Physics in Medicine & Biology

PAPER • **OPEN ACCESS**

Prompt gamma emission prediction using a long short-term memory network

To cite this article: Fan Xiao *et al* 2024 *Phys. Med. Biol.* **69** 235003

IOP Publishing Phys. Med. Biol. 69 (2024) 235003 https://doi.org/10.1088/1361-6560/ad8e2a

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

OPEN ACCESS

RECEIVED
1 July 2024

REVISED
11 October 2024

ACCEPTED FOR PUBLICATION
1 November 2024

PUBLISHED
20 November 2024

PAPER

Prompt gamma emission prediction using a long short-term memory network

Fan Xiao[1], Domagoj Radonic[1,2], Michael Kriechbaum[2], Niklas Wahl[3,4], Ahmad Neishabouri[4,5], Nikolaos Delopoulos[1], Katia Parodi[2], Stefanie Corradini[1], Claus Belka[1,6,7], Christopher Kurz[1], Guillaume Landry[1] and George Dedes[2,*]

[1] Department of Radiation Oncology, LMU University Hospital, LMU Munich, Munich, Germany
[2] Department of Medical Physics, LMU Munich, Munich, Germany
[3] Department of Medical Physics in Radiation Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany
[4] National Center for Radiation Oncology (NCRO), Heidelberg Institute for Radiation Oncology (HIRO), Heidelberg, Germany
[5] Clinical Cooperation Unit Radiation Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany
[6] German Cancer Consortium (DKTK), partner site Munich, a partnership between DKFZ and LMU University Hospital, Munich, Germany
[7] Bavarian Cancer Research Center (BZKF), Munich, Germany
[*] Author to whom any correspondence should be addressed.

E-mail: g.dedes@physik.uni-muenchen.de

Keywords: deep learning, proton therapy, prompt gamma, range verification, LSTM

Abstract

Objective: To present a long short-term memory (LSTM)-based prompt gamma (PG) emission prediction method for proton therapy. *Approach*: Computed tomography (CT) scans of 33 patients with a prostate tumor were included in the dataset. A set of 10^7 histories proton pencil beam (PB)s was generated for Monte Carlo (MC) dose and PG simulation. For training (20 patients) and validation (3 patients), over 6000 PBs at 150, 175 and 200 MeV were simulated. 3D relative stopping power (RSP), PG and dose cuboids that included the PB were extracted. Three models were trained, validated and tested based on an LSTM-based network: (1) input RSP and output PG, (2) input RSP with dose and output PG (single-energy), and (3) input RSP/dose and output PG (multi-energy). 540 PBs at each of the four energy levels (150, 175, 200, and 125–210 MeV) were simulated across 10 patients to test the three models. The gamma passing rate (2%/2 mm) and PG range shift were evaluated and compared among the three models. *Results*: The model with input RSP/dose and output PG (multi-energy) showed the best performance in terms of gamma passing rate and range shift metrics. Its mean gamma passing rate of testing PBs of 125–210 MeV was 98.5% and the worst case was 92.8%. Its mean absolute range shift between predicted and MC PGs was 0.15 mm, where the maximum shift was 1.1 mm. The prediction time of our models was within 130 ms per PB. *Significance*: We developed a sub-second LSTM-based PG emission prediction method. Its accuracy in prostate patients has been confirmed across an extensive range of proton energies.

1. Introduction

Proton radiotherapy is a well established cancer therapy method, which can achieve high-dose conformity and significant sparing of healthy tissues because of its unique physical characteristics. However, due to its steep Bragg peak dose deposition and distal dose fall-off, proton therapy is highly sensitive to proton range uncertainties that can cause an overshoot or undershoot of the tumor, resulting in non-optimal dose coverage of targets and/or overdose of healthy tissues (Unkelbach *et al* 2009, Paganetti 2012). To help minimize such proton range delivery errors during treatment, online range verification would be beneficial (Knopf and Lomax 2013).

During proton irradiation, positron-emitting isotopes and prompt gamma (PG) rays are generated by proton-nuclear interactions, which have a strong correlation with proton dose distributions (Litzenberg *et al*

1999, Min *et al* 2006). Therefore, positron emission tomography (PET) and PG imaging have been widely considered for proton range verification (Moteabbed *et al* 2011). For online scenarios, in-beam PET systems can provide positron-emitting isotope images with a given time delay, due to the isotopes' half-life, and estimate beam ranges during the course of treatment (Dendooven *et al* 2015, Purushothaman *et al* 2023). PG imaging systems allow real-time PG signal range monitoring due to the PG emissions occurring within nanoseconds after proton-nuclear interactions (Knopf *et al* 2009, Smeets *et al* 2012, Xie *et al* 2017, Berthold *et al* 2021).

Although PG distributions are highly correlated to proton dose distributions, the comparison between the measured PG distribution and the therapeutic dose distribution is not direct because the PG fall-off does not explicitly match the dose fall-off (Janssen *et al* 2014, Schmid *et al* 2015, Tian *et al* 2018). An estimated PG distribution for comparison to PG measurements can be calculated from the treatment plan by Monte Carlo (MC) simulations (Choi *et al* 2020, Wrońska *et al* 2021). These methods are usually time-consuming due to the rarity of PG events that require high numbers of protons for simulation, which does not meet the requirements for online verification. While several analytical approximation methods (Sterpin *et al* 2015, Pinto *et al* 2020) and MC simulation using massively parallel central processing unit (CPU) architectures (Souris *et al* 2016) were developed to calculate the expected PG distributions in a few seconds for clinical routine, real-time and accurate PG calculation methods are still an open research topic. Particularly for next-generation online or even real-time adaptive proton therapy (Paganetti *et al* 2021), which aims for immediate treatment corrections during radiation delivery, PG calculation would benefit from MC accuracy and sub-second speed.

Recently, several studies have shown the potential of deep learning techniques to realize fast proton dose calculation with accuracy comparable to MC simulations (Javaid *et al* 2021, Neishabouri *et al* 2021, Wu *et al* 2021, Zhang *et al* 2021, Pastor-Serrano and Perkó 2022). These methods can be broadly divided into two categories: The first one uses computationally cheap physics quantities as input, e.g. high noise MC dose (Javaid *et al* 2021), pencil beam (PB) dose (Wu *et al* 2021), or analytically derived stopping power (Zhang *et al* 2021), and learns corrections that can be applied to the cheap inputs and lead to a highly accurate MC dose. The second category (Neishabouri *et al* 2021, Pastor-Serrano and Perkó 2022), based on sequential processing of single proton PBs in the beam's eye view (BEV), using long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997) or Transformer (Vaswani *et al* 2017) models, can achieve dose prediction without the auxiliary physical input. Furthermore, due to their models exclusively focusing on the PB area, they tend to be lightweight and exhibit fast inference speeds. Given the close correlation between dose and PG distributions along the beam path, and the successful use of recurrent neural network (RNN)-based networks, such as LSTM, in predicting dose distributions for single PBs in heterogeneous geometries, this study explores their potential for predicting PG distributions.

In this work, a sub-second PG emission prediction method for proton beam spots was developed based on a previous LSTM-based proton dose calculation model (Neishabouri *et al* 2021), and its accuracy in prostate cancer patients was tested. In addition, by incorporating dose information into our model, we explored the capacity of the PG prediction model to generalize the mapping to unseen proton energies.

2. Materials and methods

2.1. Proposed framework

Based on the characteristic that PG emissions occur in the proton beam path until the energy of primary protons decreases below the reaction threshold, the PG prediction task closely follows the 3D dose distribution pattern. When viewed in the BEV coordinate system, the 3D PG distribution can be seen as a spatiotemporal sequence of 2D slices. Given the initial energy and beam direction of a PB, the framework illustrated in figure 1 proposes two approaches to process these 2D sequences, to collectively predict the 3D PG distribution through the proton penetration path based on the patient's computed tomography (CT)-derived relative stopping power (RSP) to water.

For the model from RSP to PG (RP) illustrated in figure 1(a), the 3D RSP cuboid containing the PG region of a PB is firstly resampled from the patient's CT scan. After the BEV coordinate conversion, the 3D RSP cuboid is conceptualized as a succession of 2D slices progressing from upstream to downstream and each slice is flattened into a 1D sequence, which can be processed by the uni-directional LSTM unit. The internal hidden state and cell state generated by the LSTM unit from upstream can be used as the input information for the downstream sequence prediction and the output sequences are fed into a back-end fully-connected layer network to generate the predicted PG slices. In the end, the generated PG slices are reassembled to match the 3D PG cuboid dimensions and then compared with the MC PG ground truth.

For the mixed-input model from RSP and dose to PG (RDP) depicted in figure 1(b), inspired by the dose filtering method making full use of dose distribution for PG prediction (Pinto *et al* 2020), dose is utilized as

