A long short-term memory network-based proton dose calculation method in a magnetic field

Fan Xiao[1], Domagoj Radonic[1,2], Niklas Wahl[3,4], Luke Voss[3,4,5], Ahmad Neishabouri[4,6], Nikolaos Delopoulos[1], Stefanie Corradini[1], Claus Belka[1,7,8], George Dedes[2], Christopher Kurz[1], and Guillaume Landry[1]

[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]Ruprecht Karl University of Heidelberg, Institute of Computer Science, Heidelberg, Germany
[6]Clinical Cooperation Unit Radiation Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany
[7]German Cancer Consortium (DKTK), partner site Munich, a partnership between DKFZ and LMU University Hospital, Munich, Germany
[8]Bavarian Cancer Research Center (BZKF), Munich, Germany

Abstract ***Objective*:** To present a long short-term memory (LSTM) network-based dose calculation method for magnetic resonance (MR)-guided proton therapy. ***Approach*:** 30 deformed planning computed tomography (CT) images of prostate cancer patients were collected for Monte Carlo (MC) dose calculation under a 1.5 T magnetic field. Proton pencil beams (PB) at 150, 175, and 200 MeV energies were simulated (6480 PBs at each energy). A 3D relative stopping power (RSP) cuboid covering the extent of the PB dose was extracted and flattened into the LSTM model, yielding a 3D predicted PB dose. Training and validation involved 25 patients; 5 patients were reserved for testing. ***Results*:** Test results of all PBs showed a mean gamma passing rate (2%/2 mm) of 99.89%, with an average center-of-mass discrepancy of 0.36 mm between predicted and simulated trajectories. The model inference time was 10 ms per beam. ***Conclusion*:** LSTM models for proton dose calculation in a magnetic field were developed and showed promising accuracy and efficiency for prostate cancer patients.

1 Introduction

Proton therapy has limitations in achieving high-precision dose delivery due to its inherent range uncertainty [1], which drives the need for online image guidance during proton dose delivery. Magnetic resonance imaging (MRI) has the advantages of high soft tissue contrast and non-ionizing radiation, and offers the potential for real-time anatomical imaging [2]. With the combination of real-time MRI and proton therapy, the treatment plan could be adapted or re-optimized based on the latest patient geometry before dose delivery, which would improve the target dose coverage and reduce the dose to surrounding normal tissues [3]. To make this online adaptive mode a reality, the plan adaptation and re-optimization require an accurate and fast proton dose calculation engine.
For proton dose calculation in magnetic fields, the Lorentz force causes protons to be deflected as they travel [4]. Monte Carlo (MC) methods, which compute the trajectories of individual particles by simulating particle transport physics, can consider the impact of magnetic fields on dose distributions but also require high computation times. To offer a trade-off between speed and accuracy, several analytical methods have been proposed to estimate the proton beam deflection and achieve accuracy close to MC methods in homogeneous tissues. Although the calculation time of recent MC and analytical methods can be accelerated by graphics processing units (GPUs) from several hours to several seconds [5, 6], the ultimate need for future real-time plan adaptation and re-optimization is still unmet, which requires dose engines to produce MC accuracy at millisecond speeds in the online iterative process.
Recently, several studies have shown the feasibility of deep learning techniques to realize millisecond proton dose calculations with MC accuracy. Some methods use computationally cheap physical inputs and computed tomography (CT) scans as inputs to U-Net convolutional models, which learn the correlation between physical inputs, CT scans and MC label dose [7, 8]; some other methods use long short term memory (LSTM) and transformer models, which are suitable for processing sequences with long dependencies, to learn the mapping of CT sequences to dose sequences in the beam's eye view (BEV) coordinate system [9, 10]. Although the two kinds of methods achieved good accuracy and high speed, to the best of our knowledge, they have not been applied to proton dose calculation scenarios under magnetic fields.
In this work, we applied a previously published LSTM proton dose calculation model [9] to dose calculation in a 1.5 T magnetic field, and confirmed its accuracy in patient cases with prostate cancer.

2 Materials and Methods

2.1 Proposed framework

The framework of this study is shown in Figure 1. First, based on the given initial energy and beam direction, a 3D relative stopping power (RSP) cuboid of interest containing the proton beam dose region is resampled from the patient's CT-derived RSP. Then, in the BEV coordinate system, the 3D cuboid of interest is treated as a sequence of 2D slices traveling from upstream to downstream. By flattening these

2D slices into 1D sequences, the directional LSTM unit can process the RSP sequences and generate the internal hidden states and outputs, where the former can be used as the input information for the subsequent slices. The outputs are passed to a fully-connected neural network to generate the predicted dose slices. Finally, the dose slices from our model are spliced back to the original 3D dose size and compared with the ground truth MC dose.

