PAPER • OPEN ACCESS

Improving proton dose calculation accuracy by using deep learning

View the article online for updates and enhancements.

MACHINE
LEARNING
Science and Technology

PAPER

Improving proton dose calculation accuracy by using deep learning

Chao Wu[1,2,3] 🆔, Dan Nguyen[1] 🆔, Yixun Xing[1], Ana Barragan Montero[3,4], Jan Schuemann[5] 🆔, Haijiao Shang[2,3], Yuehu Pu[2] and Steve Jiang[1]

[1] Medical Artificial Intelligence and Automation (MAIA) Laboratory, Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX, United States of America
[2] Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai, People's Republic of China
[3] University of Chinese Academy of Sciences, Beijing, People's Republic of China
[4] Molecular Imaging Radiation Oncology (MIRO) Laboratory, UCLouvain, Brussels, Belgium
[5] Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA, United States of America

E-mail: Steve.Jiang@UTSouthwestern.Edu

Abstract

Pencil beam (PB) dose calculation is fast but inaccurate due to the approximations when dealing with inhomogeneities. Monte Carlo (MC) dose calculation is the most accurate method but it is time consuming. The aim of this study was to develop a deep learning model that can boost the accuracy of PB dose calculation to the level of MC dose by converting PB dose to MC dose for different tumor sites. The proposed model uses the PB dose and computed tomography image as inputs to generate the MC dose. We used 290 patients (90 head and neck, 93 liver, 75 prostate and 32 lung) to train, validate, and test the model. For each tumor site, we performed four numerical experiments to explore various combinations of training datasets. Training the model on data from all tumor sites together and using the dose distribution of each individual beam as input yielded the best performance for all four tumor sites. The average gamma passing rate (1 mm/1%) between the converted and the MC dose was 92.8%, 92.7%, 89.7% and 99.6% for head and neck, liver, lung, and prostate test patients, respectively. The average dose conversion time for a single field was less than 4 s. The trained model can be adapted to new datasets through transfer learning. Our deep learning-based approach can quickly boost the accuracy of PB dose to that of MC dose. The developed model can be added to the clinical workflow of proton treatment planning to improve dose calculation accuracy.

1. Introduction

Proton therapy has attracted increasing attention in recent years. The main rationale for using proton therapy is the physical characteristics of the depth dose curve, which has a dose peak (Bragg peak) at a well-defined depth in tissue [1]. This physical advantage allows proton therapy to achieve higher dose conformity in the tumor volume with a lower dose to the surrounding healthy tissue than conventional radiation therapy. Nonetheless, uncertainties in dose calculation have a bigger impact on the desired dose distributions in proton therapy, so accurate dose calculation is essential for the success of a proton therapy treatment [2]. Currently, many pencil beam (PB)-based dose calculation algorithms, based on the works of Hong *et al* [3] and Schaffner *et al* [4], are widely used in clinical practice. These algorithms provide fast computation but come at the expense of lower dose calculation accuracy in the presence of tissue heterogeneity [5]. This is mainly because these algorithms adopt approximations that disregard lateral inhomogeneities to achieve fast dose calculations [6, 7]. In PB dose calculation, the proton dose is computed by using the water equivalent path length along the central path of a pencil beam, and the medium on the central axis is assumed to be laterally infinite and homogeneous [6]. These approximations lead to inaccurate

modeling of the multiple Coulomb scattering (MCS) process and, therefore, cause both dose distortion and range uncertainties, especially in the presence of complex geometries and heterogeneous environments [8]. Moreover, PB algorithms' approximations in modeling elastic and inelastic nuclear interactions can also lead to considerable errors in dose calculation, even in homogeneous geometries [6, 7].

Monte Carlo (MC) methods are recognized as the gold standard for dose calculation because they can simulate particle propagation through materials by randomly sampling the cross-section of interactions [7–10]. Several phantom studies comparing PB dose calculations and MC algorithms have demonstrated that MC algorithms can provide more accurate dose distributions than PB algorithms [11–14]. For example, in a multi-institution phantom study, Taylor *et al* reported that the dose distribution obtained with the MC algorithm matched the phantom dose measurement more closely than the dose distribution obtained with the PB algorithm [11]. Besides phantom studies, PB dose accuracy deficiencies have also been investigated in several studies comparing PB dose with MC dose in different tumor sites [5, 8, 9, 15]. Schuemann *et al* demonstrated that current PB dose calculation algorithms can cause underdosage to the target by as much as 5%, which can result in differences in tumor control probability of up to 11%. For complex geometries (head and neck cancer and lung cancer) and very deep-seated targets (prostate cancer), MC simulations should be considered instead of PB dose calculations [8]. Moreover, proton therapy can be very sensitive to patient anatomical changes that may distort the planned dose distribution and deteriorate the treatment quality. A recent study has also shown that online adaptation with MC dose calculation can further improve treatment quality for inter-fractional patient geometry changes [16]. Despite all these advantages, MC dose calculation methods have not been widely used in clinical routine because of their computational burden and implementation complexity [7].

Over the years, the research community has devoted significant efforts to accelerating MC dose calculation for proton therapy [17]. Recently, parallel computing techniques based on graphics processing units (GPU) have been employed for this application. Different groups have achieved considerable acceleration factors over conventional central processing unit (CPU)-based computations [17–24]. To date, the computation time of the fastest reported MC dose engine which adopts some approximations is in the minute range. Furthermore, some commercial treatment planning systems (e.g. RayStation and Eclipse) have already included the MC dose calculation feature for proton therapy.

Recently, deep convolutional neural networks have been successfully used to predict patient-specific dose distributions from anatomical information [25–32]. Although these works tried to learn the relationship between patient anatomy and the optimal dose distribution, they did inspire us to use deep learning methods to learn the relationship between tissue inhomogeneities and the differences between PB and MC dose distributions.

In this paper, we present a novel approach that uses deep learning techniques to achieve MC dose calculation accuracy with PB dose calculation efficiency for proton therapy. Specifically, we developed a deep learning model that can precisely and efficiently convert a 3D PB dose distribution to a dose distribution with MC-equivalent accuracy for different tumor sites. The model architecture and training details are presented in section 2. In section 3, we present the results of the proposed model. We discuss some future work and draw conclusions in section 4.

2. Methods

2.1. Model architecture

The model developed in this work is based on the 3D HD U-Net that was developed and tested for voxel-wise 3D dose prediction for patients with head and neck cancer [27]. HD U-Net combines the essence of two influential neural networks: U-Net [33] and DenseNet [34]. In general, HD U-Net's architecture combines U-Net's ability to abstract both local and global features from input images with DenseNet's efficient feature propagation and reuse, while maintaining a reasonable memory usage [27, 31]. Figure 1 shows the architecture of the HD U-Net model modified for this work. The details of the three operations between layers (Dense Convolve, Dense Downsample and U-Net Upsample) have been previously introduced elsewhere [27].