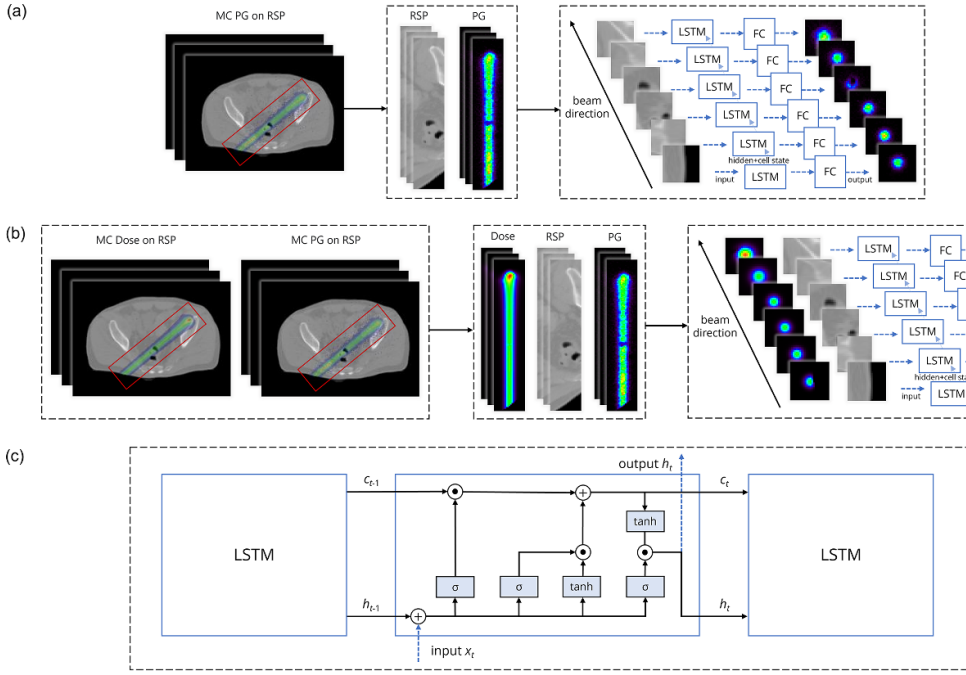

Figure 1. Training framework of this study: (a, left) The MC PG distribution of a PB with the outline of a 3D cuboid of interest. (a, middle) The 3D input RSP cuboid and ground truth PG cuboid after BEV resampling. (a, right) The LSTM model, which consists of a one-layer LSTM network and a two-layer fully-connected network (labeled as LSTM and FC in the figure), converts a sequence of RSP slices to a sequence of PG slices; (b, left) The MC dose and PG distributions of a PB with the outline of a 3D cuboid of interest. (b, middle and right) 3D RSP and dose cuboids are inputted into the model and the LSTM model converts a sequence of RSP and dose slices to PG slices. (c) The corresponding LSTM module structure: x_t is the 1D flattened input slice received at each step, while h_t represents the corresponding 1D output sequence passed to the fully-connected network. The cell states c_{t-1} and c_t (previous and current steps) and the hidden states h_{t-1} and h_t (previous and current steps) are for maintaining the LSTM's memory. The σ functions denote the input, forget, and output gates of the LSTM, whereas the tanh functions denote the updating of the cell state information.

the auxiliary information together with RSP. First, the PB dose cuboid of the same size as the corresponding RSP cuboid is extracted and resampled in the BEV coordinate system. Then, the slices of dose and RSP cuboid of the same depth in beam direction are concatenated and flattened into 1D input sequences, and the LSTM and the back-end fully-connected layer network process these combined sequences and output predicted PG.

The PG results from RP and RDP models were compared to evaluate the impact of the introduction of dose information, which would be routinely available in clinical workflows. In addition, due to the energy information carried by the dose distribution in the RDP model, the capability of this model to predict PG distributions across a range of unseen energies was investigated.

2.2. Patient datasets and MC simulations

A dataset consisting of 33 planning CT scans of prostate cancer patients treated at the Department of Radiation Oncology, LMU University Hospital, was collected for this study. All planning CT images were obtained from a single CT scanner (Aquilion LB, Canon Medical Systems, NL) and were resampled to a voxel size of 1.5 mm × 1.5 mm × 1.5 mm. All patients had no artificial implants and went through prior anonymization to adhere to the data protection law.

MC simulations in this work were performed using Geant4 version 11.00-patch-03 with the predefined QGSP_BIC_HP_EMZ Geant4 physics list. This physics list incorporates multiple scattering and ionization for charged particles, offering a suitable compromise between performance and speed. We utilized a CT scanner-specific lookup table, established in previous work (Schmid *et al* 2015), for converting CT numbers to mass density and material composition in the Geant4 MC simulations. The model input RSP maps were also generated by Geant4, using the voxelized geometry of the corresponding CT and setting the mean excitation energy of the water to 78 eV and the proton energy to 150 MeV. For the PG and dose simulation, considering that the aim of this study is to predict the PG distribution at the level of individual PBs, PG and dose for single PBs were scored simultaneously. For each PB, 10^7 initial protons, referenced from Souris *et al* (2016), were simulated with an energy spread of 0.83 MeV and a Gaussian shape in the transverse plane with a standard deviation of 4.2 mm. In each voxel, the number of emitted PGs with energies above 3 MeV (Sterpin *et al* 2015) was recorded.

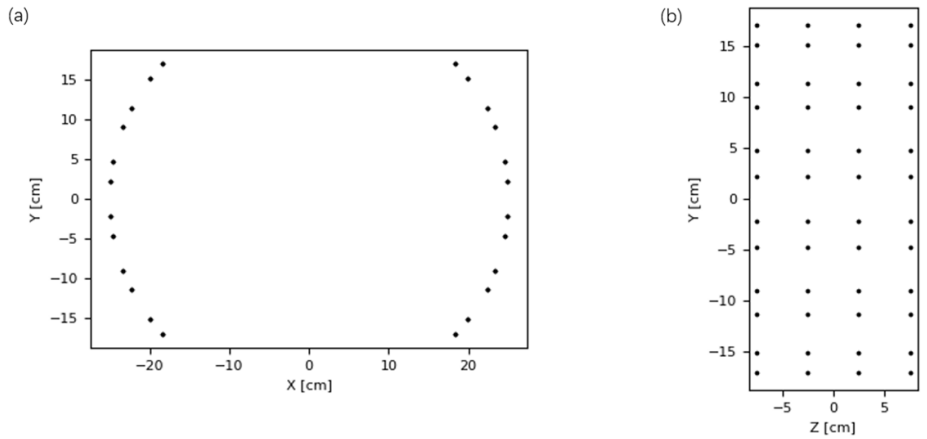

Figure 2. (a) Initial positions of PB distributions in the transverse plane (six angles on each side and two spots on each angle); (b) corresponding PB positions in the sagittal plane of the patient.

Table 1. PB energies (MeV) and PB numbers (N) of three datasets for RP and RDP models (N represents the PB number of the dataset before the overshooting PB filtering). The 125–210 MeV dataset corresponds to randomly sampled integer energies in that range.

	Dataset [MeV (N)]		
Model	Training	Validation	Test
RP	150 (1920)	150 (288)	150 (540)
RDP (SE)	150 (1920)	150 (288)	150 (540)
RDP (ME)	150 (1920), 175 (1920), 200 (1920)	125–210 (864)	150 (540), 175 (540), 200 (540), 125–210 (540)

The world coordinate system that Geant4 uses is aligned with the CT anatomy and its basis vectors $\{e_1, e_2, e_3\}$ point to the right lateral, posterior and superior direction of the patient respectively. The PB penetration was set to occur in the transverse plane and PBs were isotropically sampled across a cylindrical surface surrounding the prostate region. With the right lateral direction corresponding to 90°, PBs were confined in a predefined angular sector within the angle of $[40°, 140°] \cup [220°, 320°]$ to ensure none were positioned posterior or anterior to the patient. 12 different angles with a stepsize of 16.6° were sampled and a 2×4 spot grid with a resolution of 2.5 cm × 5 cm was then generated in the plane orthogonal to the direction from each of these 12 angles, as shown in figure 2. For each transverse plane, the PBs point towards the center of this plane, finally yielding 96 PBs for each patient.

Table 1 lists the three PB datasets that were simulated for RP and RDP models, with a consistent training/validation/test patient number split of 20/3/10. Due to limitations of the simulation time (averaged PB simulation time of about 20 hours per single core of an Intel(R) Xeon(R) Gold 6354 3.00 GHz CPU), 96 PBs per patient were simulated in all training and validation datasets, while 54 PBs randomly selected from 96 PBs per patient were simulated in all test datasets. For the mono-energetic RP model, a dataset containing RSP and PG maps of one single energy (SE) of 150 MeV was simulated, yielding 1920 PBs for training, 288 PBs for validation and 540 PBs for testing. For the energy-agnostic RDP model, to firstly assess the impact of the introduction of the dose, the same dataset of 150 MeV energy as for the RP model was used, but it included the corresponding simulated dose maps. Second, to explore the capability of the RDP model to predict PG distribution across multiple energies (ME), PBs at 150, 175, and 200 MeV energies in 20 patients were simulated for training (1920 PBs at each energy). To assess whether that model worked at arbitrary energies, 864 PBs in 3 patients were simulated for validation, each PB with an integer energy randomly sampled from 125 to 210 MeV. A total of 2160 PBs with energies of 150, 175, 200 MeV, and randomized energies ranging from 125 to 210 MeV (each sub test set comprised 540 PBs) were simulated for testing across 10 patients. Furthermore, in order to avoid the impact of overshooting PBs on the PG range modeling, 220 overshooting PBs in the 200 MeV energy dataset and 34 overshooting PBs in the 125–210 MeV energy dataset were filtered out.

2.3. Pre-processing

After the MC simulations for three datasets, the values of all the RSP, dose and PG maps outside of the patient body were set to zero. Then, for each PB, the world coordinate system was replaced with the BEV coordinate system, based on the initial position and orientation of the PB, to model the slices perpendicular to the beam direction. The 3D cuboids of RSP and PG maps for the RP model and the 3D cuboids of RSP,

dose and PG maps for the RDP models were cropped starting from the patient's surface and resampled. According to the energy range of all PBs from 125 to 210 MeV, the dimensions of all 3D cuboids were set to $320 \times 24 \times 24$, representing respectively the penetration depth z and the 2D lateral spread length x and width y of the PB, with a resolution of 1 mm × 2 mm × 2 mm. Finally, global normalization was performed for the three datasets described in table 1. The training, validation, and testing cuboids were normalized to the maximum RSP, dose and PG cuboid values of the training dataset.

2.4. Model architecture

The network architecture for the RP and RDP model is the same, corresponding to the LSTM model used in Neishabouri *et al* (2021) for proton dose calculation. The LSTM features one layer and the hidden state h_t is of size 1000 (the module architecture shown in figure 1(c)), followed by two back-end fully-connected layers with 100 neurons each and ReLu activation function. The final PG output has the size of the flattened RSP slice and gets reshaped to 2D.

For the training of the RP model and the two RDP models, the Adam optimizer was used to minimize the mean squared error (MSE) loss between the output PG and MC PG sequences. The initial learning rate was set to 10^{-5} and the batch size to 4. We used Python 3.8.10 and Pytorch 2.0.1 to implement the three models and trained them in an NVIDIA RTX A6000 GPU (48 GB memory) for 2000 epochs, which took less than 1 day for the RP and RDP (SE) models and around 2 days for the RDP (ME) model. The models with the best overall MSE validation loss were identified and used for the prediction. After training, the RP model and the two RDP models were evaluated on the corresponding test dataset described in table 1.

2.5. Evaluation

After obtaining all PG predictions from the three models, gamma passing rates, range shifts and integrated differences between predicted PGs and MC PGs were evaluated from three test datasets.

