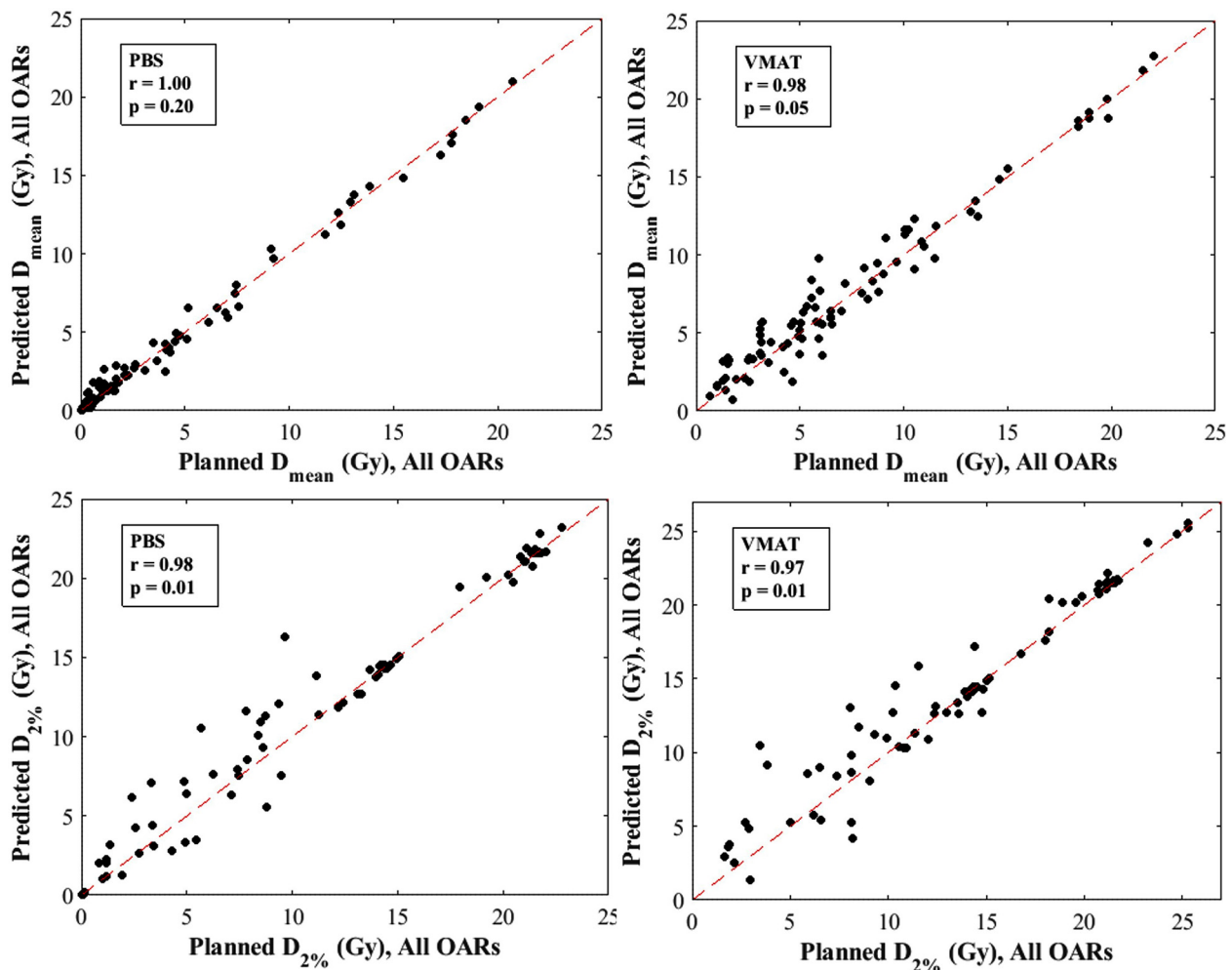


Fig. 3. Predicted versus planned D_{mean} (top)/D_{2%} (bottom) for all OARs (kidneys, liver, spleen and pancreas) for both PBS (left) and VMAT (right) modalities. Values are presented in Gy. The p -value of a two-tailed student's t -test and the Pearson's correlation coefficient (r) are also shown.

diction networks which can have an immediate application as a guidance tool for treatment modality selection but also as a quality control tool. Planners might immediately view if a dose distribution can be modified by either adjusting contour delineations and/or dose constraints preferences. This would potentially improve plan quality while reducing planning time and eliminating planning diversity between centres and physicians. Further-

more, despite the proposed networks were trained for NBL and WT patients using dose distributions with specific beam configurations and fixed robustness uncertainties, they have the potential to be used for other tumour sites and radiotherapy techniques. Thus, the creation of large and tumour-specific databases accounting for multicentre treatment planning diversity should be encouraged by the medical community.

