An AI dose engine for fast carbon ion treatment planning.

Quarz[1,2], A. De Gregorio[3], G. Franciosini[4,6], A. Schiavi[4,6], Z. Perkó[5], L. Volz[1], V. Patera[4,6] , M. Durante[1,7, 8] and C. Graeff[1,2]

[1]GSI Helmholtzzentrum für Schwerionenforschung, Biophysics Department, Darmstadt, Germany
[2]Technische Universität Darmstadt, Department of Electrical Engineering and Information Technology, Darmstadt, Germany
[3]Dipartimento di Fisica, Sapienza Università di Roma, Italy
[4]Dipartimento di Scienze di Base e Applicate per l'Ingegneria, Sapienza Universita` di Roma, Roma, Italy
[5]Delft University of Technology, Department of Radiation Science and Technology, Delft, The Netherlands
[6]INFN (Istituto Nazionale di Fisica Nucleare) section of Roma 1, Rome, Italy
[7]Technische Universität Darmstadt, Department of Condensed Matter Physics, Darmstadt, Germany
[8]University Federico II, Department of Physics "Ettore Pancini", Naples, Italy
E-mail: c.graeff@gsi.de

Keywords: Treatment planning, dose engine, RBE-weighted dose, deep-learning, carbon therapy

Abstract

Purpose

Monte Carlo (MC) simulations provide gold-standard accuracy for carbon ion therapy dose calculations but are computationally intensive, limiting their use in adaptive workflows. Analytical pencil beam algorithms offer speed but reduced accuracy in heterogeneous tissues. This study develops the first AI-based dose engine capable of predicting absorbed dose and the α and β parameters for relative biological effectiveness (RBE)–weighted optimisation in carbon ion therapy, delivering MC-level accuracy with drastically reduced computation time.

Materials and Methods

We extended the transformer-based DoTA architecture to predict absorbed dose (C-DoTA-d), α (C-DoTA-α), and β (C-DoTA-β), introducing a cross-attention mechanism for α and β to combine dose and energy inputs. The training dataset consisted of roughly 70,000 pencil beams from 187 head-and-neck patients, with ground-truth values obtained using the GPU-accelerated Monte Carlo toolkit FRED. Performance was evaluated on an independent test set using gamma pass rate (1%/1 mm), depth-dose, and isodose contour Dice coefficients. Monte Carlo dropout–based uncertainty analysis was performed.

Results

Median gamma pass rates exceeded 98% for all predictions (99.76% for dose, 99.14% for α, 98.74% for β), with minima above 85% in the most heterogeneous anatomies. The Dice coefficient was 0.95 for 1% isodose contours, with slightly reduced agreement in high-gradient regions. Compared to MC FRED, inference was over 400× faster (0.032 s vs. 14 s per pencil beam) while maintaining accuracy. Uncertainty analysis showed high stability, with mean standard deviations below 0.5% for all models.

Conclusions

This AI-based dose engine achieves MC-quality predictions of absorbed dose and RBE model parameters in ~30 milliseconds per beamlet. Its speed and accuracy support online adaptive planning, paving the way for more effective carbon ion therapy workflows. Future work will expand to additional anatomical sites, beam geometries, and clinical beamlines.

1. Introduction

Radiation therapy treatment planning depends on accurate and efficient dose calculation to ensure optimal tumour targeting while sparing surrounding healthy tissues. In particle therapy, this is particularly complex because of the sharp dose distribution of ion beams, which exhibit a steep maximum at the end of their range in the patient, known as the Bragg peak.

For carbon ion therapy, the ions' variable relative biological effectiveness (RBE) - which also shows a maximum at the Bragg peak location - introduces further complexity. Accurate dose calculation in carbon ion therapy, therefore, not only requires a precise beam model, but it also relies on the correct calculation of biological factors that depend on the spectrum of primary and secondary particles along the beam path in the patient [1,2].

Monte Carlo (MC) particle transport simulations are generally considered the gold standard for dose calculation. By considering the physics processes involved in the beam interaction with the patient, MC simulations accurately produce the lateral broadening and range mixing of the pencil beams. Recent advances in GPU-accelerated MC codes have allowed increasing introduction of MC dose engines in commercial treatment planning systems for proton therapy. However, the increased complexity of particle transport for carbon ions compared to protons – requiring the simulation of multiple fragment species along the primary ion path and the associated computational expense still render MC carbon dose engines impractical for clinical routine. Computation time for even the fastest currently available carbon ion GPU MC codes is a significant bottleneck, particularly in scenarios requiring many or rapid plan optimisations, such as for robust 4D optimisation [3] and adaptive treatment workflows [4,5].

Due to this limitation, all carbon radiotherapy treatments are still planned using variations of analytical pencil beam algorithms (PBAs) [6,7]. However, these algorithms make inherent assumptions that drastically limit their accuracy in heterogeneous tissues [8]. Several studies have compared PBAs with MC methods, highlighting trade-offs between speed and precision in proton [9] and carbon [10,11] therapy. For high-precision treatment planning that fully exploits the accuracy available with carbon ion beams, MC-quality dose engines are essential.

For proton dose calculation, artificial intelligence (AI) has been successfully applied to achieve MC accuracy at massively reduced runtimes, sometimes even outperforming traditional PBA speed. Approaches such as DoTA [12], LSTM-based dose prediction [13,14], and DiscoGAN [15] have demonstrated the feasibility of rapid dose calculations of single pencil beams (PB) as well as for the full dose distribution [16]. Recent studies have also explored the use of AI to enhance the quality of PBA for protons [17,18].

The greater complexity of carbon ions due to their higher linear energy transfer (LET) and nuclear fragmentation presents a challenge that is not yet considered in any of the available AI dose engines. As any carbon therapy treatment planning system (TPS) needs to optimise the RBE-weighted dose, carbon ion dose engines must include the relevant biological parameters used in the RBE-weighted dose calculations, namely the alpha and beta parameters in the linear quadratic model of cell survival [6].

Previous works in carbon ion dose prediction [19,20] have focused solely on physical dose, omitting the biologically crucial RBE component. To our knowledge, this study represents the first attempt to fill in this gap by developing an AI-driven dose engine capable of jointly predicting physical dose, and α and β parameters for RBE-weighted optimisation.

We extended the transformer-based DoTA [12] model, which recently achieved impressive results for proton dose calculations, and trained three models separately predicting the input parameters for the local effect model (LEM I) [21]: namely, the absorbed dose (C-DoTA-d), alpha (C-DoTA-α) and beta (C-DoTA-β). Our model operates on a per-pencil-beam (PB) basis, enabling both optimisation and forward dose calculation.