Specifically, we used the open-source library PyMedphys (Biggs *et al* 2022) to calculate the 3D global gamma passing rates γ_{pr} with a 2%/2 mm criteria, and the evaluated PG count threshold was set to 10% of the maximum PG count for each PG cuboid, similarly to the common methodology applied to dose (Song *et al* 2015).

To make the PG range shift evaluation accurate, all predicted PG and MC PG cuboids were resampled to 3200 bins in the z direction with a resolution of 0.1 mm, and a profile shifting method similar to Knopf *et al* (2008) was performed. After cropping the PB entry path (10 mm) into the body and the tail area of the PG profiles containing negligible values (PG counts lower than 1% of the PG maximum) to determine the profile evaluation interval I, the laterally integrated profiles $A(z)$ of MC and predicted PG distributions in I were evaluated. For δ ranging from −10.0 mm to 10.0 mm in steps of 0.1 mm, the root mean square error (RMSE) between the shifted prediction profile $A_{\mathrm{pred}}(\delta, z)$ and the stationary MC profile $A_{\mathrm{MC}}(z)$ was calculated at each shift voxel step. The step with the smallest RMSE value represented the step with the best overlap between the two profiles, while the PG range shift Δz can be determined from the amount of shift corresponding to that step:

$$\Delta z = \arg\min_{\delta} \left(\sqrt{\frac{1}{n} \sum_{z \in I} \left(A_{\mathrm{pred}}(\delta, z) - A_{\mathrm{MC}}(z) \right)^2} \right) [\mathrm{mm}] \tag{1}$$

where $n = |I|$ is the length of the profile evaluation interval.

In addition, laterally integrated PG-depth profiles with corresponding relative count differences ϵ_{rel} (%) through the z axis and total integrated count differences ϵ_{int} (%) from three dimensional predicted PG and MC PG cuboids were used for the relative disagreement and systematic offset evaluation, respectively.

$$\epsilon_{\mathrm{rel}} = 200 \frac{\left(P_{\mathrm{pred}} - P_{\mathrm{MC}} \right)}{\left(P_{\mathrm{pred}} + P_{\mathrm{MC}} \right)} [\%] \tag{2}$$

$$\epsilon_{\mathrm{int}} = \sum_{x} \sum_{y} \sum_{z} 100 \frac{|C_{\mathrm{pred}} - C_{\mathrm{MC}}|}{C_{\mathrm{MC}}} [\%] \tag{3}$$

where P_{pred} and P_{MC} are the predicted and MC PG count through the laterally integrated profile, and C_{pred} and C_{MC} are the predicted and MC PG count per voxel.

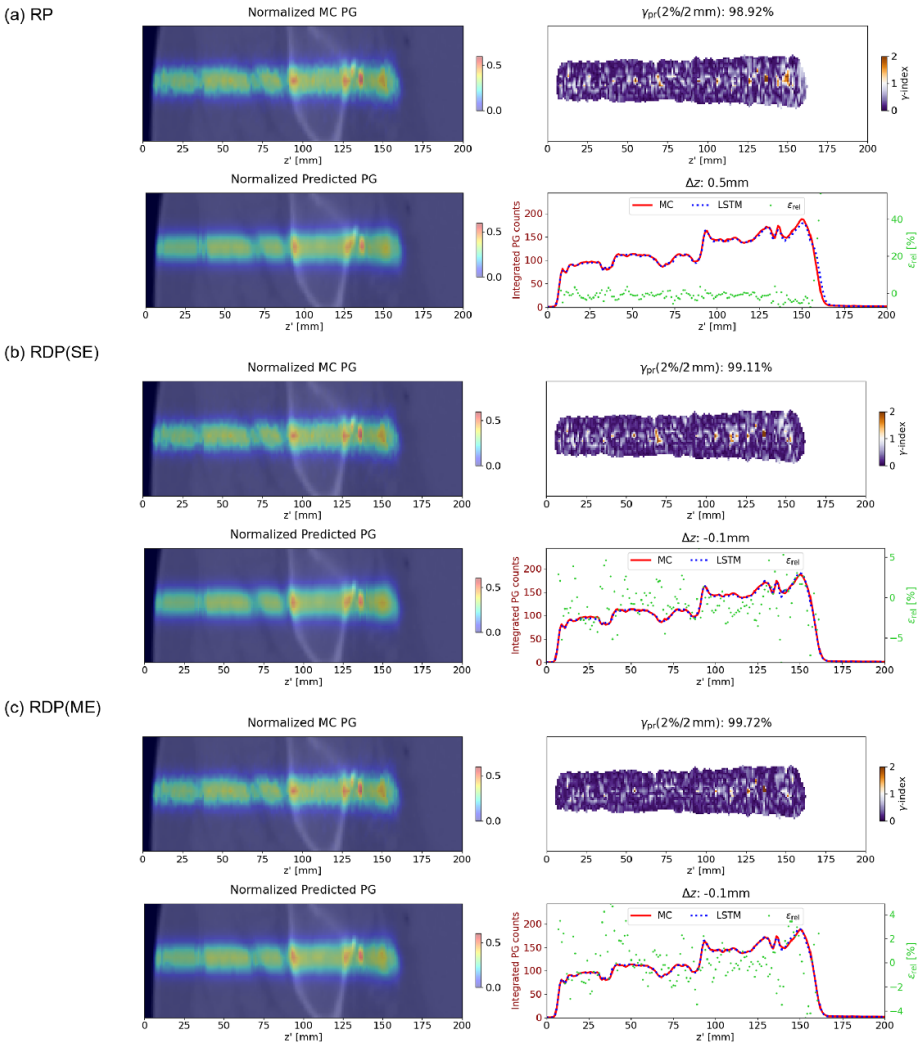

Figure 3. A nominal case in the 150 MeV test dataset: (a)–(c) *yz* slice of normalized MC PG, normalized predicted PG from (a) RP model, (b) RDP (SE) model and (c) RDP (ME) model, gamma index map crossing the maximum of MC PG, and laterally integrated normalized PG profiles with the relative count differences along the beam path.

3. Results

To evaluate and compare the predictive performance of the RP, RDP (SE) and RDP (ME) models, the test results of three models in the same 150 MeV test dataset are shown in section 3.1. To assess the ability of the RDP (ME) model to predict PG distributions of different energies, the test results of the RDP (ME) model in 150/175/200/125–210 MeV test datasets are shown in section 3.2. Finally, the prediction runtimes for RP and RDP models are reported in section 3.3.

3.1. Results for RP, RDP (SE) and RDP (ME) models in the 150 MeV test dataset

Due to the energy of all beams being 150 MeV, for visualization purposes, PG examples shown in this section were cropped from the original cuboid size (320 × 24 × 24) to 200 × 24 × 24 to focus on the main PG region. Figure 3 presents a nominal case predicted by the RP, RDP (SE), and RDP (ME) models in the 150 MeV test dataset, respectively in (a)–(c). Based on the identical RSP map, all the three models successfully predicted PG distributions with the gamma passing rate γ_{pr} higher than 98.9% and the range shift Δz within 0.5 mm. The corresponding laterally integrated profiles of MC and predicted PG also demonstrated good consistency. Moreover, in comparison to Δz of 0.5 mm and ϵ_{rel} of up to 40% of the predicted PG from the RP model, the predicted PGs from the RDP (SE) and RDP (ME) models with the auxiliary input of dose distribution showed better Δz (-0.1 mm) and better relative PG count difference ϵ_{rel} as shown in figures 3 (b) and (c).

To evaluate and compare the overall performance of the RP, RDP (SE) and RDP (ME) models in the 150 MeV test dataset, table 2 shows the mean, worst values and standard deviations (SD) of the evaluated

Table 2. The mean, worst values and SDs of evaluated metrics for the RP/RDP (SE) models in the 150 MeV test dataset and for the RDP (ME) model in 150/175/200/125–210 MeV test datasets.

		Δz (mm)				γ_{pr} (%)			ϵ_{int} (%)		
Model	E (MeV)	|mean|	SD	min	max	mean	SD	min	mean	SD	max
RP	150	0.50	0.58	−1.5	3.5	98.14	1.03	89.97	0.71	0.60	3.77
RDP(SE)	150	0.20	0.28	−0.8	1.6	98.18	1.11	91.44	0.71	0.56	3.34
RDP(ME)	150	0.15	0.22	−0.5	1.6	99.09	0.57	95.00	0.50	0.48	2.82
	175	0.13	0.16	−0.4	0.7	98.70	0.70	93.93	0.60	0.42	2.27
	200	0.11	0.13	−0.3	0.7	98.36	0.72	93.63	0.55	0.37	1.70
	125–200	0.15	0.19	−0.8	1.1	98.50	0.99	92.83	0.87	0.86	6.16