Table 2

Average (\pm standard deviation) mean relative errors between planned and predicted dose distributions ($\text{Diff}_{\text{mean}}$) and dice similarity coefficient (DSC) for selected isodose lines for both modalities.

Isodose	PBS		VMAT	
	$\text{Diff}_{\text{mean}}$ (%)	DSC	$\text{Diff}_{\text{mean}}$ (%)	DSC
10%	1.0 ± 3.4	0.93 ± 0.02	0.3 ± 3.1	0.93 ± 0.03
30%	1.5 ± 2.9	0.94 ± 0.02	1.2 ± 3.4	0.90 ± 0.03
50%	1.4 ± 2.4	0.94 ± 0.01	1.7 ± 2.9	0.89 ± 0.03
70%	1.1 ± 2.0	0.94 ± 0.01	1.4 ± 2.1	0.91 ± 0.02
90%	0.9 ± 1.3	0.89 ± 0.04	0.8 ± 0.8	0.90 ± 0.03

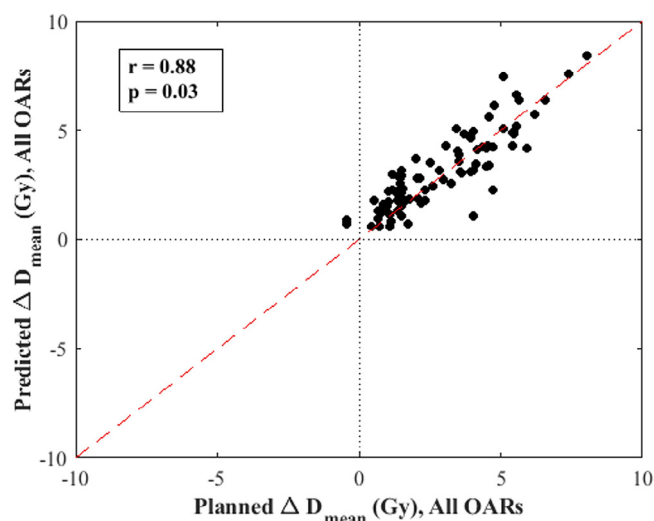


Fig. 4. Predicted versus planned ΔD_{mean} for all OARs in the inference cohort. ΔD_{mean} denotes the difference between VMAT minus PBS dose distributions. The p -value of a two-tailed student's t -test and the Pearson's correlation coefficient (r) are also shown.