In this proof-of-concept demonstration, we assess our model's feasibility on a cohort of head and neck cancer patients formerly treated within a carbon ion therapy pilot study at GSI (Darmstadt, Germany). As a baseline for comparison and to generate training data, we utilise the GPU MC toolkit FRED [20,22], which can score the alpha and beta parameters alongside the physical dose at fast per-pencil-beam calculation speeds of only a few seconds. The models deliver all three input parameters in only 32 ms, with a median gamma pass rate (GPR) of a gamma analysis with 1%/1 mm criteria and a 10% threshold >99%.

2. Materials and methods

2.1 Patient *data*

Treatment planning data were taken from the GSI carbon ion pilot project for a cohort of head and neck cancer patients treated between 1997 and 2008 at GSI with carbon ions. The pilot project was approved by the ethical committee of the University of Heidelberg in 1997. Anonymised treatment plans of all patients are stored for research purposes at GSI. Informed consent is waived by the ethical committee of the University of Heidelberg for anonymised plans used for research purposes.

For this study, 225 head-and-neck cancer patients were randomly selected from the database. For each, a new plan was generated in a scripted manner based on the original treatment plans, keeping the irradiation target unchanged, providing a starting point for the data production. The gantry angle was fixed as it was during the pilot project at GSI. For this study, a value of 0° was used across all patients to simplify model evaluation and debugging.

Instead, the couch angles were randomised in the interval $[-\phi_{Couch}, +\phi_{Couch}]$, with the range $\phi_{Couch} \in [85°, 105°]$ being patient dependent to introduce geometric variability. This approach helped mitigate training errors arising from air gaps between the patient and the immobilisation masks in the dataset. Within each ϕ_{Couch} five negative and five positive random couch angles were selected. The beam angles that contained part of a shoulder or mask mounting structure, or for which the plan optimisation was not possible due to the restricted energy range, were excluded from the 10 sampled angles. For the remaining angles, treatment plans were optimised with the in-house TPS TRiP98 [6,23] using base data for the CNAO facility in Pavia (Italy). For each successfully optimised field, multiple beam-eye-view computed topographies (BEV CTs) were sampled with a regular grid at a lateral plane with a 10 mm step. Each BEV CT with the size of 48x48 pixels laterally till the end of the CT at 1 mm resolution was centred on a PB of the plan and interpolated, spanning the extent of the field.

Two PBs of a randomly selected energy in a range of 115 to 260 MeV/u were generated for each BEV CT, adhering to the discrete beam energies available at CNAO. If pre-calculation in TRiP98 revealed that the Bragg Peak was placed outside the patient's body, the maximum possible energy range for this BEV CT was reduced accordingly, and a new beam energy was sampled. To guarantee training quality, the random energy selection was guided to reject energies already over-represented in the database. For each discrete energy level, the number of samples already generated in the training dataset was tracked. If, for a newly sampled energy, its sample count exceeded 99% of the maximum number of samples among the energy layers, the energy was resampled. Finally, as a data curation step, all BEV-CTs were automatically processed to remove air proximal to the patient.

The train test split was performed on patients rather than at a beam level, to ensure fully independent datasets. A total of 187 patients were used for training and validation, providing a dataset of 34706 BEV CTs for training and 1769 for validation (corresponding to a 95%/5% training/validation split), following the data curation step. As for each BEV CT, 2 energies were sampled, a total of 69418 pencil beams were used for training, and 3538 for validation. The

remaining 38 patients were used to create 6963 BEV CTs with 2 PB energies each, yielding a test dataset of 13926 PBs.

For each BEV-CT, the two PBs were simulated in the MC FRED [20] to create labels to train the models. The FRED MC simulation setup is explained in the next section.

2.2 FRED

Ground truth data were generated with the GPU-accelerated Monte Carlo particle transport software FRED [20,22]. FRED is capable of fast dose calculations for protons and light ions. In particular, this work relies on the recently implemented carbon ion extension of FRED [20].

The simulation for carbon ions considers three main physics blocks: ionisation energy loss, multiple scattering (both implemented analogously to the proton transport models of FRED) and a phenomenological fragmentation model. The latter was developed directly from published fragmentation data and evaluated against this data as well as FLUKA simulations. Full details on the physics processes in the simulation are provided in the overview paper by De Simoni et al [20]. In addition to the absorbed dose, FRED supports RBE-weighted dose calculation with multiple scoring options. In this work, we relied on table-based RBE calculation using tabulated alpha and beta values as a function of energy for the first eight light ion species (up to Oxygen) generated with LEM I. Internally in FRED, the effect of mixed ion fields is resolved by dose averaging [24], in the same way as TRiP98:

$$alpha = \frac{\sum_{i=1}^{n} a_i D_i}{\sum_{i=1}^{n} D_i}$$

$$beta = \left(\frac{\sum_{i=1}^{n} \sqrt{\beta_i} D_i}{\sum_{i=1}^{n} D_i} \right)^2$$

Where α_i and β_i are the energy-specific alpha and beta values for a specific ion type, n is the total number of ion species considered. D_i is the dose contribution from each ion species. The mixed alpha and beta values were saved together with the absorbed dose. In the TRiP98 treatment planning system, these values are used alongside the physical dose to calculate the RBE-weighted dose. Hence, they present the target values for the C-DoTA-α and C-DoTA-β models, respectively, while the absorbed dose output was the target for the C-DoTA-d model.

The simulation geometry consisted of the interpolated BEV CT, positioned so that its proximal entrance was located at the isocentre. The beam nozzle was simulated as a PMMA slab of 4mm thickness, positioned 10 cm in front of the BEV CT in beam direction. The beam was generated 20 cm in front of the isocentre. Each PB was simulated as a Gaussian beam with full width half maximum (FWHM) according to the CNAO beam characteristics at the respective PB energy. For all PBs, a momentum spread of 0.5% was assumed. Note that this does not reflect the characteristics of the CNAO beamline. A fully realistic simulation would introduce slight differences in the form of the beam's lateral spread (wider due to air drift) and range straggling (slightly broader). The simplified beam geometry was chosen to permit efficient training data generation. Each PB simulation used 10^7 primaries for high-quality ground truth data. Dose, alpha, and beta were scored on the same grid as the BEV CT. FRED provides the dose in Gy/primary.

2.3 Model

Figure 1 represents the C-DoTA architectures. For physical dose prediction, the vanilla DoTA architecture proved optimal for carbon ions, utilising normalised beam's-eye-view (BEV) CT scans and energy information as inputs. However, for α and β predictions, we extended the model by incorporating:

1. An additional convolutional encoder and transformer block processing both energy data and the output from the C-DoTA-*d* network (using a 1% dose threshold below which doses were zeroed out),
2. A cross-attention block to merge the BEV CT and dose information streams,
3. A final decoder block for prediction output.