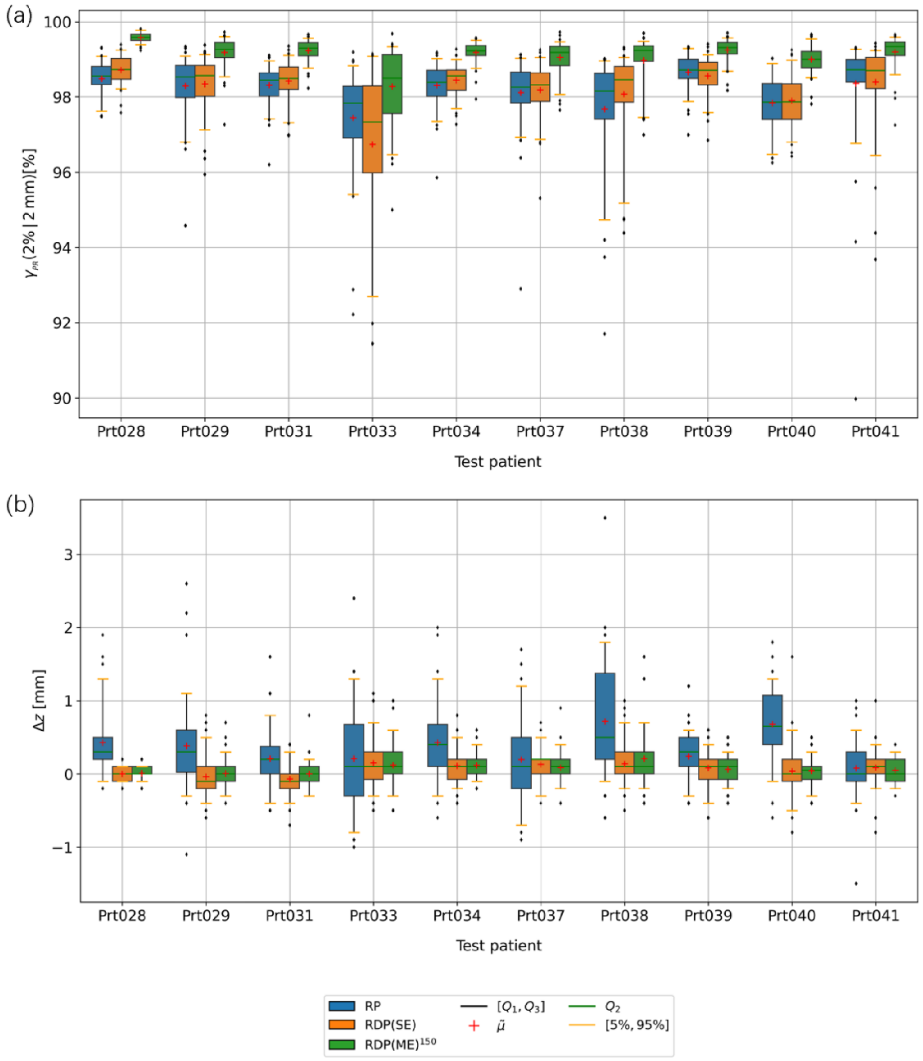

Figure 4. Patient specific boxplot for (a) γ_{pr} and (b) Δz of predicted PGs from the RP, RDP (SE) and RDP (ME) models in the 150 MeV energy test dataset. The red cross denotes the mean and the green line denotes the median value. The boxes denote the range between 25th and 75th percentiles. The orange whiskers denote the 5th and 95th percentiles, while the black points denote outliers outside of that range.

metrics for the three models. For the range shift metric Δz, the mean absolute Δz was within 0.5 mm for the RP model. In contrast, it was 0.2 mm for the RDP (SE) model and 0.15 mm for the RDP (ME) model, both outperforming the RP model. The worst Δz of the RDP (SE) and RDP (ME) models (within 1.6 mm) were also better than that of the RP model (within 3.5 mm). For the γ_{pr} and ϵ_{int} metrics, the RDP (ME) model demonstrated the best performance, and the RDP (SE) model showed marginally superior results compared to the RP model. γ_{pr} results of all three models in the 150 MeV test dataset were above 89.9% and ϵ_{int} within 3.77%.

Figure 4 further shows patient specific boxplots for the γ_{pr} and Δz metrics of predicted PGs from the three models across 10 test patients in the 150 MeV test dataset. Similar to the results analyzed in table 2, the

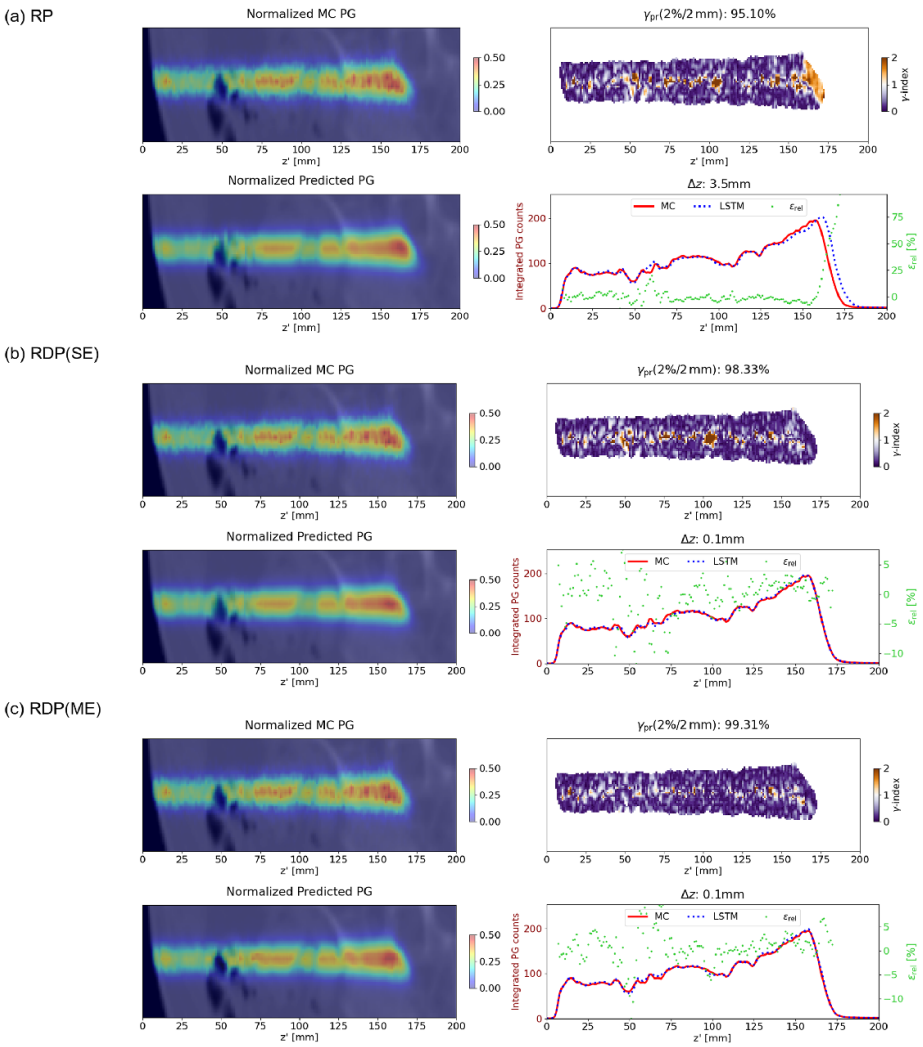

Figure 5. The worst case predicted by the RP model and the corresponding cases predicted by the RDP (SE) and RDP (ME) model in the 150 MeV test dataset: (a)–(c) *yz* slice of normalized MC PG, normalized predicted PG from (a) RP model, (b) RDP (SE) model and (c) RDP (ME) model, gamma index map crossing the maximum of MC PG, and laterally integrated normalized PG profiles with the relative count differences along the beam path.

RDP (ME) model performed the best in the γ_{pr} metric. The Δz of the 5th to 95th percentile for the RDP (SE) and RDP (ME) models (rarely larger than 1 mm) were smaller than that of the RP model. The results demonstrated that the introduction of dose distribution for the model can improve the predicted PG range accuracy.

Figure 5(a) presents the worst case in terms of Δz predicted by the RP model in the 150 MeV test dataset. It can be seen that the air cavity contributed to an overestimation of the predicted PG range by the RP model. Figures 5(b) and (c) show the corresponding case predicted by the RDP (SE) and RDP (ME) models, respectively. In this case, RDP models aided by the introduction of dose enhanced the accuracy of PG range prediction. Figures 6(a) and (b) present the worst case in terms of Δz predicted by the RDP (SE) and RDP (ME) model in the 150 MeV test dataset, which had smaller range shift Δz than the worst case predicted by the RP model.

3.2. Results for the RDP (ME) model in the 150/175/200/125–210 MeV test datasets

Figure 7 presents three nominal cases predicted by the RDP (ME) model in the 125–210 MeV test dataset (129 MeV PG in figure 7(a), 177 MeV PG in figure 7(b) and 206 MeV PG in figure 7(c)). The laterally integrated profiles of MC and predicted PG exhibited good consistency and absolute Δz were within 0.5 mm, indicating that the RDP (ME) model, trained with data from three distinct energy levels (150/175/200 MeV), has the ability to generalize the mapping to energies beyond the 150–200 MeV range. Figure 8 further shows the boxplots for the γ_{pr} and Δz metrics of predicted PGs from the RP model and the two RDP models in

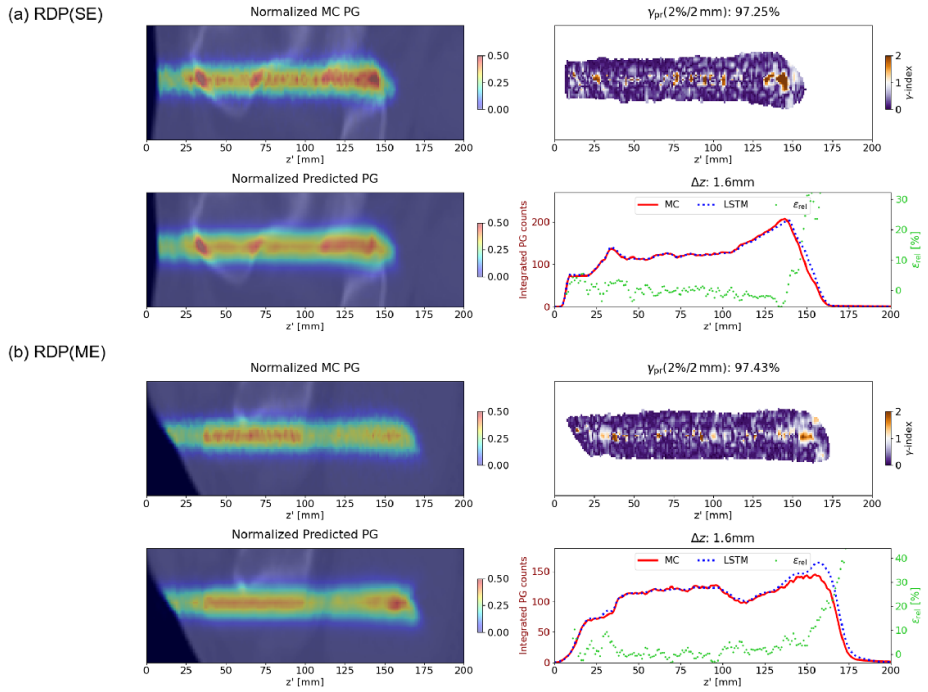

Figure 6. The worst case predicted by the RDP (SE) and RDP (ME) model in the 150 MeV test dataset: (a), (b) *yz* slice of normalized MC PG, normalized predicted PG from (a) RDP (SE) model and (b) RDP (ME) model, gamma index map crossing the maximum of MC PG, and laterally integrated normalized PG profiles with the relative count differences along the beam path.

different test datasets. From the results of four 150/175/200/125–210 MeV test sets, the RDP (ME) model consistently achieved γ_{pr} above 91%, with only three outliers exceeding Δz of 1 mm. The worst case predicted by the RDP (ME) model in the 125–210 MeV test dataset is also shown in figure 9, where the air cavity caused an overestimation of the PG range by the model. Overall, the RDP (ME) model achieved high accuracy in the corresponding four test datasets.