2.2 Dataset processing

We used a patient dataset which included deformed planning CT images after registration with MRI from 30 prostate cancer patients treated with a 0.35 T MR-linac (MRIdian, ViewRay, USA) at the Department of Radiation Oncology of the LMU University Hospital. All patients had no artificial implants and went through prior anonymization to comply with data protection law. The ground truth MC dose was generated through Geant4, and RSP maps with respect to water were then created within Geant4 and exported. To perform dose calculations in the presence of a magnetic field, a 1.5 T homogeneous magnetic field was set to be aligned with the z-axis (parallel to the superior direction) and act within a cylinder of 30 cm radius. The dose was scored on the same voxel geometry as the CT input, and all the proton beams were simulated with 10^6 histories, which made the maximal relative standard error in all regions above 10% of the maximum dose always below 1.15%.
Pencil beams (PBs) with energies of 150 MeV, 175 MeV and 200 MeV were simulated, with an energy spread of 0.83 MeV and laterally the beam was Gaussian with a standard deviation of 4.2 mm. PBs were isotropically sampled over a nearly cylindrical surface surrounding the prostate region and confined within a predefined angular sector to ensure no PBs were placed posterior or anterior to the patient. The propagation of the PBs was chosen to take place in the transverse plane, always orthogonal to the B field. Defining the right lateral direction as the 0° reference line, the sampling was done within $\alpha \in [40°, 140°] \cup [220°, 320°]$ region in $\Delta\alpha = 16.\dot{6}°$ steps, yielding 12 different angles per patient. At each of the 12 angles, a 3×6 PB grid was generated in the plane orthogonal to the PB's direction. In total, 216 PBs were simulated per patient and 6480 PBs for each energy.
After applying the body mask to RSP and dose maps, utilizing the transformation relationship between the patient and BEV coordinate systems, the 3D RSP-dose cuboid pairs were cropped and resampled according to the initial position and orientation of PBs. Cuboid sizes of $170 \times 43 \times 23$, $188 \times 49 \times 25$, and $205 \times 55 \times 25$ were used for the PBs with energies of 150 MeV, 175 MeV and 200 MeV, respectively. For each energy, 4320 PBs from 20 patients were used to train the model, 1080 PBs from 5 patients for validation and the remaining 5 patients with 1080 PBs for testing. Then, a global normalization was performed for the training, validation, and testing cuboid datasets using the maximum RSP and dose cuboid values of the training dataset. Finally, a reflection transformation across the xy plane was used, which doubled the original training dataset from 4320 PBs to 8640 PBs for each energy.

2.3 Model training

The network architecture used in this work corresponds to the LSTM model used in [9] for proton dose calculation in the absence of a B field in lung cancer patients. The LSTM features one layer with 1000 neurons as internal layer, followed by two fully connected layers with 100 neurons and ReLu activation layers.
For model training, we used the Adam optimizer to minimize the mean squared error (MSE) loss between the output dose and ground truth dose sequences. The initial learning rate was set to 10^{-5} and batch size to 8. The model was implemented in Python 3.8 and Pytorch 2.0.1 and trained with an NVIDIA RTX A6000 GPU (48 GB memory) for around 15000 epochs, which took about a week. The model with the best overall MSE validation loss was identified and used for the prediction in the test dataset. Each energy was trained separately.

2.4 Evaluation metrics

After obtaining model predictions, all dose voxels below a threshold of 0.01% of the global maximum were set to 0, similarly to [10]. Then, 198 PBs that overshot the patient were filtered out in the test dataset with energy of 200 MeV and all the output doses were reverse-normalized for the final evaluation.
Comparison metrics used in our study were according to the metrics in [11]. The 3D global gamma pass rate γ_{pr} was calculated to compare the 3D dose distributions, which was implemented within an open-source project, PyMedphys [12]. The distance-to-agreement threshold was set to 2 mm, the dose difference threshold to 2% and with a dose cutoff to 10% of the maximum dose. The relative error ε_{rel} was used to quantify the relative disagreement between two laterally integrated dose profiles at a certain depth. The integrated dose difference ε_{int} was evaluated to allow a comparison between PBs of different energies with each other. To compare the depth range of PBs, $\Delta R_{D_{80}}$ was used to calculate the difference of the depth of the 80% dose distal to the Bragg peak between two dose distributions. Finally, to analyze magnetic deflection modeling and the agreement of two PBs' lateral profiles at a given depth, the difference in the center of mass F_{COM} and full width at half maximum value FW_{50} of the PBs were calculated.