Other model design choices in the original DoTA, such as the choice of convolutional and transformer block setups, hyperparameters (including learning rate schedules) and the optimiser configuration, were found to be the most optimal for the study.

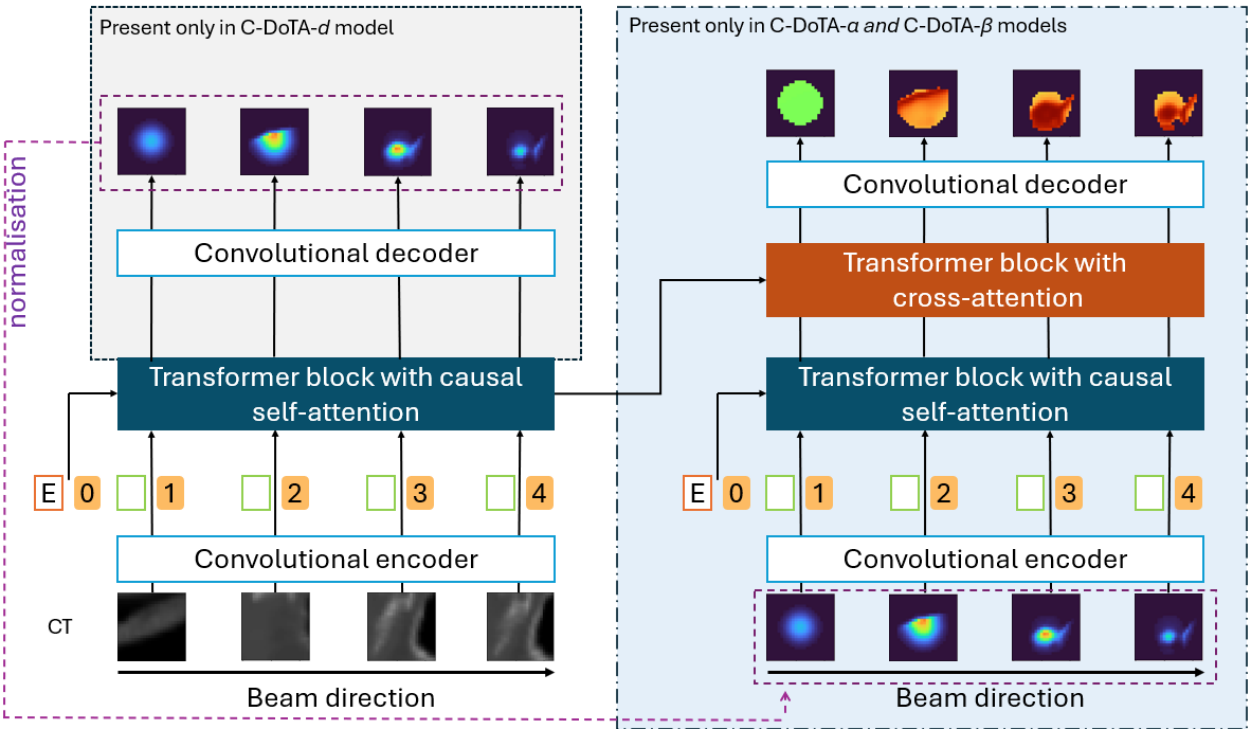

Figure 1 C-DoTA architectures. Left: Vanilla DoTA, with a convolutional encoder processing CT as a sequence of 2D images along the beam's travel, and the [0,1] normalised energy E inputed as the 0th token before the transformer block. Right and bottom: Modified DoTA architecture for predicting α and β parameters. Two separate convolutional encoders process the CT and predicted dose as input, and the normalised energy is appended to both sets of tokens before 2 separate transformer blocks process them. The two sets of output tokens are then processed via a third transformer block using cross attention. The convolutional encoder for the CT is retrained along with all other parts of the C-DoTA-α and C-DoTA-β models.

The loss function used for the vanilla DoTA, however, was not sufficient for the model to learn the sharper gradients at the carbon Bragg peaks. Thus, the C-DoTA models were first trained for 2

times with 28 epochs each using only a mean squared error (MSE) loss (similar to the training schedule of the vanilla DoTA model), then were further trained 2 more times with 28 epochs each using a custom loss. The learning rate was rescheduled after each block of 28 epochs to prevent convergence in a local minimum. Unlike the vanilla DoTA model, the C-DoTA models do not use normalised values as labels. We noticed that normalisation to the typical [0,1] range causes blurry output, especially near Bragg peak regions. Instead, we scaled the FRED labels in our loss function by multiplying the raw values as follows: absorbed dose values by the number of simulated primaries, alpha values by 10 and beta values by 10^3 (as the latter two values are independent of the particle number), both for the training and evaluation data. It resulted in the following value ranges for the absorbed dose (0 – 2.6 Gy), alpha (0 -15.8 Gy^{-1}), beta (0 – 26.5 Gy^{-2})

The custom loss function was designed based on MSE with three additional components, where y is the ground truth (GT) value and $\tilde{y}$ is the predicted value:

Masked MSE:

$$Masked\ MSE\ = \frac{1}{n} \sum_{i:\,|y_i - \tilde{y}_i| > \epsilon^2} (y_i - \tilde{y}_i)^2$$

where n is the cardinality of the set of voxels with a squared error above the threshold ϵ^2 (0.001 Gy for dose, 1/Gy for alpha and $1/Gy^2$ for beta) at the given epoch. Unlike the traditional MSE loss, this masked loss excludes a high number of voxels with very small differences, forcing the model to pay more attention to the regions with high squared dose error, i.e., emphasising the narrow Bragg Peak where gradients are high.

Depth-wise partial gradient loss

$$Depth\ Loss\ = \frac{1}{m} \sum_{i}^{m} \left(\frac{\partial y_i}{\partial z} - \frac{\partial \tilde{y}_i}{\partial z} \right)^2$$

where the partial derivatives of the true and predicted responses with respect to depth z for each voxel i are calculated as

$$\frac{\partial y_i}{\partial z} = \frac{y_{j:\, z_j = z_i + \Delta z} - y_i}{\Delta z}$$

$$\frac{\partial \tilde{y}_i}{\partial z} = \frac{\tilde{y}_{j:\, z_j = z_i + \Delta z} - \tilde{y}_i}{\Delta z},$$

with j being the index of the voxel in the next depth layer adjacent to voxel i , Δz is the depth spacing and m is the total number of voxels.

Region masked MSE

$$Region\ masked\ MSE\ = \frac{1}{l} \sum_{i:y_i > \theta} (y_i - \tilde{y}_i)^2$$

where *l* is the cardinality of the set of voxels where GT values are above the threshold θ (60% of the sample maximum) at the given epoch. The loss increases the weight of the error in the regions with high values.