3.3. Model prediction runtimes

The model prediction runtime includes the cuboid extraction time and the model inference time. All computation measurements were carried out on a workstation equipped with an Intel(R) Xeon(R) Gold 6354 3.00 GHz CPU and an NVIDIA RTX A6000 GPU. The averaged extraction time for a RSP or dose cuboid was about 55 ms, and the averaged inference time of the RP and RDP model for a single PB was around 10 ms (10.3 ms) and 11 ms (10.6 ms), respectively. In total, the single PB prediction time for the RP model was within 70 ms, and for the RDP model was within 130 ms. Geant4 MC PG simulations took approximately 20–30 h on a single CPU core for a single PB comprising 1×10^7 protons, with the duration varying depending on the initial proton energy. Although multi-core CPUs (e.g. 32 cores) can accelerate this process significantly, the runtime of Geant4 simulations that model the full set of physics, still requires tens of minutes.

4. Discussion

In this study, we proposed a sub-second PG prediction approach for proton beam spots based on a previous LSTM-based proton dose calculation model (Neishabouri *et al* 2021). Our experiments indicate that both methods, either using RSP as the only model input or RSP plus dose as model inputs, can achieve accurate PG predictions in prostate cancer patients. The comparative test results of the RP and RDP (SE) models demonstrate that incorporating dose information into the LSTM model can drastically improve the accuracy of PG range prediction. Despite the dose and RSP slices being flattened and concatenated, no loss of spatial coherence is observed in the outcome. Besides, the superior performance of the RDP (ME) model over the RDP (SE) model on the 150 MeV test dataset suggests that the model mapping acquired from analyzing more data with different energies can enhance the outcome for single energies when considered separately, demonstrating increased robustness. Moreover, the results of the RDP (ME) model on the 125–200 MeV test

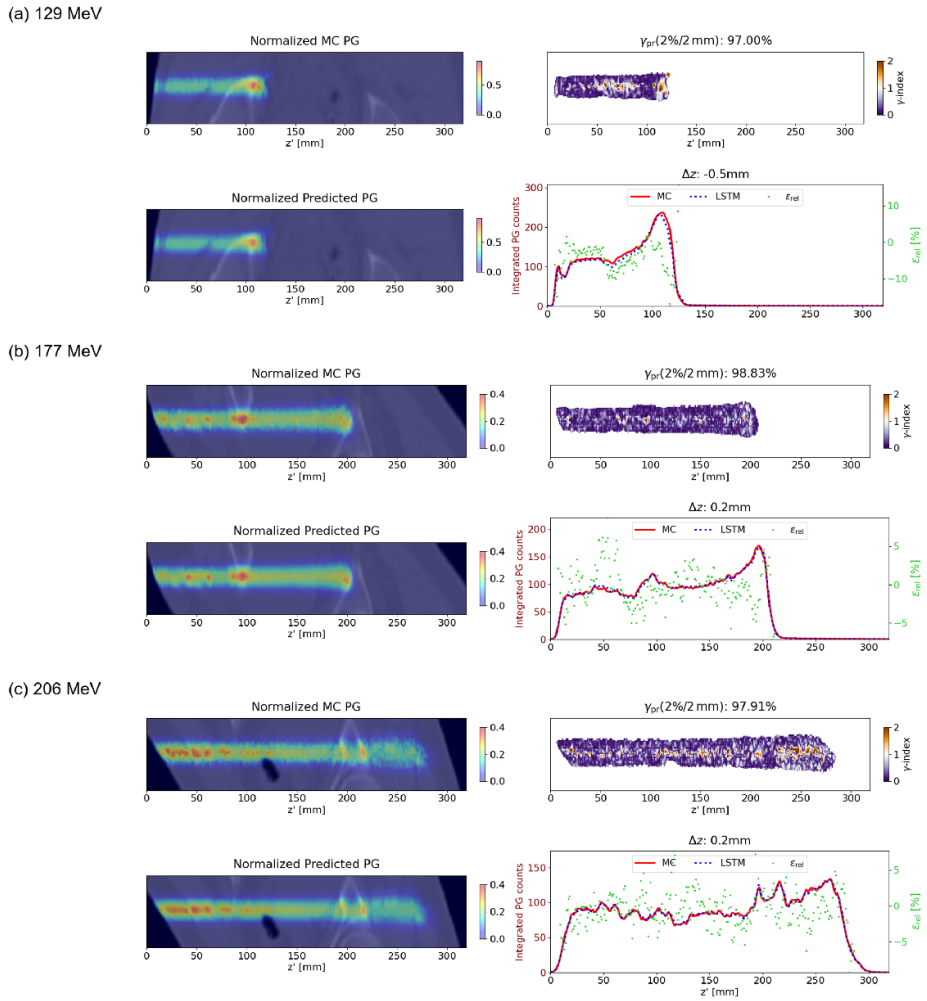

Figure 7. Three nominal cases predicted by the RDP (ME) model in the 125–210 MeV test dataset: (a)–(c) *yz* slice of normalized MC PG, normalized predicted PG for energies of (a) 129 MeV, (b) 177 MeV and (c) 206 MeV, gamma index map crossing the maximum of MC PG, and laterally integrated normalized PG profiles with the relative count differences along the beam path.

dataset indicate that the model trained on three separate energies can generalize the mapping across a wide range of unseen proton energies, which is quite interesting. Although the RDP (ME) model showed its mean PG range shift is lower than 0.15 mm across all test datasets, the presence of air cavities sometimes still resulted in increased errors (e.g. figure 9, Prt033 and Prt038 in figure 4). This may be due to the insufficient data, as only 433 out of 6404 PBs in the training and validation datasets include cases with air cavities.

Once the LSTM model was trained, the PG prediction for an individual beam spot can be finished within 70 ms with the RP model (averaged model inference time of 10.3 ms) and 130 ms with the RDP model (averaged model inference time of 10.6 ms), including data preparation. The inference time comparison between the RP and RDP model is also interesting: the RDP model introduced only 0.3 ms runtime surplus while it processed double amount of input vector (from 576 to 1152) for each time step. This is mainly because the feature dimension of the input has a much smaller influence on the inference time of an LSTM compared to the time step size (320 in our case) and the hidden layer size, allowing it to handle long-vector inputs efficiently. In our case, this means the LSTM model can process more channels of inputs or larger 2D slices with a negligible increase in inference time. Additionally, the inference of both the RP and RDP models exhibited low GPU memory requirements, with each model utilizing less than 1.5 GB.

Our method is compared with two fast analytical PG approximation methods (Sterpin *et al* 2015, Pinto *et al* 2020) and the multi-CPU MC method (Souris *et al* 2016) in terms of accuracy and runtimes. The first analytical method (Sterpin *et al* 2015) can approximate PG emission profiles based on look-up tables. In a given geometry, the local PG emission can be derived by adding the contribution of each chemical element according to the local energy of the proton and the local composition. Compared with MC PG emission simulations, their method could achieve a 3%/1 mm symmetric γ_{pr} (Dhakal and Yepes 2014) (with the PG

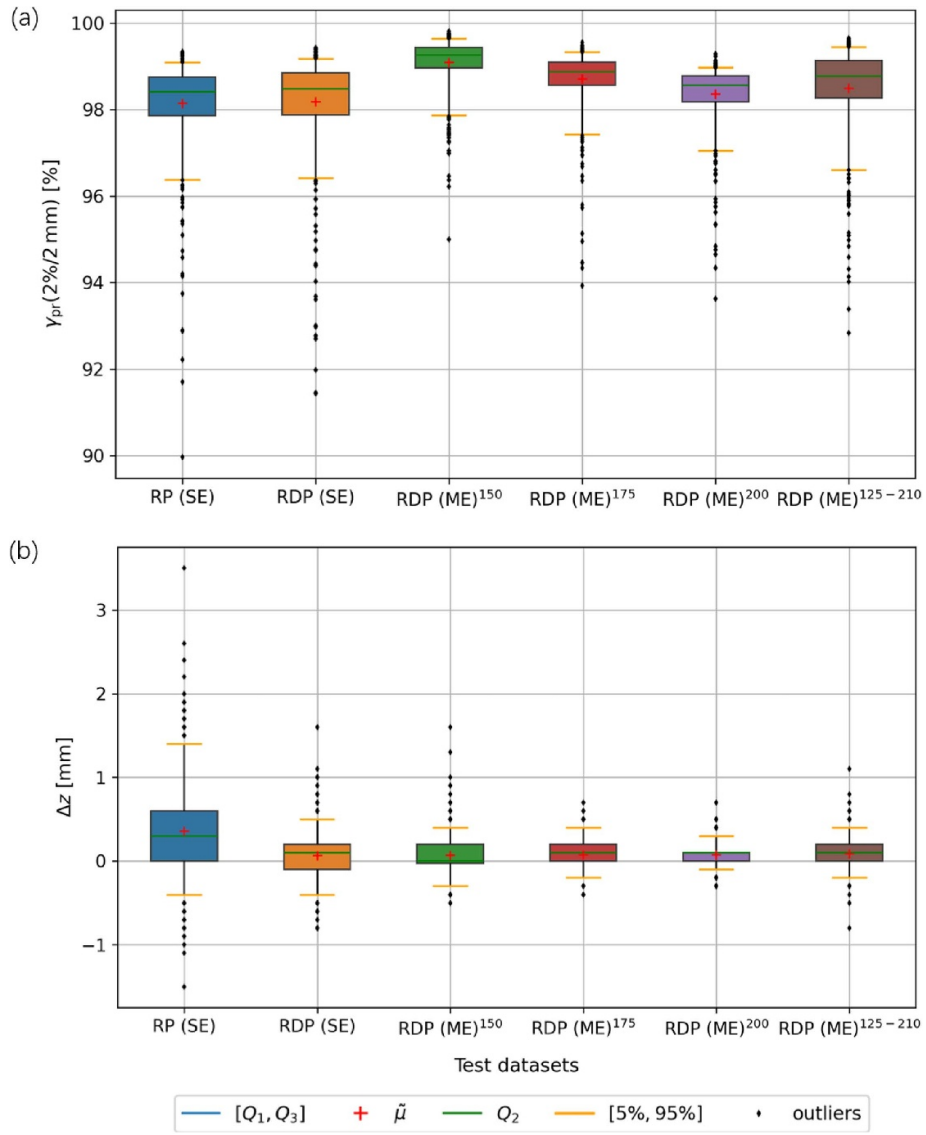

Figure 8. Boxplot for (a) γ_{pr} and (b) Δz metrics of predicted PGs from the RP and RDP (SE) model in the 150 MeV test dataset, and RDP (ME) model in 150/175/200/125–210 MeV test datasets. The red cross denotes the mean and the green line denotes the median value. The boxes denote the range between 25th and 75th percentiles. The orange whiskers denote the 5th and 95th percentiles, while the black points denote outliers outside of that range.