3 Results

Figure 2 (a) and (b) show a simulated and predicted PB from a 175 MeV energy model on a test patient. The LSTM can accurately predict the entry point and angle of the PB

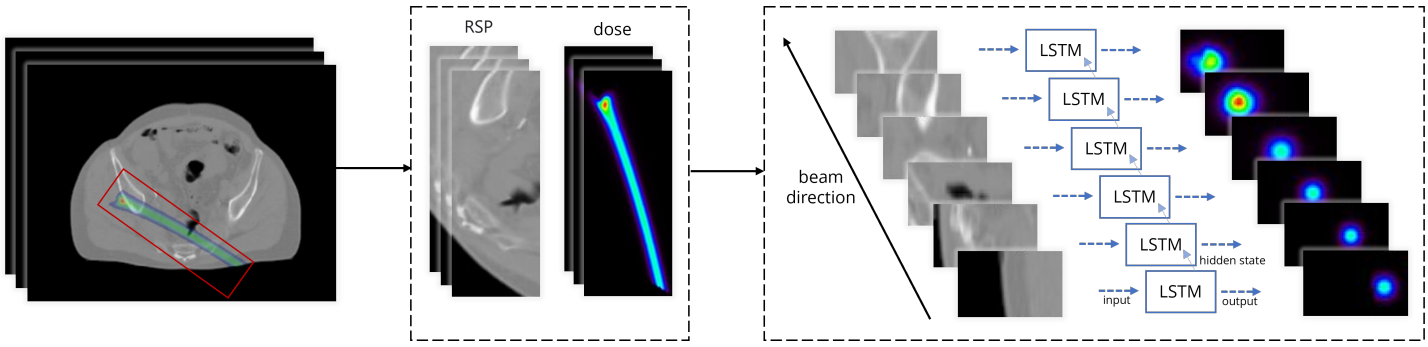
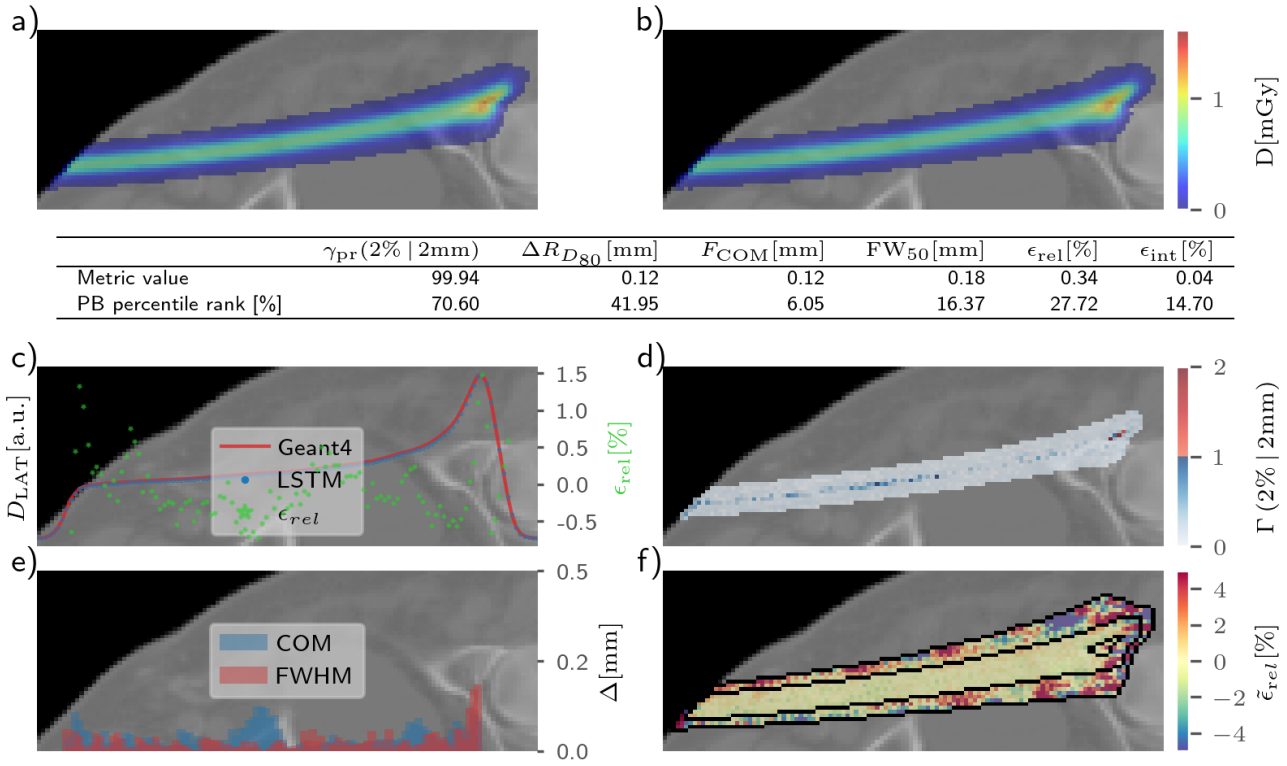