The proposed model contains two input channels: one for the 3D proton PB dose distribution and the other for the corresponding computed tomography (CT) image. The model has 5 max pooling and 5 upsampling operations, which decrease the image patch size from $128 \times 128 \times 16$ voxels to $8 \times 8 \times 1$ voxels, then increase it back to $128 \times 128 \times 16$ voxels. The convolutional kernel size is set to $3 \times 3 \times 3$, and the max pooling size is set to $2 \times 2 \times 1$. Batch normalization is added after the convolution and before the rectified linear unit operations. The dropout rate is set to 0 because no overfitting issue has been observed during training.
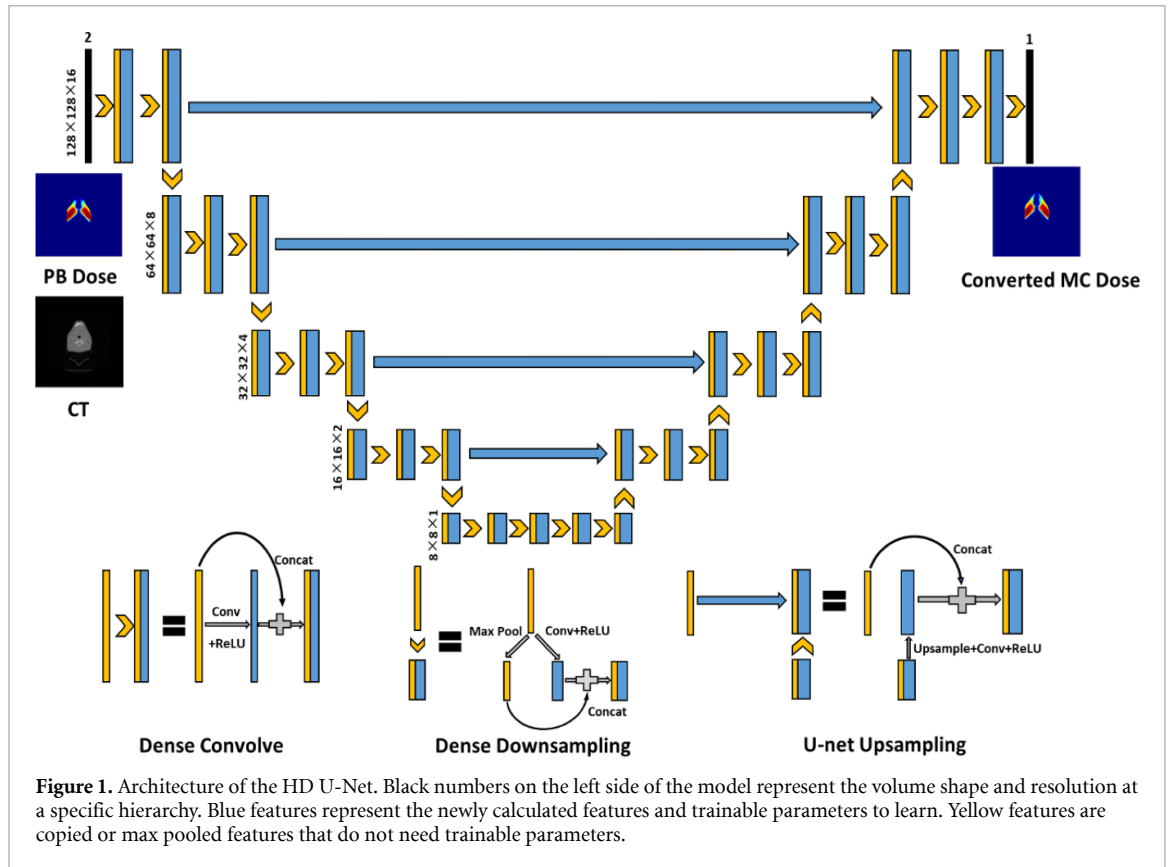
Figure 1. Architecture of the HD U-Net. Black numbers on the left side of the model represent the volume shape and resolution at a specific hierarchy. Blue features represent the newly calculated features and trainable parameters to learn. Yellow features are copied or max pooled features that do not need trainable parameters.

Table 1. Beam numbers and distribution details of the testing set and the training & validation set for each tumor site.
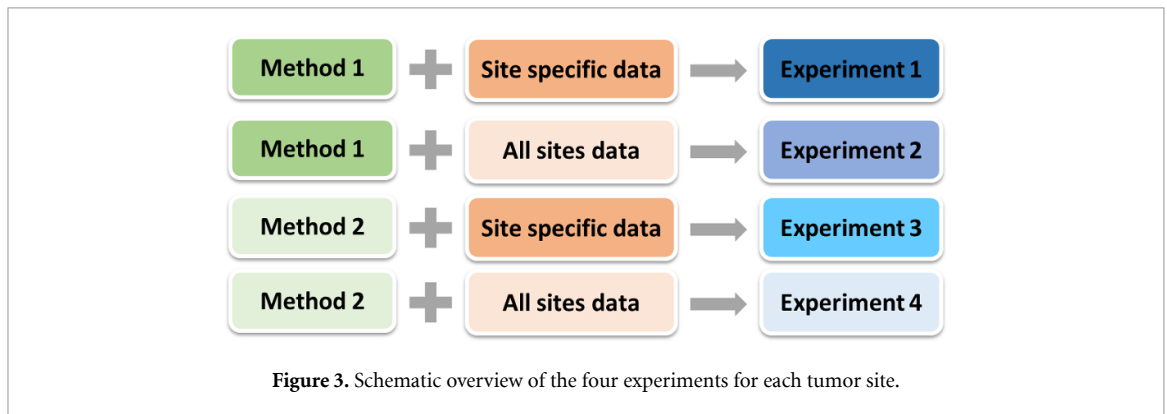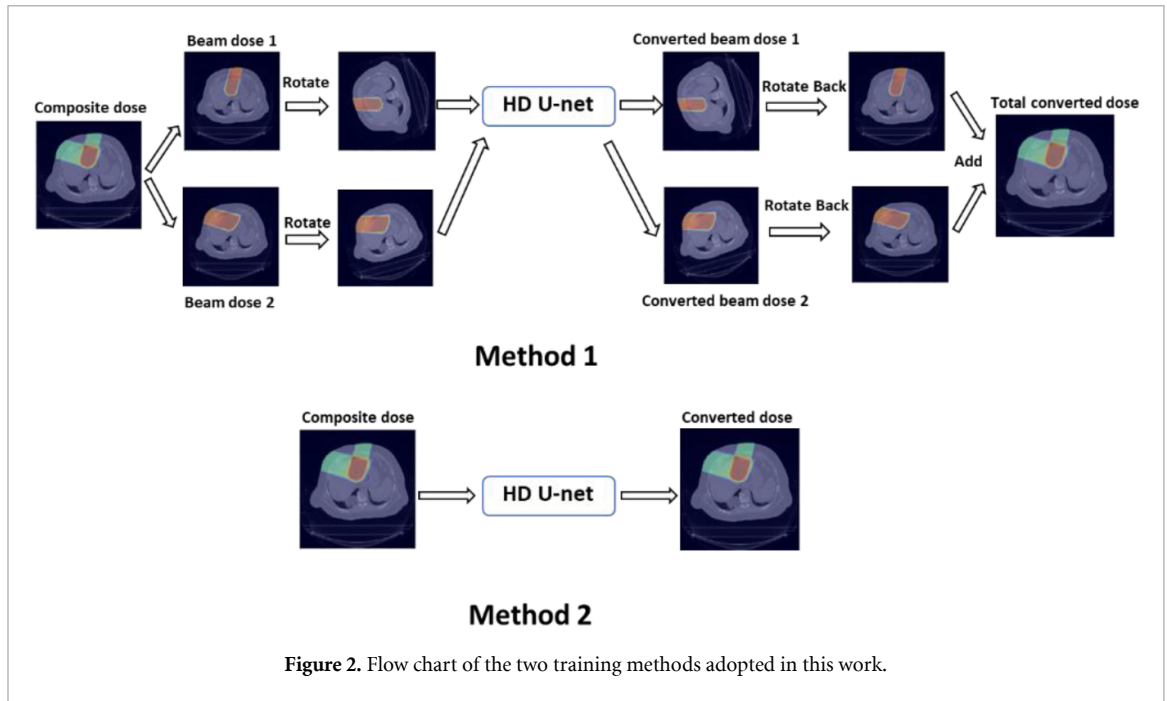
	Head and neck	Liver	Lung	Prostate	Total
Number of patients	90	93	32	75	290
Training & validation	72	75	26	61	234
Testing	18	18	6	14	56
Number of beams	720	215	88	260	1283