The final loss is defined as the weighted sum of the separate MSE losses with the weight factors listed in Table S 1:

$$Custom\ loss = MSE + w_{Mask} \cdot Masked\ MSE + w_{Depth} \cdot Depth\ Loss\ MSE + w_{Region} \cdot Region\ masked\ MSE$$

To find the optimal weight, we trained models with different sets of weights in the interval [0,1] and a step of 0.2. The optimal weight values were chosen based on the evaluation of the validation dataset. Not all the weights were used in all models.

2.4 Data preprocessing

After the simulation, each beamlet was cropped symmetrically to 24x24x240 mm^3. To prevent training on the noise at the beam penumbra, dose values are below 1% of the maximum absorbed dose as obtained in the MC simulation for the sample. This mask, further denoted MC mask, was applied exclusively during C-DoTA-d training. For the correct functioning of the model as a carbon ion therapy dose engine, it is crucial that every voxel with a non-zero predicted dose is paired with α and β values. Therefore, to ensure the model output remains consistent between C-DoTA-d, C-DoTA-α and C-DoTA- β, instead of the MC mask, a mask was generated from a 1% threshold of the C-DoTA-d prediction (further denoted as “AI dose mask”). This mask was used for the dose input and labels of C-DoTA-α and C-DoTA-β. Figure S 1 graphically illustrates the data workflow for a clearer understanding of the applied masks.

For easier handling, we normalized input data to the [0,1] range from their original raw ranges, which for CT data was [-1024 HU, 3071 HU], for beam energy was [115 MeV/u, 260 MeV/u], and for C-DoTA-d output (when it was used as input to C-DoTA- α and C-DoTA- β) was [0 Gy, 2.61 Gy].

2.5 Evaluation

As evaluation metrics, we used the gamma pass rate from a gamma analysis with 1%/1 mm criteria and a 10% threshold. The 10% threshold was applied to mitigate the influence of the metric in low-dose regions, as low values artificially inflated the pass rate. In the results, the GPR was computed only for voxels where the C-DoTA-d prediction exceeded 1% of its maximum, i.e., within the AI dose mask. This was necessary due to the impact of the AI dose mask on the alpha and beta predictions. The GPR based on the MC mask for the MC label can be found in the supplementary materials.

Furthermore, the consistency between prediction and GT was checked with dice coefficients [25] by calculating iso-dose contours for C-DoTA-d at 1%, 30% and 70% of the sample maximum. The maximum of the GT sample was used to calculate relative metrics such as MSE. To estimate the impact of voxel exclusion due to differences in the 1% dose level between AI and MC, we counted voxels present in the MC mask but not in the AI dose mask and vice versa.

In addition to 3D-dose analyses, integrated depth-dose (IDD) profiles were assessed. Unlike GPR for IDD metrics, the MC mask was applied to GT, and the AI dose mask to the AI output to calculate the carbon range R80 of the Bragg peak, relative dose difference (ΔDmax) and MSE.

To automatically compute the carbon range R80 of the Bragg peak, the Python Shapely package was used to find the interpolated intersection value at 80% of the sample maximum. This allowed us to evaluate the range accuracy of the model dose output.

To estimate the model uncertainty distribution, we applied Monte Carlo dropout (MCD) [26]. MCD is an efficient method for estimating the stability of a model's performance through multiple forward inferences with an activated dropout layer, thereby simulating multiple trained models. The evaluation was performed with a dropout rate of 0.2 on the complete prediction chain (and the dropout layers for both training and inference were the same as in the vanilla DoTA model).

In our setup, the alpha and beta predictions are strongly correlated to the quality of the C-DoTA-d dose input. In this work, we evaluated the propagated uncertainty using the MCD C-DoTA-d outputs performed with activated dropout layers. It means that the output of C-DoTA-d is used as

input for C-DoTA-α and C-DoTA-β models. After 30 runs, the average value and standard deviation per sample were analysed. For the graphic representation of the MCD-based uncertainty, a standard dose prediction was used as input to show the uncertainty regions of the models only. Quantitative statements on model uncertainty were obtained by evaluating the gamma passing rate of the 30 individual predictions.

3 Results

Table 1 shows the results of GPR 1%/1 mm with the 10% threshold applied to the output of all three C-DoTA models (absorbed dose (D), alpha (α) and beta (β)). The median value for all three outputs is around 99%. 5%-tile values are 98.09% for absorbed dose, 95.60% for alpha and 95.73% for beta, which shows that most of the data is close to the expected values. The results for the MC mask are presented in the Supplementary Materials.

Table 1 GPR 1%/1 mm results. AI dose mask was applied to both GT and AI output.

Parameter	**Value range**	**Gamma pass rate (1%/1 mm) [%]**				
		Median	Min	10%-tile	90%-tile	Max
D	0-2.6 Gy	99.76	92.06	98.70	99.96	100
α	0-1.6 Gy^{-1}	99.14	86.36	96.72	99.87	100
β	0-0.03 Gy^{-2}	98.74	85.26	96.71	99.45	99.87

The top of Figure 2 shows the output of the best absorbed dose prediction with GPR 1%/1 mm of 100% as well as the corresponding alpha (100%) and beta (99.87%) predictions. In contrast, the bottom one shows the difference between GT and the prediction, as well as the corresponding MCD uncertainty maps. The top of Figure 3 shows a test sample with the worst GPR for physical dose prediction, at 92.06%, with alpha at 89.16% and beta at 90.62%. The bottom part illustrates its difference and MCD uncertainty. The worst-case sample has a highly inhomogeneous structure, featuring bones and a complex nasal cavity, where the Bragg peak is positioned exactly on the boundary between the regions of high and low density, resulting in a complex shape of the beamlet. The main uncertainty region for the dose is around the Bragg peak region for all the models, where the gradient is sharp. This somewhat correlates with the difference in the best-case scenario; however, no such relation is seen in the worst-case scenario.

The worst case of alpha (Figure S 4) and beta (Figure S 5) predictions can be found in the supplementary materials.