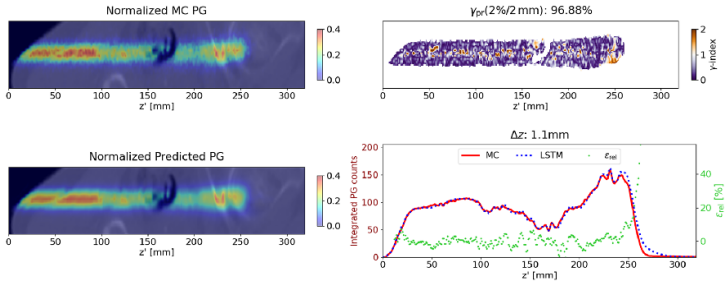

Figure 9. *yz* slice of normalized MC PG, normalized predicted PG, gamma index map crossing the maximum of MC PG, and laterally integrated normalized PG profiles with the relative count differences along the beam path of the worst case predicted by the RDP (ME) model in the 125–210 MeV test dataset.

count threshold set to 10% of the maximum PG count) above 97.6% for PBs in different heterogeneous phantoms and 90.2% for PBs in a highly heterogeneous lung patient case, demonstrating very good PG accuracy. Their evaluated energy range was limited to six discrete energies (100, 110, 115, 145, 150, and 200 MeV). The computation time for this method on a single CPU core, considering 10^7 incident protons, ranged from 0.3 s (without super sampling) to 10 s (super sampling with 113 rays), depending on the number of sample rays per spot. The second analytical method (Pinto *et al* 2020) utilized TPS PB dose as input for convolution filtering operations to output PG distribution. This filtering approach was compared to MC simulations in terms of shifts of the longitudinal PG profiles for each spot in patient cases, involving abdomen and more heterogeneous head-and-neck areas, achieving mean shifts ranging from −1.3 to 0.9 mm across four patient cases. The computation time of this method in single threaded CPU mode was around 63 ms (30/477 = 0.063 s) per spot. It should be noted that it is unfair to directly compare the PG results of a method using a less accurate analytical dose algorithm as input with our LSTM method, which uses MC dose as input, given that MC PG served as the ground truth. However, fast MC dose inputs are becoming more accessible with the availability of many fast MC-based dose calculation algorithms nowadays. The MC algorithm of a commercial TPS (RayStation) has been validated to achieve a median computation time of 5.2 s for clinical plans with 1.0% statistical uncertainty (Fracchiolla *et al* 2021). As the current state-of-the-art for MC dose calculation speed is 10^6–10^7 protons per second, MC methods may overtake analytical approaches in the future (Holmes *et al* 2024). For the multi-CPU MC method (Souris *et al* 2016), although no shift metrics and test results for patient cases were reported, the method achieved maximum deviations of less than 20% in lateral PG emission profiles of a 150 MeV PB in a heterogeneous phantom compared to Geant4 simulations, and deviations mainly occurred at interfaces in the phantom. The simulation time of this multi-CPU MC method with its fastest configuration was less than 25 s to simulate 10^7 protons in soft tissue. Compared to the aforementioned methods, our approach utilizing the RDP(ME) model achieved an average shift within 0.15 mm per spot in 10 prostate patient cases, with the test energy range covering from 125 to 210 MeV. It also demonstrates the speed competitiveness with a prediction time within 130 ms (model inference time of 11 ms) for 10^7 simulated protons, and its runtime does not scale with the number of protons in the PB.

Due to computing resource limitations, the number of initial protons simulated per spot in this study was set to 10^7, consistent with Souris *et al* (2016). This can be sufficient for PG emission simulation, but may not be adequate for the PG detection, where usually more than 10^8 proton simulations are required (Draeger *et al* 2018). It's worthwhile exploring if our RDP model trained with the data simulated by 10^7 protons can demonstrate robust performance in different proton number levels. To investigate this, we conducted a preliminary experiment: One random selected PB (167 MeV) from the 125–210 MeV test dataset was simulated using 10^6 and 10^8 protons respectively, resulting in two sets of dose and PG distributions (due to the MC simulation of 10^8 protons taking approximately 6 days, only one PB was simulated). Subsequently, the cuboid extraction and normalization was done to get the model input dose and RSP cuboids, and the ground truth PG cuboid. It needs to be mentioned that, compared to dose and PG simulated by 10^7 protons, those simulated by 10^6 and 10^8 protons were normalized by a factor of 1/10 and 10 respectively. The test results of RDP (ME) model for a 167 MeV PB simulated by 10^6, 10^7 and 10^8 protons are shown in the appendix with figure A1. Our model demonstrated its robust performance in terms of range prediction, both for 10^8 protons (less noisy dataset) and for 10^6 protons (noisier dataset). This is important, given that the primary use of PG distributions is envisaged to be range monitoring. Concerning the γ_{pr} metric, the model achieved high performance for the 10^8 protons dataset, but lower for the 10^6 protons dataset. As the model prediction is in general smoother than the ground truth data, we can conclude that for the 10^6 protons dataset, where the prediction and MC differ the most in terms of fluctuations, the γ_{pr} tends to be worse.

Although our method showed promising results in prostate patient cases, more extensive sites and geometries need to be included in future studies. It should be noted that our method could also work with the Hounsfield unit maps of patient's CT, but since RSP maps are consistent with dose and with PG generated from MC simulations and can be easily obtained from the treatment planning system (TPS), we chose to directly use RSP maps. As the PG distributions are strongly correlated with the chemical composition, adding that kind of information (which can be derived from CT or even dual energy CT (Niepel *et al* 2024)) might yield better results than RSP alone, and multiple channel inputs would not significantly increase the inference time of our LSTM model. The proposed mixed-input method (RDP model) currently uses MC dose as one of the model inputs. For practical application, the model's dose input must be adjusted according to the specific dose algorithm used by the TPS, and the model needs to be retrained. This ensures that the predicted PG distribution from our method would be consistent with the dose calculated by the TPS, which requires verification. It is worthwhile to investigate the model performance when using a dose distribution

less accurate than the one from MC simulation (e.g. calculation errors due to heterogeneities) as our model input. Since we do not have access to an analytical PB algorithm with beam splitting from a commercial TPS, we developed a dose prediction model based on the LSTM method instead (Neishabouri *et al* 2021, Radonic *et al* 2024) using the 150 MeV dataset (see figure B1 in appendix B for γ_{pr} performance). We believe that an LSTM dose prediction, especially in highly heterogeneous regions (prediction errors due to air cavities in our cases), is a good approximation of a state-of-the-art TPS analytical dose. The comparison of the RDP model using MC dose and LSTM dose input in the 150 MeV energy dataset are reported in appendix B. From figures B2 and B3, we found that the errors introduced by the input dose reduced the PG prediction accuracy of the RDP PG prediction model, and its accuracy was comparable to the RP PG prediction model.

Considering the time constraints of the clinical workflow, our LSTM approach and several analytical methods can offer alternative solutions for the slow PG MC simulations. Similar to the analytical methods, our method can also be used for PB selection before treatment delivery. After rapid PG predictions of different PBs from our model, new treatment planning strategies (Tian *et al* 2018, 2020) can be employed to select PBs with high correlation as suitable candidates for reliable PG detection, before delivering the complete plan. Our method can be further utilized for real-time plan adaptation, which would also require real-time PG prediction. We believe that this is feasible given current advancements and could be of interest, for example, in magnetic resonance imaging (MRI) guided proton therapy (Hoffmann *et al* 2020). Since 3D+t CTs with a temporal resolution of 3.65 Hz may be generated from orthogonal 2D cine MRI (Rabe *et al* 2023), the real-time update of 3D+t RSPs should also be achievable. By employing an AI-accelerated dose engine, such as an LSTM dose model (Neishabouri *et al* 2021), real-time dose distributions could be obtained. Consequently, real-time PG prediction using our model could be achieved. Fast PG prediction is only one of the steps in the workflow above, but nonetheless it would be an important one. Finally, our current implementation predicts PG distributions at the emission level inside the patient body, therefore, validation with PG detection data from experimental PG camera also need to be explored in the future.

5. Conclusions

We successfully developed a sub-second LSTM-based PG emission prediction method. Its accuracy in prostate patients has been confirmed across an extensive range of proton energies. The proposed method has the potential to improve the efficiency of online proton range verification in the adaptive proton therapy workflow.

Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

The work of Fan Xiao was supported by China Scholarship Council (No. 202308440107).

Conflict of Interest

The Department of Radiation Oncology of the University Hospital of LMU Munich has research agreements with Brainlab, Elekta and C-Rad.