2.2. Patient database

The patient database used in this work consists of 90 head and neck, 93 liver, 75 prostate and 32 lung patients treated using the double scattering method at the Massachusetts General Hospital (MGH) proton therapy center. For each tumor site, about 20% of the patients were randomly selected as the testing set, and the rest were used for training and validation. The specific numbers of test patients and training & validation patients for each tumor site are presented in table 1. It should be noted that some treatment plans were only used for research or as alternate plan options or beam arrangements and the corresponding treatment fields were not actually delivered to patients. For each patient, the proton PB dose was calculated using an algorithm implemented on the XiO treatment planning system (by Computerized Medical Systems Inc. now by ELEKTA), and the MC dose was obtained by using TOPAS version 3.0.1 [35], which is based on Geant4.10.3 [36]. All the dose data and their corresponding CT images were resampled to have a voxel resolution of $2 \times 2 \times 2.5$ mm^3.

2.3. Training experiment design

The patient data of the four tumor sites used in this work have different beam configurations, so the model must be able to maintain high performance across different beam settings. To address this issue, for each patient, we rotated the dose distribution of each beam and the CT image to the same default angle and then used them as inputs for the model. As the model output, the converted dose distribution of each beam is rotated back and added to the dose distributions of other beams to obtain the total converted dose. In this work, the default angle is set to a 270° gantry angle and a 0° couch angle. We believe this method allows the model to better learn the upstream and downstream characteristics when the proton beam passes through the patient's body and deposits energy along the path, and therefore helps the model to better learn the accurate mapping between the PB dose and the MC dose in relation to tissue inhomogeneities. We also

Figure 2. Flow chart of the two training methods adopted in this work.

Figure 3. Schematic overview of the four experiments for each tumor site.

implemented and tested an alternative method that directly uses the composite dose distribution as the model input and output. In this article, we refer to the method that uses the beam dose as the model input and output as Method 1 and the method that uses the composite dose as Method 2. The schematic flow chart of the two methods is shown in figure 2.

Because the prescription dose varies from patient to patient and from tumor site to tumor site, all the dose and CT images are normalized before being input into the model. For both methods, the CT voxel values are normalized to be between 0 and 1, and the dose distribution voxel values are normalized by dividing the 95th percentile dose value of all voxels receiving dose values greater than 5% of the maximum PB dose. On the output side of the model, the converted beam or composite dose is multiplied by the same normalization factor.

To develop a general model that can convert PB dose distribution to MC dose distribution for each tumor site, in addition to using site-specific data for training, we implemented joint training that used data from all sites for each method. For each tumor site, four different numerical experiments were carried out to investigate which model had the best performance. Take head and neck patients as an example: for Experiment 1, the beam dose distributions (Method 1) of head and neck patients were used as the model input, and for Experiment 2, the beam dose distributions (Method 1) of all four tumor sites were used as the model input. For Experiment 3, the composite dose distributions (Method 2) of head and neck patients were used as the model input, and for Experiment 4, the composite dose distributions (Method 2) of all four tumor sites were used as the model input. The details of the four experiments are summarized in figure 3. In total, 10 models were trained and evaluated (two general models trained with all-sites data and eight models trained with site-specific data). For each tumor site, the models of all four experiments were tested on the same test dataset.

Table 2. Gamma passing rate (1%/1 mm) and MSE results of all the experiments for head and neck and liver test patients (above a threshold of 10% of the maximum MC dose).

	Head and neck		Liver	
	Gamma passing rate	MSE(Gy)	Gamma passing rate	MSE(Gy)
PB dose	(73.3 ± 6.3) %	(4.89 ± 3.39)	(79.2 ± 5.1) %	(1.72 ± 0.65)
Experiment 1	(90.4 ± 2.8) %	(1.28 ± 0.92)	(91.9 ± 4.8) %	(0.43 ± 0.31)
Experiment 2	(92.8 ± 2.9) %	(1.14 ± 0.82)	(92.7 ± 2.9) %	(0.31 ± 0.15)
Experiment 3	(77.9 ± 9.6) %	(2.64 ± 1.84)	(89.5 ± 4.9) %	(0.66 ± 0.43)
Experiment 4	(83.0 ± 5.1) %	(2.26 ± 1.58)	(88.9 ± 5.1) %	(0.69 ± 0.42)

Table 3. Gamma passing rate (1%/1 mm) and MSE results of all the experiments for lung and prostate test patients (above a threshold of 10% of the maximum MC dose).

	Lung		Prostate	
	Gamma passing rate	MSE(Gy)	Gamma passing rate	MSE(Gy)
PB dose	(65.4 ± 5.3) %	(3.17 ± 1.80)	(73.3 ± 2.7) %	(2.04 ± 1.10)
Experiment 1	(86.3 ± 5.9) %	(0.66 ± 0.62)	(99.5 ± 0.3) %	(0.13 ± 0.08)
Experiment 2	(89.7 ± 3.8) %	(0.48 ± 0.39)	(99.6 ± 0.3) %	(0.12 ± 0.08)
Experiment 3	(70.3 ± 5.3) %	(2.12 ± 1.66)	(98.6 ± 0.9) %	(0.23 ± 0.16)
Experiment 4	(75.5 ± 5.2) %	(1.60 ± 0.97)	(97.2 ± 1.4) %	(0.30 ± 0.16)

2.4. Model generalizability analysis

To investigate the model generalizability, we performed Experiment 5 based on Experiment 2. In this experiment, for a tumor site A, we first implemented joint training using the beam dose distributions of the other three tumor sites B, C and D as the model input. The model was trained for 200 epochs, and the learning rate was adjusted to minimize the validation loss as a function of epochs. Then, the trained model was fine-tuned by using the beam dose of tumor site A for 100 epochs. The learning rate for fine-tuning was set to 1×10^{-4}. After joint training and fine-tuning, the model was tested on the test dataset for tumor site A, same as in the other experiments.

To further test the model generalizability to see if it can be adapted to a new dataset from another hospital, we performed Experiment 6 based on Experiment 2. The new dataset we used in Experiment 6 includes 27 patients with non-small cell lung cancer (NSCLC) from a hospital in China. We randomly selected 22 of these patients as the training set to fine-tune the model and used the remaining five patients as the testing set. For each patient, RayStation 9A was used to develop intensity modulated proton therapy (IMPT) treatment plans, and to calculate the PB and MC dose distributions. We set the total number of epochs to 100 and the learning rate to 1×10^{-4}.