Figure 1: The framework of this study. (left) Dose distribution from a single pencil beam obtained from MC simulation with outline of 3D cuboid of interest. (middle) Resampling of input RSP and output dose for the cuboid of interest. (right) LSTM model converting sequence of RSP slices to a sequence of dose slices.



	$\gamma_{\mathrm{pr}}(2\% \mid 2\mathrm{mm})$	$\Delta R_{D_{80}}$ [mm]	F_{COM} [mm]	FW_{50} [mm]	ϵ_{rel} [%]	ϵ_{int} [%]
Metric value	99.94	0.12	0.12	0.18	0.34	0.04
PB percentile rank [%]	70.60	41.95	6.05	16.37	27.72	14.70

Figure 2: An example of a predicted 175 MeV PB in test patient. An *xy* slice of a) Geant4 simulated and b) LSTM predicted dose distribution is plotted over the RSP map input. The table shows the PB's performance regarding different metrics and percentile ranks with respect to all other PBs from the test set. A percentile rank of 0% corresponds to the best- and 100% to the worst-ranked PB. c) Integrated lateral dose profiles over depth and $\varepsilon_{\mathrm{rel}}$ of each slice. d) Γ (2%|2mm) index. e) F_{COM} and FW_{50} at each depth. f) Voxel-based relative dose difference of dose distribution integrated over z capped at ±5%. Black lines denote 80%, 10% and 1% isodose lines with respect to z integrated dose in the Bragg peak.

in the patient and the deflection of the PB caused by the magnetic field. In Figure 2 (c), laterally integrated profiles of the simulated and predicted PB are shown and agree well. The $\Delta R_{D_{80}}$ and mean $\varepsilon_{\mathrm{rel}}$ were less than 0.5 mm and 0.5%, respectively. The Γ index (2%|2 mm) map is shown in Figure 2 (d) and its corresponding γ_{pr} was 99.94%. To analyze the predicted PB trajectory and its lateral profiles, F_{COM} difference and FW_{50} difference are shown in Figure 2 (e), and were both less than 0.5 mm. In Figure 2 (f), the 2D distribution of relative dose differences integrated over z is displayed.

To further analyze the overall model prediction performance for the whole test dataset, Table 1 shows the worst and average values of each evaluation metric for the models at the three energies. The worst γ_{pr} (2%|2 mm, $D > 10\% D_{\mathrm{max}}$) was above 95.9%, the maximum $R_{D_{80}}$, F_{COM} and FW_{50} differences were 2.07%, 1.99 mm and 3.28 mm, respectively. High accuracy was achieved with respect to almost every metric we evaluated. For calculation time, the cuboid extraction and interpolation (not optimized) took 45 to 60 ms, while LSTM inference time was 10 ms; our method can thus achieve proton dose calculation per spot within 70 ms.

Table 1: The performance of all evaluated metrics summarized by worst and mean values across all five test patients combined and for each energy model individually. Red cells indicate the worst value.

	$\gamma_{(2\%\,\vert\,2\,\mathrm{mm})}$ [%]		$R_{D_{80}}$ [%]		F_{COM} [mm]		FW_{50} [mm]		$\vert\varepsilon_{\mathrm{rel}}\vert$ [%]		$\varepsilon_{\mathrm{int}}$ [%]
	min	mean	mean	max	mean	max	mean	max	mean	max	mean
E [MeV]											
150	95.90	99.90	0.27	1.60	0.40	1.99	0.45	2.62	0.60	4.55	0.26
175	96.64	99.88	0.25	1.85	0.37	1.46	0.46	3.28	0.54	2.75	0.24
200	98.16	99.89	0.17	2.07	0.31	1.76	0.45	2.54	0.46	2.30	0.19
Overall	95.90	99.89	0.23	2.07	0.36	1.99	0.45	3.28	0.54	4.55	0.23

4 Discussion

This study is an application of a previously developed architecture [9] to magnetic field dose calculation while using more extensive and rigorous criteria. To our knowledge, this work is the first one to utilize an AI architecture to speed up proton dose calculation in magnetic fields. Compared with recent transformer-based dose calculation methods [10], our study demonstrated that the LSTM architecture can also achieve state-of-the-art accuracy but has lighter trainable parameters; the GPU memory requirements of our model never increased above 1 GB and our network is easy to train and converged with a constant learning rate of 10^{-5}.