Appendix A

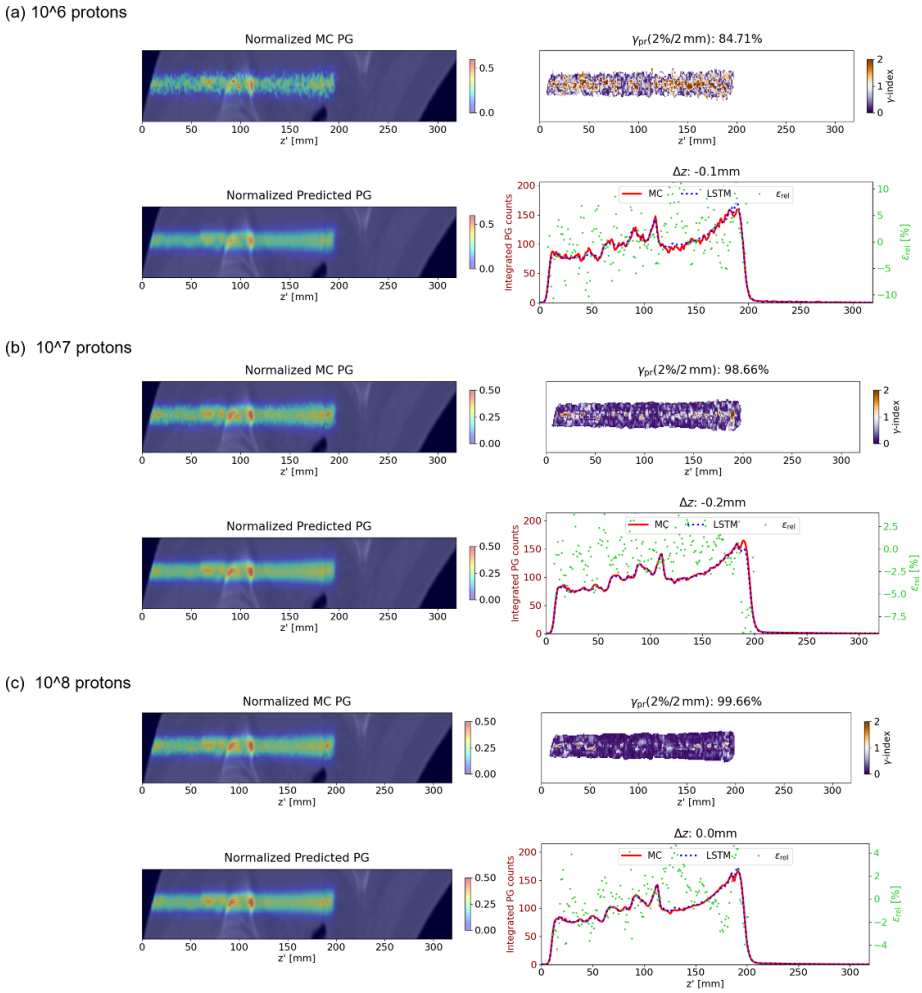

Figure A1. One 167 MeV PB case simulated with different number of protons: (a)-(c) *yz* slice of normalized MC PG simulated with (a) 10^6, (b) 10^7 and (c) 10^8 protons, normalized predicted PG from the RDP (ME) model, gamma index map crossing the maximum of MC PG, and laterally integrated normalized PG profiles with the relative count differences along the beam path.

Appendix B

We trained the dose prediction model based on the LSTM method (Neishabouri *et al* 2021, Radonic *et al* 2024) using the same training, validation, and test datasets in the 150 MeV energy dataset shown in table 1 in the main text. With the input of RSP and output of approximated MC dose, the LSTM dose model was trained using the same training parameters as the RP model. The model with the lowest MSE validation loss was saved. The patient specific boxplots for γ_{pr} of predicted doses from the LSTM dose model in the 150 MeV test dataset are shown in figure B1, and the predicted dose of one selected outlier test case with an air cavity is shown in figure B2(a). The γ_{pr} values for all predicted cases are above 90%, with Prt033 and Prt038 being relatively lower due to the higher presence of air cavities in the cuboids in these two test cases. Figure B2(a) demonstrates that the LSTM dose may introduce prediction errors due to air cavities compared to MC dose in our cases.

To generate the training and validation datasets for the RDP model using LSTM dose input, doses from the 150 MeV training and validation datasets were predicted using the LSTM dose model. Then the RDP (SE, LSTM dose) PG prediction model was trained using the same training parameters as the RDP (SE, MC dose) PG prediction model and saved with the lowest MSE validation loss. The predicted PG results from the RDP (SE, LSTM dose) PG prediction model in 10 test patients were compared with the results from the RDP (SE,

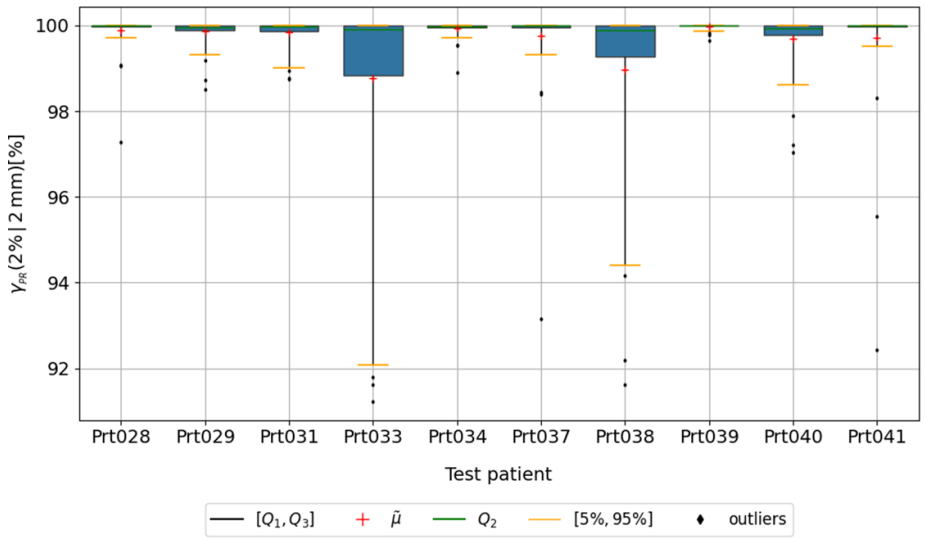

Figure B1. Patient specific boxplots for γ_{pr} of predicted doses from the LSTM dose model in the 150 MeV energy test dataset. The red cross denotes the mean and the green line denotes the median value. The boxes denote the range between 25th and 75th percentiles. The orange whiskers denote the 5th and 95th percentiles, while the black points denote outliers outside of that range.

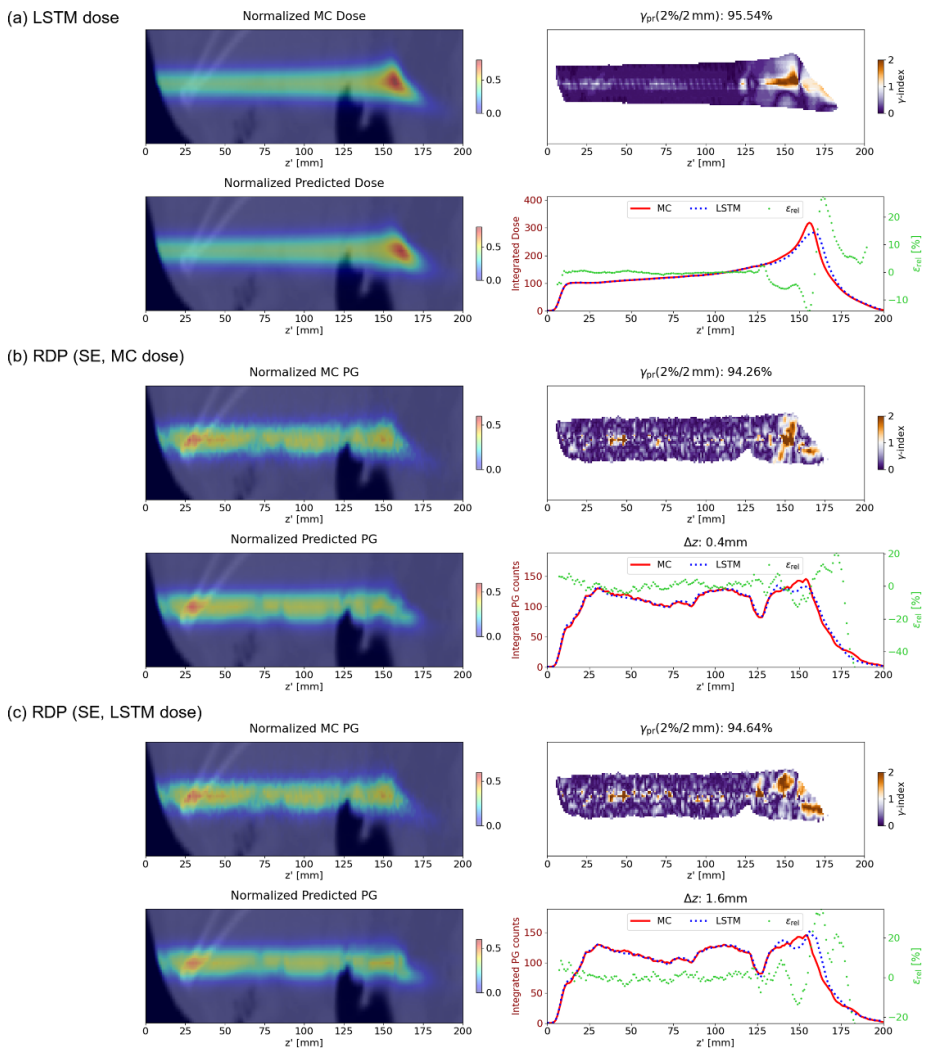

Figure B2. A selected outlier test case with an air cavity in the 150 MeV test dataset: (a) *yz* slice of normalized MC dose, normalized predicted dose, gamma index map and laterally integrated normalized dose profiles with relative differences along the beam path; (b), (c) *yz* slice of normalized MC PG, normalized predicted PG from (b) RDP (SE, MC dose) model and (c) RDP (SE, LSTM dose) model, gamma index map crossing the maximum of MC PG, and laterally integrated normalized PG profiles with the relative count differences along the beam path.