2.5. Model training details

The mean squared error (MSE) between the converted dose distribution and the MC dose distribution was used as the loss function for training each model. The learning rate of each model was adjusted to minimize the validation loss as a function of epochs. During each training iteration, a patch of size $128 \times 128 \times 16$ was randomly selected from the patient volume. This patch-wise training method, similar to data augmentation, can reduce overfitting [27]. The Adam algorithm [37] was selected as the optimizer to minimize the loss function. All the deep learning models were built and implemented in Keras with Tensorflow [38] as the back end. Each model was trained for 200 epochs on one NVIDIA Tesla V100 GPU card with 32 GB RAM.

3. Results

3.1. Gamma index analysis and MSE results

To assess the accuracy of the PB-to-MC converted dose, we computed the MSE between the converted dose and the MC dose above a threshold of 10% of the maximum MC dose in the test dataset. We also computed the 3D gamma index (γ), which can evaluate the dosimetric accuracy of voxels by combining the distance difference and dose difference metrics [39]. The gamma passing rate with 1%/1 mm criteria and the MSE results of all experiments for head and neck and liver test patients are presented in table 2; the corresponding results for lung and prostate test patients are presented in table 3. Both tables include the PB dose metrics for comparison. For every tumor site, the gamma passing rate and MSE results are substantially better for all the experiments than for the same metrics of the PB dose. Method 1 (Experiments 1 and 2), which uses the beam dose as the model input, clearly outperformed Method 2 (Experiments 3 and 4), which directly uses the

Table 4. Average single-field MC simulation, PB dose calculation, and dose conversion time results for all four tumor sites used in this work (all in seconds).

	Head and neck	Liver	Lung	Prostate
MC simulation	14 487 ± 12 819	28 652 ± 20 041	27 372 ± 28 206	77 530 ± 25 882
PB dose calculation	123 ± 91	253 ± 199	303 ± 390	172 ± 59
Dose conversion	3.3 ± 0.8	2.2 ± 0.7	2.1 ± 0.2	2.2 ± 1.0

composite dose as the model input. Experiment 2, which used the beam dose distribution data from all sites as the model input, clearly outperformed Experiment 1, which used the beam dose distributions of site-specific data as the model input, except in the case of prostate test patients. In general, Experiment 2, which used the beam dose distribution data from all sites as the model input, had the best performance across all four types of test tumor sites. The average gamma passing rate between the converted and the MC dose distributions for head and neck, liver, lung and prostate test patients was 92.8%, 92.7%, 89.7% and 99.6%, respectively.

3.2. Dosimetry analysis
Dose color washes and dose volume histograms (DVHs) curves of the converted and the MC dose distributions were compared to test the performance of the model trained in Experiment 2. Figures 4 and 6 show dose color washes of one example patient from the head and neck and prostate test sets, respectively. In each of these two figures, one axial slice of the PB dose distribution, the converted dose distribution, the absolute differences between the PB and MC dose distribution and the absolute differences between the converted and MC dose distribution are illustrated in sub-figures (a), (b), (c) and (d), respectively. Figures 5 and 7 show DVHs curves of the same example head and neck and prostate test patient, respectively. Upon visual inspection of the dose color washes, the converted dose distributions are almost identical to the MC dose distributions for each example patient. The DVH curves of the converted dose are found to be very close to those of the MC dose for both targets and OARs, but an obvious gap can be observed between the DVH curves of PB and MC doses. The results of the other test patients and other tumor sites are not presented here because of space limits in the manuscript, but the behaviors are similar for all of them.

We also compared the voxel-based dose difference between the PB dose and the MC dose and between the converted dose and the MC dose for all test patients. If any voxel in the PB or the converted dose distribution received more than 3% of the maximum MC dose and its absolute dose difference with the MC dose distribution was larger than 3% of the maximum MC dose, then that voxel was used for evaluation. Dose difference histograms for 18 head and neck, 18 liver, 6 lung and 14 prostate test patients are shown in figure 8. We can clearly see that, for each type of test patient, the number of voxels with dose differences larger than 3% of the maximum MC dose is substantially lower for the converted dose than for the PB dose.

3.3. Model efficiency
Table 4 shows the average time required for MC simulation, PB dose calculation and dose conversion of a single field for each tumor site. The MC simulations were implemented in TOPAS, and the PB doses were calculated using an algorithm implemented on the XiO system, both calculations were performed at the MGH proton therapy center. The conversion processes were executed on a NVIDIA Tesla V100 card with 16 GB dedicated RAM. The total time to obtain the converted MC dose is roughly the sum of the PB dose calculation time and the conversion time. Compared with the MC simulation time, the average total time to obtain the converted MC dose is over 89 times shorter.

3.4. Model generalizability
The gamma passing rate with 1%/1 mm criteria and MSE results of Experiments 2 and 5 for the four types of test patients are presented in table 5, both using MC results as reference. In general, Experiments 2 and 5 had similar performance. The models trained in Experiment 5 showed good performance for each tumor site, with the average gamma passing rate between the converted dose and the MC dose above 90%. These results show that the models that use the beam dose distributions of three tumor sites as input can be easily adapted to a new tumor site, which suggests that joint training using beam dose distributions of multiple tumor sites as the model input has good generalizability. This supports our assumption that the model that uses the beam dose distributions of all four tumor sites as input in Experiment 2 can be adapted to new datasets in clinical practice.