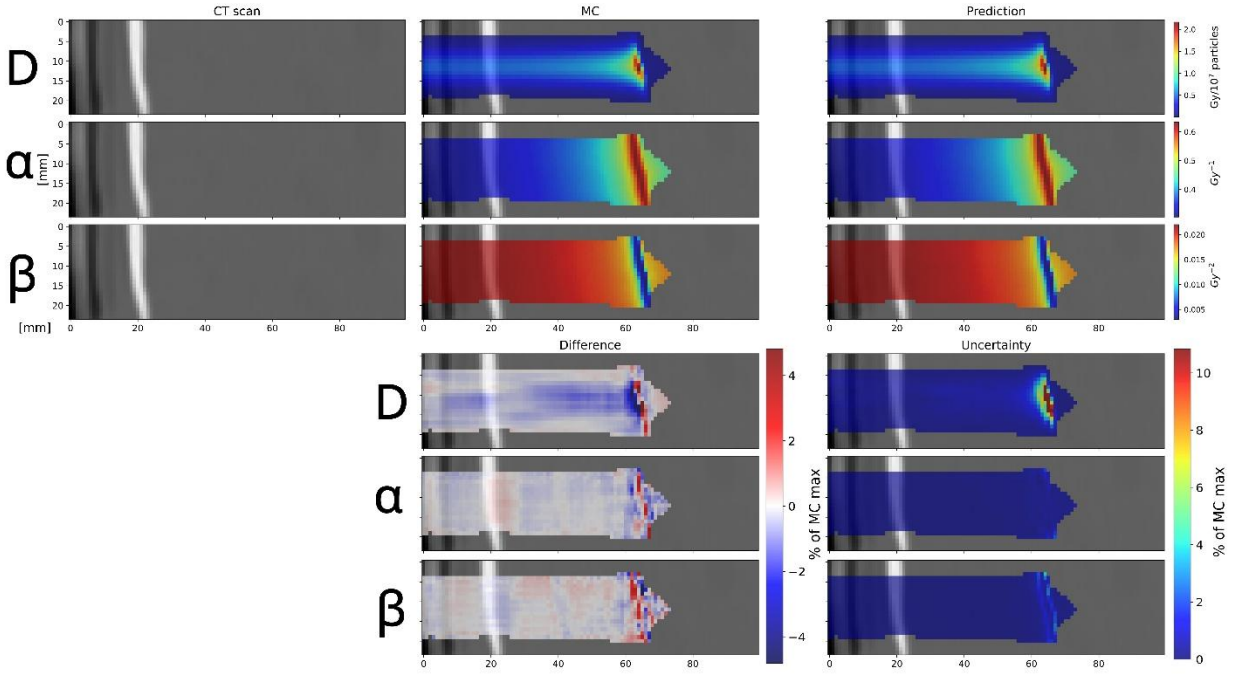

Figure 2 ***Best result in terms of the GPR for the physical dose. The GPR with a 1%/1 mm was top row absorbed dose (D): 100%, middle row alpha (α): 100%, bottom row beta (β): 99.87%. The top part: on the left, a slice of the BEV CT is shown; in the centre, the MC ground truth is shown; and the C-DoTA predictions on the right. The bottom part: on the left side, the difference (MC minus C-Dota); on the right side, MC Dropout uncertainty. For better visibility, the maximum and minimum values of the colour scale in the difference plots were set to the maximum between the absolute values of the 99.9%-tile and the 0.1%-tile.***

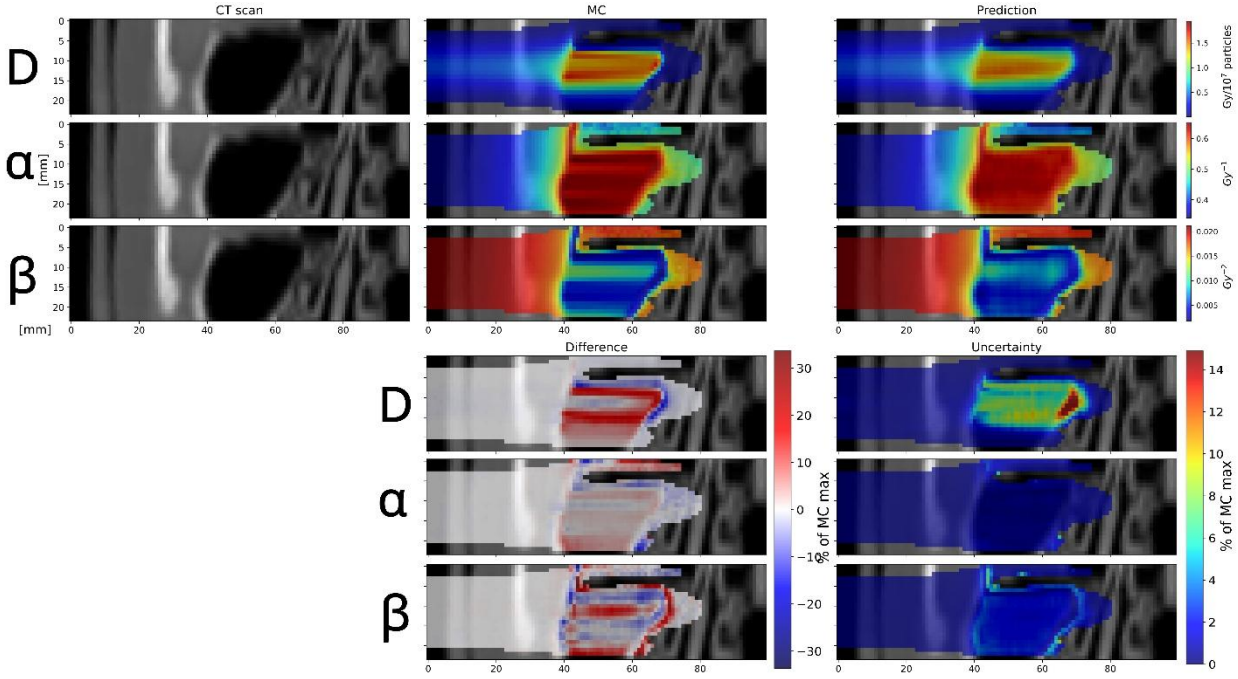

Figure 3 Worst result in terms of the GPR for the physical dose. The GPR with a 1%/1 mm was top row absorbed dose (D): 92.06%, middle row alpha (α): 89.16%, bottom row beta (β): 90.62%. The top part: on the left, a slice of the BEV CT is shown; in the centre, the MC ground truth is shown; and the C-DoTA predictions on the right. The bottom part: on the left side, the difference

(MC minus C-Dota); on the right side, MC Dropout uncertainty. For better visibility, the maximum and minimum values of the colour scale in the difference plots were set to the maximum between the absolute values of the 99.9%-tile and the 0.1%-tile.

To illustrate the effect of the custom loss, the results of the GPR 1%/1 mm for the models trained for 56 and 112 epochs on MSE can be found in the Supplementary materials (

Table S 3 and Table S 4). While the physical dose prediction does not benefit much in GPR, the alpha and beta have gained ≈18% in minimum values after 112 epochs. The additional 56 epochs improved the GPR by further ≈3% at minimum.

The result of the MC dropout uncertainty prediction is summarised in Table 2 in the form of the metrics for the GPRs evaluated for each of the individual 30 MCD predictions. The median of the mean standard deviation was below 0.5% for all the models, which shows high reliability of the prediction.

Table 2 Mean (top row) and standard deviation (bottom row) of GPR per sample after 30 runs with MC dropout of 0.2. AI dose mask was applied to both GT and AI output.