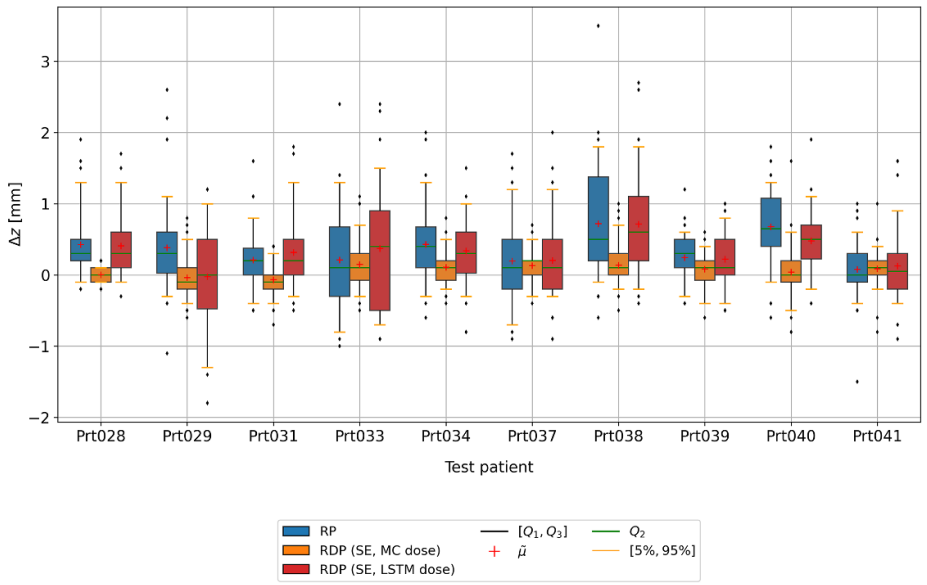

Figure B3. Patient specific boxplots for Δz of predicted PGs from the RP, RDP (SE, MC dose) and RDP (SE, LSTM dose) models in the 150 MeV energy test dataset. The red cross denotes the mean and the green line denotes the median value. The boxes denote the range between 25th and 75th percentiles. The orange whiskers denote the 5th and 95th percentiles, while the black points denote outliers outside of that range.

MC dose) PG prediction model. Figure B3 shows the patient specific boxplots for Δz of predicted PGs from the RP, RDP (SE, MC dose) and RDP (SE, LSTM dose) PG prediction models in the 150 MeV energy test dataset, and the predicted PGs from the RDP (SE, MC dose) and RDP (SE, LSTM dose) PG prediction models in the selected outlier test case with an air cavity are shown in figures B2(b) and (c). From figures B2 and B3, we found that the errors introduced by the input dose reduced the PG prediction accuracy of the RDP PG prediction model, and its accuracy is comparable to the RP PG prediction model.

ORCID iDs

Fan Xiao https://orcid.org/0000-0002-7502-0730
Domagoj Radonic https://orcid.org/0000-0002-0180-7995
Niklas Wahl https://orcid.org/0000-0002-1451-223X
Ahmad Neishabouri https://orcid.org/0000-0001-8060-2957
Katia Parodi https://orcid.org/0000-0001-7779-6690
Guillaume Landry https://orcid.org/0000-0003-1707-4068

References

Berthold J *et al* 2021 First-in-human validation of CT-based proton range prediction using prompt gamma imaging in prostate cancer treatments *Int. J. Radiat. Oncol. Biol. Phys.* **111** 1033–43
Biggs S *et al* 2022 Pymedphys: a community effort to develop an open, python-based standard library for medical physics applications *J. Open Source Softw.* **7** 4555
Choi H J, Jang J W, Shin W-G, Park H, Incerti S and Min C H 2020 Development of integrated prompt gamma imaging and positron emission tomography system for *in vivo* 3-D dose verification: a Monte Carlo study *Phys. Med. Biol.* **65** 105005
Dendooven P, Buitenhuis H, Diblen F, Heeres P, Biegun A, Fiedler F, Van Goethem M, Van der Graaf E and Brandenburg S 2015 Short-lived positron emitters in beam-on PET imaging during proton therapy *Phys. Med. Biol.* **60** 8923
Dhakal T R and Yepes P 2014 A symmetric probabilistic γ-index for Monte Carlo dose comparisons *Phys. Med. Biol.* **59** 153
Draeger E, Mackin D, Peterson S, Chen H, Avery S, Beddar S and Polf J 2018 3D prompt gamma imaging for proton beam range verification *Phys. Med. Biol.* **63** 035019
Fracchiolla F *et al* 2021 Clinical validation of a GPU-based Monte Carlo dose engine of a commercial treatment planning system for pencil beam scanning proton therapy *Phys. Med.* **88** 226–34
Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735–80
Hoffmann A *et al* 2020 MR-guided proton therapy: a review and a preview *Radiat. Oncol.* **15** 129
Holmes J, Feng H, Zhang L, Fix M K, Jiang S B and Liu W 2024 Fast Monte Carlo dose calculation in proton therapy *Phys. Med. Biol.* **69** 17TR01
Janssen F, Landry G, Lopes P C, Dedes G, Smeets J, Schaart D, Parodi K and Verhaegen F 2014 Factors influencing the accuracy of beam range estimation in proton therapy using prompt gamma emission *Phys. Med. Biol.* **59** 4427
Javaid U, Souris K, Huang S and Lee J A 2021 Denoising proton therapy Monte Carlo dose distributions in multiple tumor sites: a comparative neural networks architecture study *Phys. Med.* **89** 93–103

Knopf A-C and Lomax A 2013 *In vivo* proton range verification: a review *Phys. Med. Biol.* **58** R131
Knopf A, Parodi K, Bortfeld T, Shih H A and Paganetti H 2009 Systematic analysis of biological and physical limitations of proton beam range verification with offline PET/CT scans *Phys. Med. Biol.* **54** 4477
Knopf A, Parodi K, Paganetti H, Cascio E, Bonab A and Bortfeld T 2008 Quantitative assessment of the physical potential of proton beam range verification with PET/CT *Phys. Med. Biol.* **53** 4137
Litzenberg D, Roberts D, Lee M, Pham K, Vander Molen A, Ronningen R and Becchetti F 1999 On-line monitoring of radiotherapy beams: experimental results with proton beams *Med. Phys.* **26** 992–1006
Min C-H, Kim C H, Youn M-Y and Kim J-W 2006 Prompt gamma measurements for locating the dose falloff region in the proton therapy *Appl. Phys. Lett.* **89** 183517
Moteabbed M, Espana S and Paganetti H 2011 Monte Carlo patient study on the comparison of prompt gamma and PET imaging for range verification in proton therapy *Phys. Med. Biol.* **56** 1063
Neishabouri A, Wahl N, Mairani A, Köthe U and Bangert M 2021 Long short-term memory networks for proton dose calculation in highly heterogeneous tissues *Med. Phys.* **48** 1893–908
Niepel K, Tattenberg S, Marants R, Hu G, Bortfeld T, Verburg J, Sudhyadhom A, Landry G and Parodi K 2024 Validation of dual-energy CT-based composition analysis using fresh animal tissues and composition-optimized tissue equivalent samples *Phys. Med. Biol.* **69** 165033
Paganetti H 2012 Range uncertainties in proton therapy and the role of Monte Carlo simulations *Phys. Med. Biol.* **57** R99
Paganetti H *et al* 2021 Roadmap: proton therapy physics and biology *Phys. Med. Biol.* **66** 05RM01
Pastor-Serrano O and Perkó Z 2022 Millisecond speed deep learning based proton dose calculation with Monte Carlo accuracy *Phys. Med. Biol.* **67** 105006
Pinto M, Kröniger K, Bauer J, Nilsson R, Traneus E and Parodi K 2020 A filtering approach for PET and PG predictions in a proton treatment planning system *Phys. Med. Biol.* **65** 095014
Purushothaman S *et al* 2023 Quasi-real-time range monitoring by in-beam PET: a case for 15o *Sci. Rep.* **13** 18788
Rabe M *et al* 2023 Continuous time-resolved estimated synthetic 4D-CTs for dose reconstruction of lung tumor treatments at a 0.35 T MR-linac *Phys. Med. Biol.* **68** 235008
Radonic D *et al* 2024 Proton dose calculation with LSTM networks in presence of a magnetic field *Phys. Med. Biol.* **69** 215019
Schmid S, Landry G, Thieke C, Verhaegen F, Ganswindt U, Belka C, Parodi K and Dedes G 2015 Monte Carlo study on the sensitivity of prompt gamma imaging to proton range variations due to interfractional changes in prostate cancer patients *Phys. Med. Biol.* **60** 9329
Smeets J *et al* 2012 Prompt gamma imaging with a slit camera for real-time range control in proton therapy *Phys. Med. Biol.* **57** 3371
Song J-H, Kim M-J, Park S-H, Lee S-R, Lee M-Y, Lee D S and Suh T S 2015 Gamma analysis dependence on specified low-dose thresholds for VMAT QA *J. Appl. Clin. Med. Phys.* **16** 263–72
Souris K, Lee J A and Sterpin E 2016 Fast multipurpose Monte Carlo simulation for proton therapy using multi-and many-core CPU architectures *Med. Phys.* **43** 1700–12
Sterpin E, Janssens G, Smeets J, Vander Stappen F, Prieels D, Priegnitz M, Perali I and Vynckier S 2015 Analytical computation of prompt gamma ray emission and detection for proton range verification *Phys. Med. Biol.* **60** 4915
Tian L, Landry G, Dedes G, Kamp F, Pinto M, Niepel K, Belka C and Parodi K 2018 Toward a new treatment planning approach accounting for *in vivo* proton range verification *Phys. Med. Biol.* **63** 215025
Tian L, Landry G, Dedes G, Pinto M, Kamp F, Belka C and Parodi K 2020 A new treatment planning approach accounting for prompt gamma range verification and interfractional anatomical changes *Phys. Med. Biol.* **65** 095005
Unkelbach J, Bortfeld T, Martin B C and Soukup M 2009 Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning *Med. Phys.* **36** 149–63
Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I 2017 Attention is all you need *Adv. Neural Inf. Process. Syst.* **30** 6000–10
Wrońska A *et al* 2021 Prompt-gamma emission in GEANT4 revisited and confronted with experiment *Phys. Med.* **88** 250–61
Wu C, Nguyen D, Xing Y, Montero A B, Schuemann J, Shang H, Pu Y and Jiang S 2021 Improving proton dose calculation accuracy by using deep learning *Mach. Learn.: Sci. Technol.* **2** 015017
Xie Y *et al* 2017 Prompt gamma imaging for *in vivo* range verification of pencil beam scanning proton therapy *Int. J. Radiat. Oncol. Biol. Phys.* **99** 210–8
Zhang X, Hu Z, Zhang G, Zhuang Y, Wang Y and Peng H 2021 Dose calculation in proton therapy using a discovery cross-domain generative adversarial network (DiscoGAN) *Med. Phys.* **48** 2646–60