In Experiment 6, we first fine-tuned the model trained in Experiment 2 using the training set of a new NSCLC dataset and then evaluated the model on the test set. For the patients in the testing set, the mean
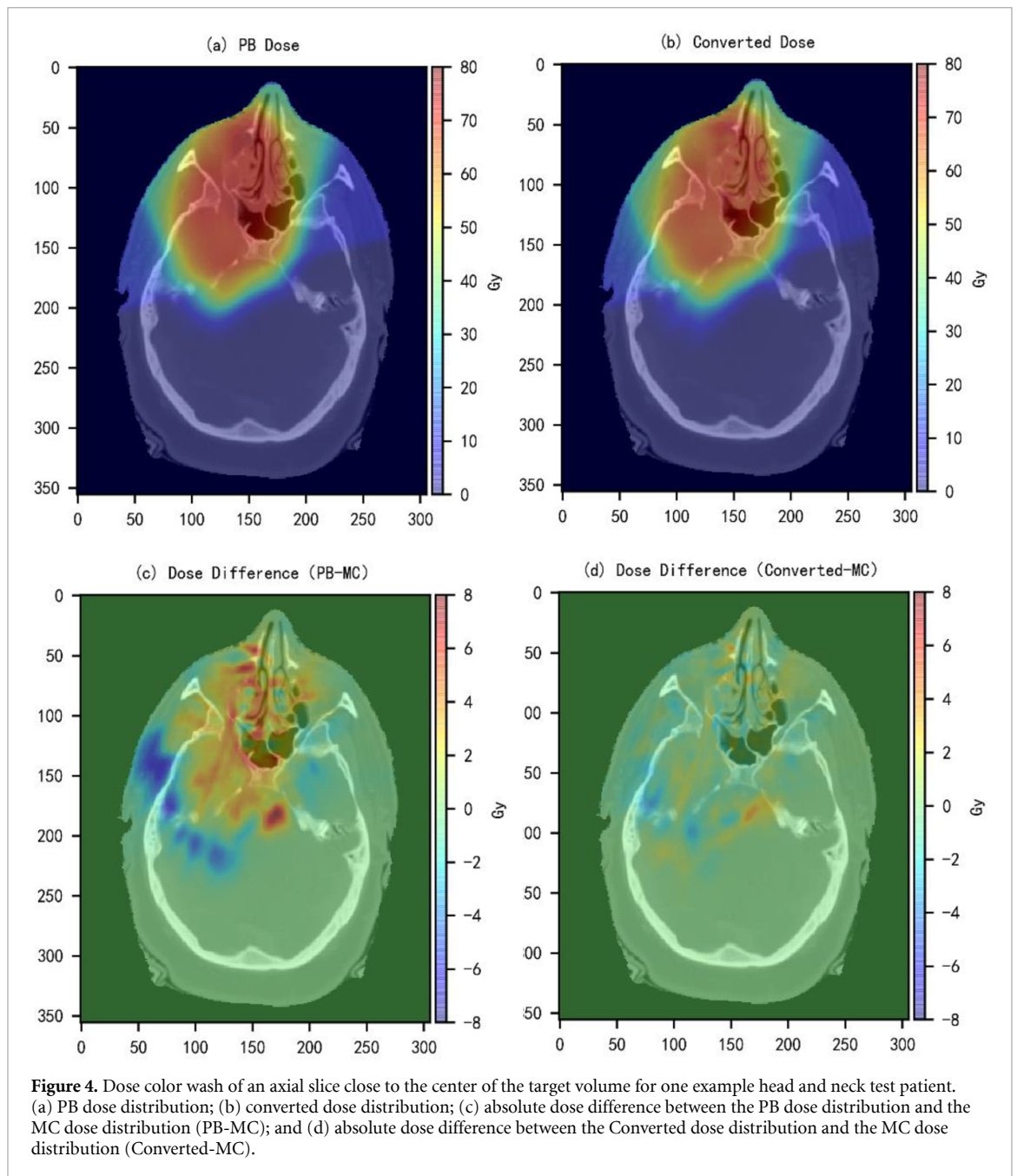
Figure 4. Dose color wash of an axial slice close to the center of the target volume for one example head and neck test patient. (a) PB dose distribution; (b) converted dose distribution; (c) absolute dose difference between the PB dose distribution and the MC dose distribution (PB-MC); and (d) absolute dose difference between the Converted dose distribution and the MC dose distribution (Converted-MC).

Table 5. Gamma passing rate (1%/1 mm) and MSE results of Experiments 2 and 5 for all test patients (above a threshold of 10% of the maximum MC dose).

	Experiment 2		Experiment 5	
	Gamma passing rate	MSE (Gy)	Gamma passing rate	MSE (Gy)
Head and Neck	(92.8 ± 2.9) %	(1.14 ± 0.82)	(91.9 ± 3.1) %	(1.21 ± 0.82)
Liver	(92.7 ± 2.9) %	(0.31 ± 0.15)	(93.7 ± 2.8) %	(0.31 ± 0.17)
Lung	(89.7 ± 3.8) %	(0.48 ± 0.39)	(90.3 ± 4.0) %	(0.46 ± 0.38)
Prostate	(99.6 ± 0.3) %	(0.12 ± 0.08)	(99.6 ± 0.3) %	(0.11 ± 0.07)

gamma passing rate (1%/1 mm) between PB and MC dose distributions was 75.5% $\pm$ 8.9% (standard deviation). We found that the converted dose distributions showed substantial improvement over the PB dose distributions and had strong agreement with the MC dose distributions. The mean gamma passing rate between the converted dose and MC dose (1%/1 mm) was 91.8% $\pm$ 4.8% (standard deviation). This result proves that the model trained in Experiment 2 can be adapted to a new dataset from different hospitals through simple fine-tuning.
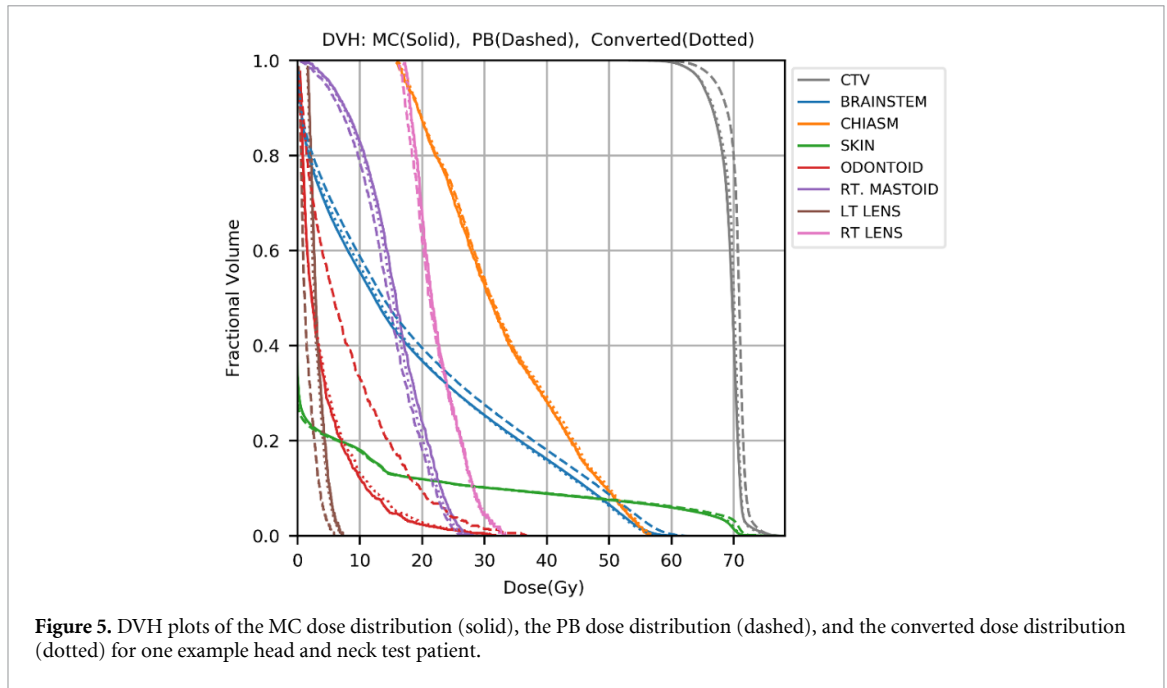
Figure 5. DVH plots of the MC dose distribution (solid), the PB dose distribution (dashed), and the converted dose distribution (dotted) for one example head and neck test patient.
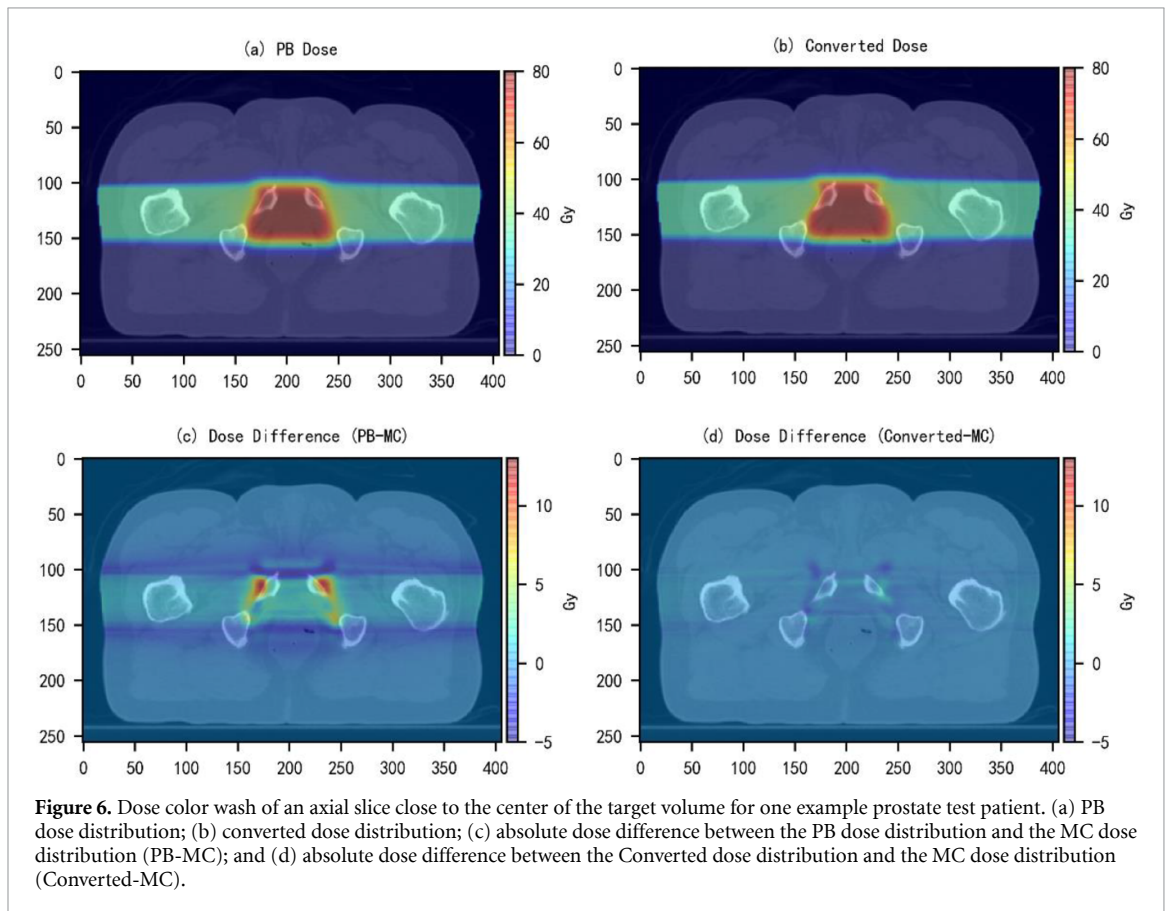