Declaration of Competing Interest

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

References

- Maris JM, Hogarty MD, Bagatell R, Cohn SL. Neuroblastoma. *Lancet* 2007;369:2106–20. [https://doi.org/10.1016/S0140-6736\(07\)60983-0](https://doi.org/10.1016/S0140-6736(07)60983-0).
- Brok J, Treger TD, Gooskens SL, et al. Biology and treatment of renal tumours in childhood. *Eur J Cancer* 2016;68:179–95. <https://doi.org/10.1016/j.ejca.2016.09.005>.
- van den Heuvel-Eibrink MM, Hol JA, Pritchard-Jones K, van Tinteren H, Furtwängler R, Verschuur AC, et al. Position paper: rationale for the treatment of Wilms tumour in the UMBRELLA SIOP-RTSG 2016 protocol. *Nat Rev Urol* 2017;14:743–52. <https://doi.org/10.1038/nrurol.2017.163>.
- Ladenstein R, Pötschger U, Pearson ADJ, Brock P, Luksch R, Castel V, et al. Busulfan and melphalan versus carboplatin, etoposide, and melphalan as highdose chemotherapy for high-risk neuroblastoma (HR-NBL1/SIOPEN): an international, randomised, multi-arm, open-label, phase 3 trial. *Lancet* 2017;18:500–14. [https://doi.org/10.1016/S1470-2045\(17\)30070-0](https://doi.org/10.1016/S1470-2045(17)30070-0).
- Sasso G, Greco N, Murino P, Sasso FS. Late toxicity in Wilms tumor patients treated with radiotherapy at 15 years of median follow-up. *J Pediatr Hematol Oncol* 2010;32:264–7. <https://doi.org/10.1097/mpb.0b013e3181e7931a>.
- Ducassou A, Gambart M, Munzer C, Padovani L, Carrie C, Haas-Kogan D, et al. Long-term side effects of radiotherapy for pediatric localized neuroblastoma: results from clinical trials NB90 and NB94. *Strahlenther Onkol* 2015;191:604–12. <https://doi.org/10.1007/s00066-015-0837-z>.
- Madenci AL, Fisher S, Diller LR, Goldsby RE, Leisenring WM, Oeffinger KC, et al. Intestinal obstruction in survivors of childhood cancer: A report from the childhood cancer survivor study. *J Clin Oncol* 2015;33:2893–900. <https://doi.org/10.1200/JCO.2015.61.5070>.
- Meadows AT, Friedman DL, Neglia JP, Mertens AC, Donaldson SS, Stovall M, et al. Second neoplasms in survivors of childhood cancer: findings from the Childhood Cancer Survivor Study cohort. *J Clin Oncol* 2009;27:2356–62. <https://doi.org/10.1200/JCO.2008.21.1920>.
- Timmermann B, Schuck A, Niggli F, Weiss M, Lomax AJ, Pedroni E, et al. Spot-scanning proton therapy for malignant soft tissue tumors in childhood: First experiences at the Paul Scherrer Institute. *Int J Radiat Oncol Biol Phys* 2007;67:497–504. <https://doi.org/10.1016/j.ijrobp.2006.08.053>.
- Lomax AJ, Boehringer T, Coray A, Egger E, Goitein G, Grossmann M, et al. Intensity modulated proton therapy: A clinical example. *Med Phys* 2001;28:317–24. <https://doi.org/10.1118/1.1350587>.
- Journy N, Indelicato DJ, Withrow DR, Akimoto T, Alapetite C, Araya M, et al. Patterns of proton therapy use in pediatric cancer management in 2016: an international survey. *Radiother Oncol* 2018;132:155–61. <https://doi.org/10.1016/j.radonc.2018.10.022>.
- Langendijk JA, Lambin P, De Ruyscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: The model-based approach. *Radiother Oncol* 2013;107:267–73. <https://doi.org/10.1016/j.radonc.2013.05.007>.
- Widder J, Van Der Schaaf A, Lambin P, Marijnien CAM, Pignol JP, Rasch CR, et al. The quest for evidence for proton therapy: Model-based approach and precision medicine. *Int J Radiat Oncol Biol Phys* 2016;95:30–6. <https://doi.org/10.1016/j.ijrobp.2015.10.004>.
- Stokes CL, Stokes WA, Kalapurakal JA, Paulino AC, Cost NG, Cost CR, et al. Timing of radiation therapy in pediatric wilms tumor: A report from the national cancer database. *Int J Radiat Oncol Biol Phys* 2018;101:453–61. <https://doi.org/10.1016/j.ijrobp.2018.01.110>.
- Guerrero F, Seravalli E, Janssens GO, van de Ven CP, van den Heuvel-Eibrink MM, Raaymakers BW. Intra- and inter-fraction uncertainties during IGRT for Wilms' tumor. *Acta Oncol* 2018;57:941–9. <https://doi.org/10.1080/0284186X.2018.1438655>.
- Fredriksson A, Forsgren A, Hårdemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. *Med Phys* 2011;38:1672–84. <https://doi.org/10.1118/1.3556559>.
- Korevaar EW, Habraken SJM, Scandurra D, Kierkels RGJ, Unipan M, Eenink MGC, et al. Practical robustness evaluation in radiotherapy – A photon and proton-proof alternative to PTV-based plan evaluation. *Radiother Oncol* 2019;141:267–74. <https://doi.org/10.1016/j.radonc.2019.08.005>.
- Hoeben BA, Carrie C, Timmermann B, Mandeville HC, Gandola L, Dieckmann K, et al. Management of vertebral radiotherapy dose in paediatric patients with cancer: consensus recommendations from the SIOPE radiotherapy working group. *Lancet Oncol* 2019;20:155–66. [https://doi.org/10.1016/S1470-2045\(19\)30034-8](https://doi.org/10.1016/S1470-2045(19)30034-8).
- Kontaxis C, Bol GH, Lagendijk JJW, Raaymakers BW. DeepDose: Towards a fast dose calculation engine for radiation therapy using deep learning. *Phys Med Biol* 2020;65. <https://doi.org/10.1088/1361-6560/ab7630705103>.
- Nguyen D, Jia X, Sher D, Lin MH, Iqbal Z, Liu H, et al. 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture. *Phys Med Biol* 2019;64. <https://doi.org/10.1088/1361-6560/ab039b065020>.
- Kontaxis C, Bol GH, Lagendijk JJW, Raaymakers BW. PV-0423: Fast automated IMRT sequencing using deeplearned dose from generative adversarial networks. *Radiother Oncol* 2019;133:S220–1.
- Barragán-Montero AM, Nguyen D, Lu W, Lin M-H, Norouzi-Kandalan R, Geets X, et al. Three-dimensional dose prediction for lung IMRT patients with deep neural networks: robust learning from heterogeneous beam configurations. *Med Phys* 2019;46:3679–91. <https://doi.org/10.1002/mp.13597>.
- Wu C, Nguyen D, Xing Y, Barragán-Montero AM, Schuemann J, Shang H et al. Improving Proton Dose Calculation Accuracy by Using Deep Learning. <https://arxiv.org/ftp/arxiv/papers/2004/2004.02924.pdf> [2020].