Parameter	Value range	Gamma pass rate (1%/1 mm) [%]				
		Median	Min	10%-tile	90%-tile	Max
D	0-2.6 Gy	99.59	92.62	96.62	99.89	100
		0.18	0.01	0.06	0.67	3.49
α	0-1.6 Gy^{-1}	98.62	86.97	96.42	99.57	99.96
		0.47	0.04	0.15	1.52	6.28
β	0-0.03 Gy^{-2}	98.20	85.59	96.12	99.05	99.52
		0.39	0.05	0.16	1.42	9.57

Table 3 presents a more detailed analysis of the test dataset. The mode of the range difference is almost 0 mm, while the automatic detection of R80 causes some extreme differences. The median relative difference in the dose maximum is 0.9%. The RMSE indicates that the average difference between the GT and prediction is 0.63%, or cumulatively, 0.62 $Gy/10^7$ primaries per 1 mm depth.

Examples of the extremes are shown in Figure S 6. The left plot depicts the maximum difference in R80, while the middle one shows the minimum difference. It is clearly visible that the predicted IDD profile closely resembles the MC one, but due to its complex shape, the R80 value obtained is highly sensitive to minor changes, in this case causing a shift between the first and second peak in the IDD profile. In the right plot, the dose loss in the Bragg peak region is caused by a

very bony CT structure, which also poses a challenge for PBA algorithms to calculate. The corresponding C-DoTA outputs can be found in Figure S 7.

The evaluation of the Dice coefficient between MC and C-DoTA predictions indicates a high level of agreement between the isodose lines. A small number of voxels primarily caused low values in some samples, where the relative impact of a small error is greater. The 10 and 90 percentile values are close to the median, demonstrating the stability of the C-DoTA results.

There is also a fraction of voxels that, due to the threshold, are present in GT but not in the prediction, and vice versa. This fraction of the lost voxels due to thresholding is 8% with mean of 0.02 Gy/10^7 primaries and a standard deviation of 0.003 Gy/10^7 primaries.

As an additional test, not included in the tables, we checked the behaviour of our models on some extreme cases such as a lot of air in the nasal cavities due to a possible surgery before the therapy or for beam paths that pass many bone-tissue interfaces. This data is presented in the supplementary material to this manuscript (Figure S 8).

Finally, we analysed the speed of the C-DoTA dose engine versus the FRED GPU MC. The time measurements were performed on a GPU Node with 8 AMD Radeon Instinct Mi100 32 GB video cards, 96 CPUs and 384 GB RAM. Only 1 out of 8 GPUs was used, and the whole node was still blocked for other users to obtain a clean measurement of the runtime. The time for C-DoTA, including loading the samples into the model and dose pre-processing for alpha and beta models (but not the interpolation for obtaining the BEV CT), was 0.032±0.001 s (0.007±0.0002 s for the absorbed dose only). The samples were randomly selected from the test dataset. C-DoTA is capable of predicting>400 times faster than MC FRED (14 s) for our setup.

*Table 3 Additional evaluation metrics. *-the values calculated with AI dose mask applied*

		Median	Min	10%-tile	90%-tile	Max
IDD Curve	ΔR80 mm	-0.02	-24.48	-0.44	0.32	26.20
	ΔDmax %	0.90	-11.51	-2.25	4.32	18.38
	RMSE %*	0.97	0.25	0.65	1.61	6.12
FRED Vs. C-Dota	RMSE %*	0.14	0.05	0.09	0.23	0.63
	Dice (1% isodose) (n voxels in GT)	0.95 (19241)	0.84 (11531)	0.93 (14455.5)	0.96 (24566)	0.98 (38586)
	Dice (30% isodose) (n voxels in GT)	0.90 (640)	0.29 (301)	0.77 (418)	0.96 (2016.5)	0.99 (5907)
	Dice (70% isodose) (n voxels in GT)	0.83 (39)	0.16 (15)	0.68 (29)	0.92 (77)	1 (499)
Voxels presented in GT and not presented in the prediction	Count	1479	292	983	2416	8530
	(in %)	8	1.32	5.48	11.42	23.38
	Total voxels in GT	19240.5	11531	14455.5	24566	38586
	Mean, Gy/10^7 primaries	0.02	0.01	0.02	0.02	0.03
	Mean± Standard deviation, %	1.17±0.17	1.08±0.07	1.14±0.12	1.25±0.29	1.83±1.31
Voxels presented in the prediction and not presented in GT	count	206	1	71	690	6973
	(in %)	1.15	0.00	0.37	4.23	21.09
	Total voxels in prediction	18032.5	10669	13577	22817	36523
	Mean, Gy/10^7 primaries	0.02	0.01	0.02	0.03	0.05
	Mean± Standard deviation, %	1.33±0.40	0.97±0.00	1.15±0.18	1.61±0.86	2.46±2.57

4 Discussion

This work presents the first AI-based dose engine for carbon ion therapy capable of predicting both absorbed dose and the α and β parameters required for RBE-weighted dose calculation and optimisation. By extending the transformer-based DoTA architecture and introducing a cross-attention mechanism, we achieved MC-comparable accuracy across all outputs while reducing computation time by over 400-fold compared with GPU MC simulations. This performance addresses a critical bottleneck in carbon ion treatment planning, where MC accuracy has traditionally been compromised for the sake of speed in analytical algorithms.

The high gamma pass rates (>98% median) and strong spatial agreement (Dice coefficient 0.95 at 1% isodose) demonstrate that the proposed framework delivers robust predictions across a wide range of anatomical configurations. Accuracy loss in highly heterogeneous regions—also a challenge for pencil beam algorithms—was modest and primarily affected high-gradient dose falloff. These cases highlight the importance of targeted data augmentation or hybrid MC–AI workflows to further improve performance in extreme anatomies, such as post-surgical cavities or highly bony structures.

Clinically, the model's independence from beam energy and anatomy for inference time makes it well-suited for online adaptive planning and 4D robust optimisation. Parallelisation across GPUs could potentially reduce runtime to milliseconds for full plans, enabling real-time replanning during treatment sessions.

However, several limitations must be acknowledged. The training data were restricted to head-and-neck patients, fixed gantry angles, and a therapeutic energy range of 115–260 MeV/u. A broader generalisation will require expanding the dataset to include more anatomical sites, beam geometries, and beamlines. Additionally, while α and β predictions matched MC values, their biological accuracy depends on the local effect model and requires further validation against experimental radiobiology data. Future research should address the following issues: the effect of the absence of the low-dose region in the C-DoTA prediction (8% of voxels on average) and the error introduced by alpha and beta predictions, both of which need to be investigated in the context of a full plan. The produced output can be directly used for optimisation in comparison to the AI-based prediction of the full plan. Moreover, the consistency between the base data for optimisation and dose calculation reduces the error of the final plans. Finally, to generalise the model for various treatment sites, the inclusion of tissue-specific alpha/beta ratios needs to be studied.