Figure 6. Dose color wash of an axial slice close to the center of the target volume for one example prostate test patient. (a) PB dose distribution; (b) converted dose distribution; (c) absolute dose difference between the PB dose distribution and the MC dose distribution (PB-MC); and (d) absolute dose difference between the Converted dose distribution and the MC dose distribution (Converted-MC).

4. Discussion & conclusions

In this work, we present a novel method that uses deep learning methods to convert PB dose distributions to MC dose distributions for different tumor sites. We performed four different experiments to compare different combinations of training datasets, and we found that joint training using the beam dose distributions data from all sites as the model input had the best performance. The converted dose distributions showed substantial improvement over the PB dose distributions in all evaluation criteria that we considered: gamma index, MSE, DVH and dose difference histogram. The developed model can learn
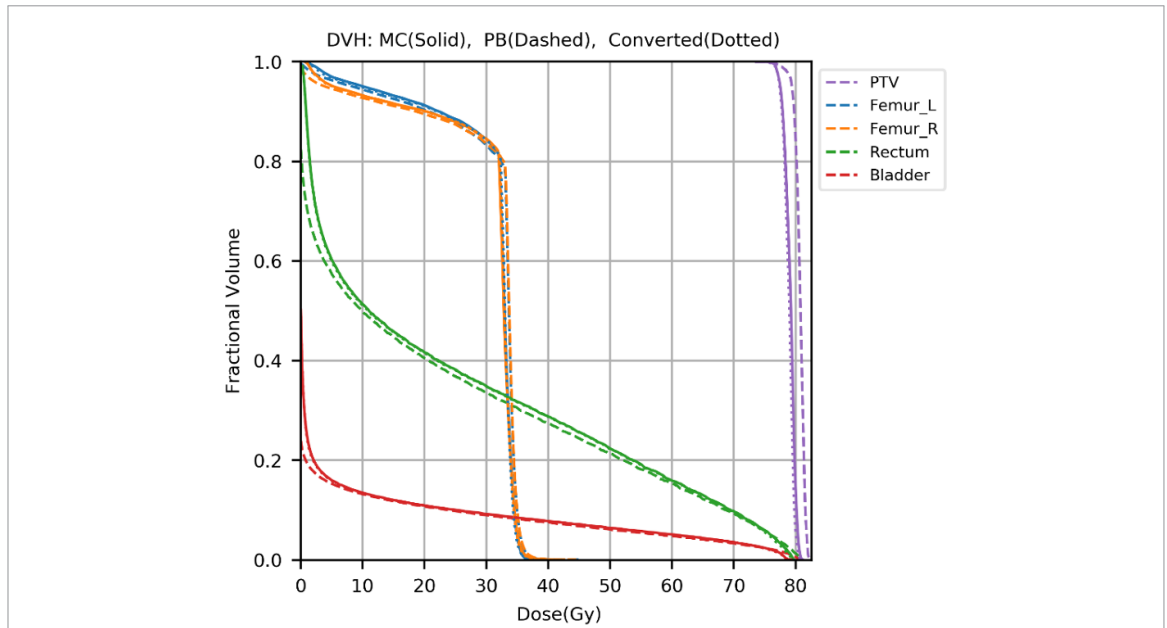
Figure 7. DVH plots of the MC dose distribution (solid), the PB dose distribution (dashed), and the converted dose distribution (dotted) for one example prostate test patient.
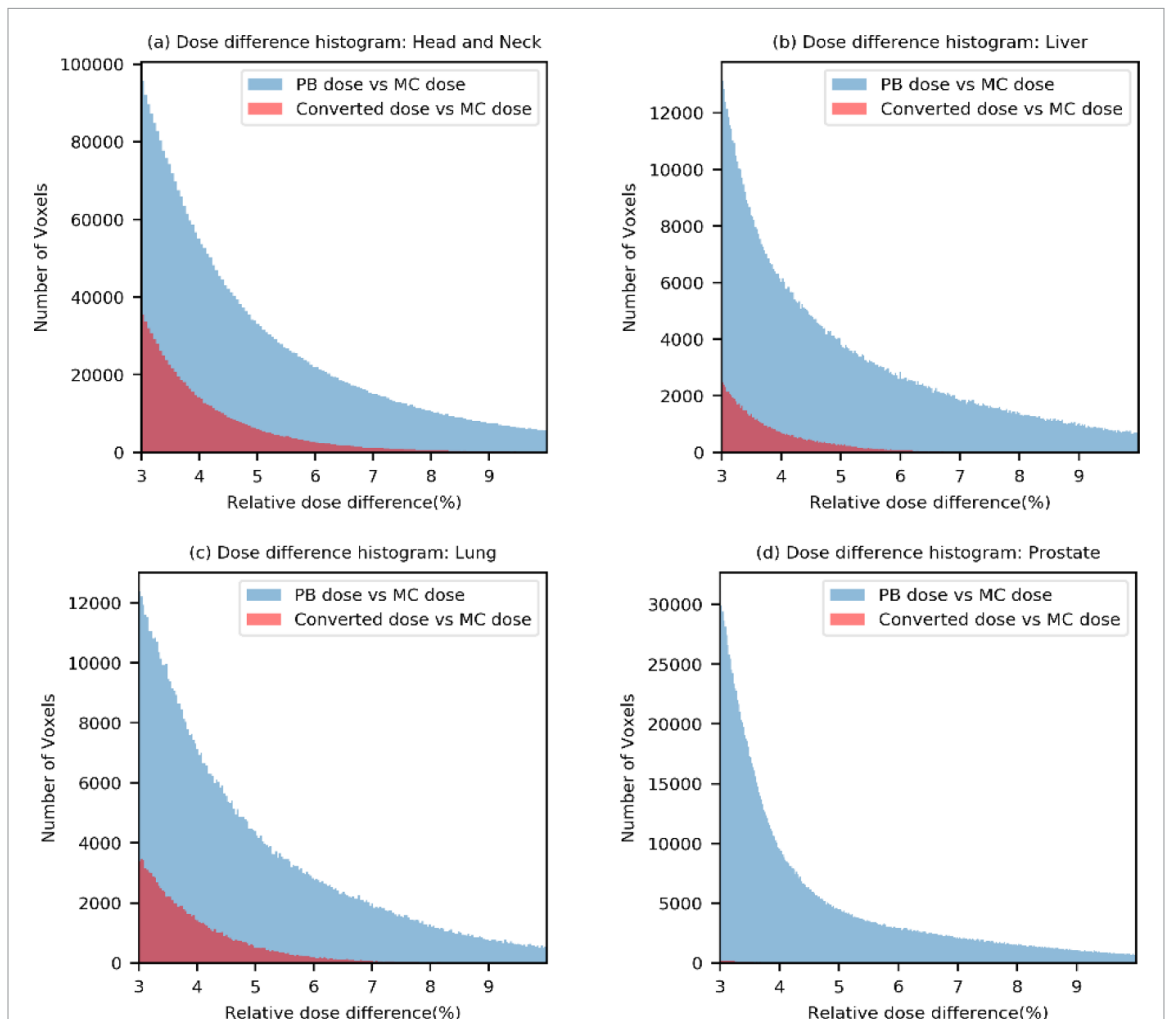