We can compare with recent analytical dose calculation methods for MRI-guided proton therapy [5], where maximum observed differences in water between their approach and MC ground truth were reported. For $\Delta R_{D_{80}}$, they reported 0.7 mm or 0.22% for the highest tested energy of 229 MeV. Furthermore, the maximum reported F_{COM} was 1.2 mm. In contrast, we reported for the highest energy model of 200 MeV a mean $\Delta R_{D_{80}}$ of only 0.17 mm and mean F_{COM} between 0.3 and 0.4 mm. This was in real test patients rather than in homogeneous water. Reference [6] presented a recent GPU-based MC algorithm for MRI-guided proton therapy called ARCHER. They report achieving a $\Delta R_{D_{80}} < 0.06$ mm and a mean $\varepsilon_{\mathrm{rel}}$ of 0.81% in water. For tissue and bone materials, which were also evaluated, they reported a mean $\varepsilon_{\mathrm{rel}}$ of 1.12% and $F_{\mathrm{COM}} < 0.05$ mm. Their 3D-gamma pass rate with a 2 mm/2% criterion in the region with dose greater than 10% of the maximum dose was over 99% for all tested cases. Although the computational time of ARCHER is found to range from 0.82 to 4.54 s for 10^7 proton histories, the potential advantage of the deep-learning method is that we can use more particles to simulate MC doses with lower statistical uncertainty as labels without increasing the prediction time. Finally, just like a MC engine has to be carefully adapted and commissioned for use in a specific treatment setting, a deep-learning model would have to be trained or at least fine-tuned. Unique settings and beam profiles of a specific machine necessitate a specific model per machine.

5 Conclusion

We successfully developed an LSTM network-based proton dose calculation method in a magnetic field and confirmed its accuracy for prostate patients. This sub-second dose engine has the potential to improve the efficiency of online plan optimization in the MR-guided proton therapy workflow.

References

[1] H. Paganetti. “Range uncertainties in proton therapy and the role of Monte Carlo simulations”. *Physics in Medicine & Biology* 57.11 (2012), R99.

[2] M. A. Schmidt and G. S. Payne. “Radiotherapy planning using MRI”. *Physics in Medicine & Biology* 60.22 (2015), R323.

[3] A. Hoffmann, B. Oborn, M. Moteabbed, et al. “MR-guided proton therapy: a review and a preview”. *Radiation Oncology* 15.1 (2020), pp. 1–13.

[4] J. Hartman, C. Kontaxis, G. Bol, et al. “Dosimetric feasibility of intensity modulated proton therapy in a transverse magnetic field of 1.5 T”. *Physics in Medicine & Biology* 60.15 (2015), p. 5955.

[5] A. Duetschler, C. Winterhalter, G. Meier, et al. “A fast analytical dose calculation approach for MRI-guided proton therapy”. *Physics in Medicine & Biology* 68.19 (2023), p. 195020.

[6] S. Li, B. Cheng, Y. Wang, et al. “A GPU-based fast Monte Carlo code that supports proton transport in magnetic field for radiation therapy”. *Journal of Applied Clinical Medical Physics* 25.1 (2024), e14208.

[7] G. Zhang, X. Chen, J. Dai, et al. “A plan verification platform for online adaptive proton therapy using deep learning-based Monte–Carlo denoising”. *Physica Medica* 103 (2022), pp. 18–25.

[8] C. Wu, D. Nguyen, Y. Xing, et al. “Improving proton dose calculation accuracy by using deep learning”. *Machine learning: science and technology* 2.1 (2021), p. 015017.

[9] A. Neishabouri, N. Wahl, A. Mairani, et al. “Long short-term memory networks for proton dose calculation in highly heterogeneous tissues”. *Medical physics* 48.4 (2021), pp. 1893–1908.

[10] O. Pastor-Serrano and Z. Perkó. “Millisecond speed deep learning based proton dose calculation with Monte Carlo accuracy”. *Physics in Medicine & Biology* 67.10 (2022), p. 105006.

[11] P. Lysakovski, A. Ferrari, T. Tessonnier, et al. “Development and benchmarking of a Monte Carlo dose engine for proton radiation therapy”. *Frontiers in Physics* (2021), p. 655.

[12] S. Biggs, M. Jennings, S. Swerdloff, et al. “PyMedPhys: A community effort to develop an open, Python-based standard library for medical physics applications”. *Journal of Open Source Software* 7.78 (2022), p. 4555.