Overall, this proof-of-concept demonstrates that deep learning can deliver MC-quality, biologically-weighted dose predictions for carbon ion therapy with clinical runtimes, paving the way for integration into routine workflows and wider adoption of this modality.

5 Conclusions

We developed an AI model able to calculate RBE-weighted dose in carbon-ion therapy with MC accuracy but at 400-fold faster computation time. The model's speed enables rapid online re-planning, which is particularly valuable for addressing inter-fractional anatomical changes. By removing the need for MC, C-Dota will improve carbon ion therapy accuracy, enabling safer and more effective treatments. Future directions include lowering the prediction threshold to 0.5% (clinically required); collaboration with C-ion clinical centres to test robustness across beamlines and patient cohorts; and embedding the model into clinical workflows (e.g., TRiP98) to evaluate end-to-end plan quality.

Conflict of interest

Zoltan Perko is an associate professor at Delft University of Technology and works as Senior Applied Scientist at Radformation Inc. His industry employment has no connection to the presented work. The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Durante M, Orecchia R, Loeffler JS. Charged-particle therapy in cancer: clinical uses and future perspectives. Nat Rev Clin Oncol 2017;14:483–95. https://doi.org/10.1038/nrclinonc.2017.30.

[2] Durante M, Paganetti H. Nuclear physics in particle therapy : a review. Reports Prog Phys 2016;79:096702. https://doi.org/10.1088/0034-4885/79/9/096702.

[3] Wolf M, Anderle K, Durante M, Graeff C. Robust treatment planning with 4D intensity modulated carbon ion therapy for multiple targets in stage IV non-small cell lung cancer. Phys Med Biol 2020;65. https://doi.org/10.1088/1361-6560/aba1a3.

[4] Paganetti H, Botas P, Sharp GC, Winey B. Adaptive proton therapy. Phys Med Biol 2021;66:22TR01. https://doi.org/10.1088/1361-6560/ac344f.

[5] Graeff C, Volz L, Durante M. Emerging technologies for cancer therapy using accelerated particles. Prog Part Nucl Phys 2023:104046. https://doi.org/10.1016/j.ppnp.2023.104046.

[6] Krämer M, Scholz M. Rapid calculation of biological effects in ion radiotherapy. Phys Med Biol 2006;51:1959–70. https://doi.org/10.1088/0031-9155/51/8/001.

[7] Tan J, Parisi A, Furutani KM, Yagi M, Hartzell S. Pioneering MCF MKM RBE-weighted dose calculation for carbon ion radiotherapy : development of a pencil beam algorithm and validation against Monte Carlo simulation on clinical CT data Physics in Medicine & Biology Pioneering MCF MKM RBE-weighted dose calc n.d.

[8] Ruangchan S, Palmans H, Knäusl B, Georg D, Clausen M. Dose calculation accuracy in particle therapy: Comparing carbon ions with protons. Med Phys 2021;48:7333–45. https://doi.org/10.1002/mp.15209.

[9] Deng Y, Chen Z, Wang Q, Yepes P, Shen Z, Chen H, et al. Comparison of pencil beam and Monte Carlo calculations with ion chamber array measurements for patient-specific quality assurance. Radiat Med Prot 2022;3:115–22. https://doi.org/10.1016/j.radmp.2022.07.002.

[10] Deng Y, Chen Z, Wang Q, Yepes P, Shen Z, Chen H, et al. Comparison of pencil beam and Monte Carlo calculations with ion chamber array measurements for patient-speci fi c quality assurance. Radiat Med Prot 2022;3:115–22. https://doi.org/10.1016/j.radmp.2022.07.002.

[11] Zhang H, Li Q, Liu X, Ma Y, He P. Physica Medica Validation and testing of a novel pencil-beam model derived from Monte Carlo simulations in carbon-ion treatment planning for different scenarios. Phys Medica 2022;99:1–9. https://doi.org/10.1016/j.ejmp.2022.04.018.

[12] Pastor-Serrano O, Perkó Z. Millisecond speed deep learning based proton dose calculation with Monte Carlo accuracy. Phys Med Biol 2022;67. https://doi.org/10.1088/1361-6560/ac692e.

[13] Radonic D, Xiao F, Wahl N, Voss L, Neishabouri A, Delopoulos N, et al. Proton dose calculation with LSTM networks in presence of a magnetic field. Phys Med Biol 2024;69.

https://doi.org/10.1088/1361-6560/ad7f1e.

[14] Neishabouri A, Wahl N, Mairani A, Köthe U, Bangert M. Long short-term memory networks for proton dose calculation in highly heterogeneous tissues. Med Phys 2021;48. https://doi.org/10.1002/mp.14658.

[15] Zhang X, Hu Z, Zhang G, Zhuang Y, Wang Y, Peng H. Dose calculation in proton therapy using a discovery cross-domain generative adversarial network (DiscoGAN). Med Phys 2021;48. https://doi.org/10.1002/mp.14781.

[16] Zhang L, Holmes JM, Liu Z, Vora SA, Sio TT, Vargas CE, et al. Beam mask and sliding window-facilitated deep learning-based accurate and efficient dose prediction for pencil beam scanning proton therapy. Med Phys 2024;51:1484–98. https://doi.org/10.1002/mp.16758.

[17] Wu C, Nguyen D, Xing Y, Montero AB, Schuemann J, Shang H, et al. Improving proton dose calculation accuracy by using deep learning. Mach Learn Sci Technol 2021;2:15017. https://doi.org/10.1088/2632-2153/abb6d5.

[18] Xing Y, Zhang Y, Nguyen D, Lin MH, Lu W, Jiang S. Boosting radiotherapy dose calculation accuracy with deep learning. J Appl Clin Med Phys 2020;21:149–59. https://doi.org/10.1002/acm2.12937.

[19] He R, Zhang H, Wang J, Shen G, Luo Y, Zhang X, et al. Deep learning-based prediction of Monte Carlo dose distribution for heavy ion therapy. Phys Imaging Radiat Oncol 2025;34:100735. https://doi.org/10.1016/j.phro.2025.100735.

[20] De Simoni M, Battistoni G, De Gregorio A, De Maria P, Fischetti M, Franciosini G, et al. A Data-Driven Fragmentation Model for Carbon Therapy GPU-Accelerated Monte-Carlo Dose Recalculation. Front Oncol 2022;12:1–16. https://doi.org/10.3389/fonc.2022.780784.