Figure 8. Dose difference histograms for all test patients. (a) Head and neck test patients; (b) liver test patients; (c) lung test patients; and (d) prostate test patients.

from a heterogeneous database that includes beam dose distributions from four tumor sites with different beam configurations and maintain high performance for each tumor site. Using beam dose distributions as the model input (Experiments 1 and 2) yielded substantially better model performance than directly using composite dose distributions as the model input (Experiments 3 and 4). These results suggest that rotating the normalized beam dose and the corresponding CT to the same angle and using them as the model input helps the model to better learn the difference between the PB and the MC dose distributions in relation to tissue heterogeneity and thus improves the model's performance. We also noticed that the converted dose of prostate test patients had the best performance, and the converted dose of lung test patients had the worst, in terms of gamma passing rate and MSE results. One possible reason is that all the prostate patient data used in this work have similar beam settings. The beam angles for prostate patients are set to be either 90° or 270° gantry angle, which we believe makes the prostate patients' data size needed for the model to achieve high performance smaller than other tumor sites. The gamma passing rate and MSE results of Experiments 1 and 2 (see table 3) show that, for prostate patients, adding the beam dose of other tumor sites for training did not improve the model's performance. This indicates that, in this work, the prostate patient data size is big enough for the model to learn the mapping between the PB and the MC dose distributions. For the other three treatment sites—especially for lung patients which have the smallest data size, adding the training data of the other tumor sites improved the model's performance. Thus, joint training using beam dose distributions from all the tumor sites as the model input can improve the model's performance when the beam configurations are heterogeneous and the data size is relatively small.

The average time needed to convert a single field for all four types of patients in this work is less than 4 s. To date, the fastest proton PB dose calculations are in the sub-second range. Da Silva *et al* achieved PB dose calculations with a double Gaussian kernel in 0.22 s [40]. Because the number of fields used in proton therapy is usually no more than 4, the total time required to obtain the converted dose can potentially be kept within half a minute. This will allow the implementation of online adaption of MC calculations to further improve the treatment quality for inter-fractional geometry changes [16]. We have not yet dedicated any efforts to improving the model's efficiency, which can be easily achieved through methods like model compression; we will explore such methods in our future work. In addition, by accurately and efficiently generating converted MC doses, the proposed model could also be used as a plan evaluation for physicians to decide whether to implement MC simulation or not.

We have also demonstrated generalizability of the model. In Experiment 5, we showed that the trained model can be easily used for another tumor site through transfer learning. With Experiment 6, we also showed that the trained model can be deployed to different hospitals through transfer learning, even when using a different treatment plan delivery, i.e. IMPT vs. double scattering

The main patient database used in this work comprises 90 head and neck, 93 liver, 75 prostate and 32 lung patients treated with double scattering methods at MGH proton center. Currently, IMPT which uses scanned proton pencils to shape the delivered dose distributions is adopted by most new proton therapy facilities. We plan to extend this work by applying the developed model to more IMPT plans of different tumor sites and testing the performance.

In conclusion, we used a deep learning-based approach to improve the accuracy of proton therapy by converting PB dose distributions to MC-equivalent dose distributions. We carried out four different experiments to explore various combinations of composite versus single beam dose distributions and all-site data versus site-specific data. The results showed that joint training using beam dose distributions of all four tumor sites data available had the best performance. The developed model used in this experiment can accurately and efficiently convert PB dose distributions to MC dose distributions for different tumor sites. The trained model was proved to have good generalizability and can be readily adapted to new datasets for different tumor sites and from different hospitals through transfer learning. This model can be added as a plug-in to the clinical workflow of proton therapy treatment planning to improve the accuracy of proton dose calculation.

Acknowledgments

Conflict of interest statement

None.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Chao Wu ⓘ https://orcid.org/0000-0003-4988-6255
Dan Nguyen ⓘ https://orcid.org/0000-0002-9590-0655
Jan Schuemann ⓘ https://orcid.org/0000-0002-7554-8818

References

[1] Paganetti H 2012 Range uncertainties in proton therapy and the role of Monte Carlo simulations *Phys. Med. Biol.* **57** R99-117
[2] Jia X, Schümann J, Paganetti H and Jiang S B 2012 GPU-based fast Monte Carlo dose calculation for proton therapy *Phys. Med. Biol.* **57** 7783–97
[3] Hong L, Goitein M, Bucciolini M, Comiskey R, Gottschalk B, Rosenthal S, Serago C and Urie M 1996 A pencil beam algorithm for proton dose calculations *Phys. Med. Biol.* **41** 1305–30
[4] Schaffner B, Pedroni E and Lomax A 1999 Dose calculation models for proton treatment planning using a dynamic beam delivery system: an attempt to include density heterogeneity effects in the analytical dose calculation *Phys. Med. Biol.* **44** 27–41
[5] Maes D, Saini J, Zeng J, Rengan R, Wong T and Bowen S R 2018 Advanced proton beam dosimetry part II: Monte Carlo vs. pencil beam-based planning for lung cancer *Transl. Lung Cancer Res.* **7** 114–21
[6] Huang S *et al* 2018 Validation and application of a fast Monte Carlo algorithm for assessing the clinical impact of approximations in analytical dose calculations for pencil beam scanning proton therapy *Med. Phys.* **45** 5631–42
[7] Teoh S, Fiorini F, George B 1, Ka V and Van den Heuvel F 2019 Is an analytical dose engine sufficient for intensity modulate proton therapy in lung cancer? *Br. J. Radiol.* **93**
[8] Schuemann J, Giantsoudi D, Grassberger C, Moteabbed M, Min C H and Paganetti H 2015 Assessing the clinical impact of approximations in analytical dose calculations for proton therapy *Int. J. Radiat. Oncol. Biol. Phys.* **92** 1157–64
[9] Liang X, Li Z, Zheng D, Bradley J A, Rutenberg M and Mendenhall N 2019 A comprehensive dosimetric study of Monte Carlo and pencil-beam algorithms on intensity-modulated proton therapy for breast cancer *J. Appl. Clin. Med. Phys.* **20** 128–36
[10] Paganetti H 2014 Monte Carlo simulations will change the way we treat patients with proton beams today *Br. J. Radiol.* **87** 20140293
[11] Taylor P A, Kry S F and Followill D S 2017 Pencil beam algorithms are unsuitable for proton dose calculations in lung *Int. J. Radiat. Oncol. Biol. Phys.* **99** 750–6
[12] Bueno M, Paganetti H, Duch M A and Schuemann J 2013 An algorithm to assess the need for clinical Monte Carlo dose calculation for small proton therapy fields based on quantification of tissue heterogeneity *Med. Phys.* **40** 081704
[13] Grassberger C, Daartz J, Dowdell S, Ruggieri T, Sharp G and Paganetti H 2014 Quantification of proton dose calculation accuracy in the lung *Int. J. Radiat. Oncol. Biol. Phys.* **89** 424–30
[14] Saini J, Maes D, Egan A, Bowen S R, St James S, Janson M, Wong T and Bloch C 2017 Dosimetric evaluation of a commercial proton spot scanning Monte-Carlo dose algorithm: comparisons against measurements and simulations *Phys. Med. Biol.* **62** 7659–81
[15] Yepes P, Adair A, Grosshans D, Mirkovic D, Poenisch F, Titt U, Wang Q and Mohan R 2018 Comparison of Monte Carlo and analytical dose computations for intensity modulated proton therapy *Phys. Med. Biol.* **63** 045003
[16] Botas P, Kim J, Winey B and Paganetti H 2018 Online adaption approaches for intensity modulated proton therapy for head and neck patients based on cone beam CTs and Monte Carlo simulations *Phys. Med. Biol.* **64**
[17] Qin N, Pinto M, Tian Z, Dedes G, Pompos A, Jiang S B, Parodi K and Jia X 2017 Initial development of goCMC: a GPU-oriented fast cross-platform Monte Carlo engine for carbon ion therapy *Phys. Med. Biol.* **62** 3682–99
[18] Qin N, Botas P, Giantsoudi D, Schuemann J, Tian Z, Jiang S B, Paganetti H and Jia X 2016 Recent developments and comprehensive evaluations of a GPU-based Monte Carlo package for proton therapy *Phys. Med. Biol.* **61** 7347–62
[19] Yepes P, Randeniya S, Taddei P J and Newhauser W D 2009 Monte Carlo fast dose calculator for proton radiotherapy: application to a voxelized geometry representing a patient with prostate cancer *Phys. Med. Biol.* **54** N21–8
[20] Jia X, Ziegenhein P and Jiang S B 2014 GPU-based high-performance computing for radiation therapy *Phys. Med. Biol.* **59** R151
[21] Ma J, Beltran C, Seum Wan Chan T, Seung H and Herman M G 2014 GPU-accelerated and Monte Carlo-based intensity modulated proton therapy optimization system *Med. Phys.* **41** 121707
[22] Wan Chan Tseung H, Ma J and Beltran C 2015 A fast GPU-based Monte Carlo simulation of proton transport with detailed modeling of nonelastic interactions *Med. Phys.* **42** 2967–78
[23] Tian Z, Shi F, Folkerts M, Qin N, S B J and Jia X 2015 A GPU OpenCL based cross-platform Monte Carlo dose calculation engine (goMC) *Phys. Med. Biol.* **60** 7419–35
[24] Souris K, J A L and Sterpin E 2016 Fast multipurpose Monte Carlo simulation for proton therapy using multi-and many-core CPU architectures *Med. Phys.* **43** 1700–12
[25] Mcintosh C and Purdie T 2017 Voxel-based dose prediction with multi-patient atlas selection for automated radiotherapy treatment planning *Phys. Med. Biol.* **62** 415–31
[26] Nguyen D, Long T, Jia X, Lu W, Gu X, Iqbal Z and Jiang S 2019 Dose prediction with U-net: a feasibility study for predicting dose distributions from contours using deep learning on prostate IMRT patients *Sci. Rep.* **9** 1076
[27] Nguyen D, Jia X, Sher D, Lin M-H, Iqbal Z, Liu H and Jiang S 2019 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture *Phys. Med. Biol.* **64**
[28] Chen X, Men K, Li Y, Yi J and Dai J 2019 A feasibility study on an automated method to generate patient-specific dose distributions for radiotherapy using deep learning *Med. Phys.* **46** 56–64
[29] Fan J, Wang J, Chen Z, Hu C, Zhang Z and Hu W 2019 Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique *Med. Phys.* **46** 370–81
[30] Kearney V, Chan J W and Haaf S Descovich M and Solberg TD 2018 DoseNet: a volumetric dose prediction algorithm using 3D fully- convolutional neural networks *Phys. Med. Biol.* **63**

[31] Barragán-Montero A M, Nguyen D, Lu W, Lin M H, Norouzi-Kandalan R, Geets X, Sterpin E and Jiang S 2019 Three-dimensional dose prediction for lung IMRT patients with deep neural networks: robust learning from heterogeneous beam configurations *Med. Phys.* **46** 3679–91

[32] Xing Y, Nguyen D, Lu W, Yang M and Jiang S 2019 Technical note: a feasibility study on deep learning-based radiotherapy dose calculation *Med. Phys.* **47** 753–8

[33] Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *Lect. Notes Comput. Sci.* **9351** 234–41

[34] Huang G, *et al* 2017 Densely connected convolutional networks *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*

[35] Perl J, Shin J, Schumann J, Faddegon B and Paganetti H 2012 TOPAS: an innovative proton Monte Carlo platform for research and clinical applications *Med. Phys.* **39** 6818–37

[36] Agostinelli S *et al* 2003 GEANT4 - a simulation toolkit *Nuclear Instrum. Methods Phys. Res.* A **506** 250–303

[37] Kingma D and Ba J 2015 Adam: A Method for Stochastic Optimization. *Proc. Int. Conf. on Learning Representations (San Diego, 2015)*

[38] Abadi M, *et al* 2016 Tensorflow: large-scale machine learning on heterogeneous distributed systems (arXiv: 1603.04467)

[39] Wendling M, Zijp L J, McDermott L N, Smit E J, Sonke J J, Mijnheer B J and van Herk M 2007 A fast algorithm for gamma evaluation in 3D *Med. Phys.* **34** 1647–54

[40] da Silva J, Ansorge R and Jena R 2015 Fast pencil beam dose calculation for proton therapy using a double-Gaussian beam model *Front. Oncol.* **5** 281