[21] Krämer M, Scholz M. Treatment planning for heavy-ion radiotherapy: calculation and optimization of biologically effective dose. Phys Med Biol 2000;45:3319–30. https://doi.org/10.1088/0031-9155/45/11/314.

[22] Schiavi A, Senzacqua M, Pioli S, Mairani A, Magro G, Molinelli S, et al. Fred: A GPU-accelerated fast-Monte Carlo code for rapid treatment plan recalculation in ion beam therapy. Phys Med Biol 2017;62:7482–504. https://doi.org/10.1088/1361-6560/aa8134.

[23] Hild S, Graeff C, Trautmann J, Kraemer M, Zink K, Durante M, et al. Fast optimization and dose calculation in scanned ion beam therapy. Med Phys 2014;41. https://doi.org/10.1118/1.4881522.

[24] Zaider M, Rossi HH. The synergistic effects of different radiations. Radiat Res 1980;83. https://doi.org/10.2307/3575352.

[25] Dice LR. Measures of the Amount of Ecologic Association Between Species. Ecol Soc Am 1945;26:297–302.

[26] Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. 33rd Int Conf Mach Learn ICML 2016 2016;3:1651–60.

Supplementary materials

Table S 1 Weights used in the custom loss function

Parameter	w_{Mask}	w_{Depth}	w_{Region}
C-DoTA-d	0.4	1	0
C-DoTA- α	0	0	1
C-DoTA- β	0	0	0.4

Table S 2 GPR 1%/1 mm results without AI dose mask applied

Parameter	Value range	Gamma pass rate (1%/1 mm) [%]*				
		Median	Min	10%-tile	90%-tile	Max
D	0-2.5 Gy	99.76	92.06	98.70	99.96	100
α	0-1.6 Gy^{-1}	91.17	71.49	87.28	93.77	97.58
β	0-0.03 Gy^{-2}	91.37	71.69	87.66	93.80	97.69

Table S 3 GPR 1%/1 mm results after 56 epochs with MSE. AI dose mask was applied to both GT and AI output.

Parameter	Value range	Gamma pass rate (1%/1 mm) [%]				
		Median	Min	10%-tile	90%-tile	Max
D	0-2.5 Gy	99.61	88.66	98.02	99.94	100
α	0-1.6 Gy^{-1}	95.76	64.53	88.80	98.44	99.91
β	0-0.03 Gy^{-2}	95.39	62.65	87.90	98.20	99.61

Table S 4 GPR 1%/1 mm results after 112 epochs with MSE. AI dose mask was applied to both GT and AI output.

Parameter	Value range	Gamma pass rate (1%/1 mm) [%]				
		Median	Min	10%-tile	90%-tile	Max
D	0-2.5 Gy	99.66	91.54	98.46	99.94	100
α	0-1.6 Gy^{-1}	97.28	69.61	92.25	98.99	99.89
β	0-0.03 Gy^{-2}	96.26	65.93	90.61	98.56	99.73

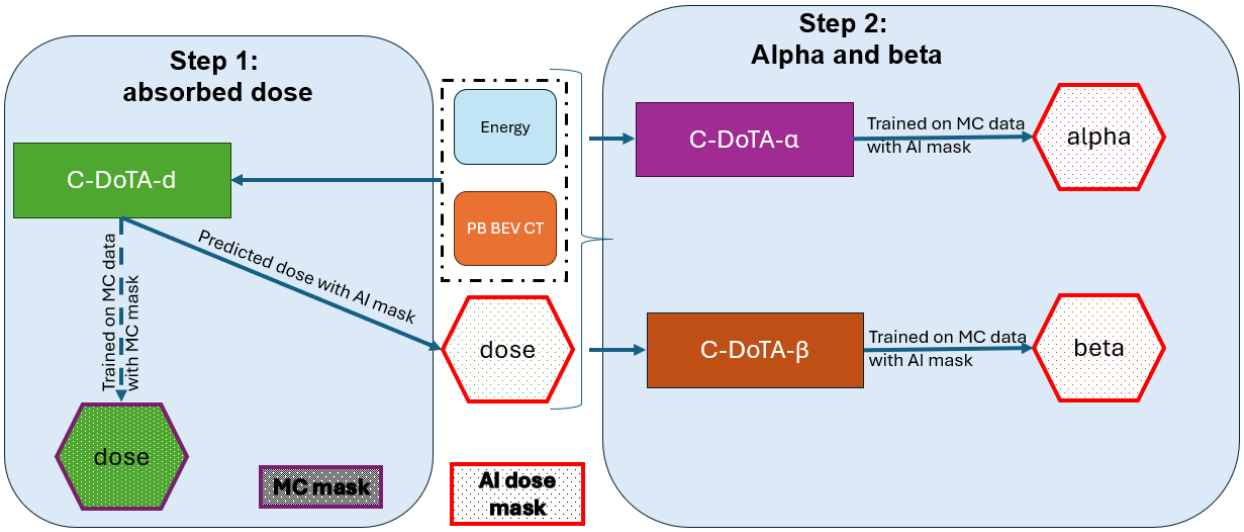

Figure S 1 Data workflow. The output of the C-DoTA-d model is normalised to be used as an input.

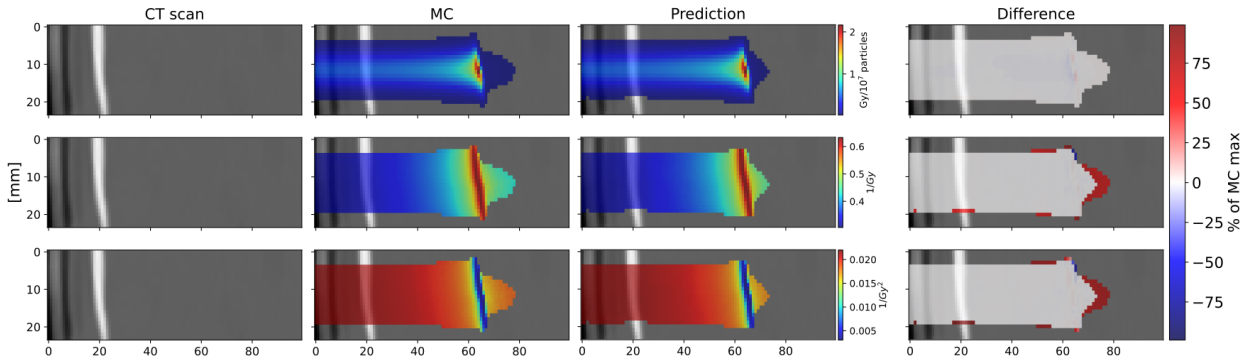

Figure S 2 Best physical dose GPR result. The top figure only AI AI-selected voxels, bottom-original thresholds. GPR 1%/1 mm (top row d:100%, middle row alpha:100%, bottom row beta:99.87%)

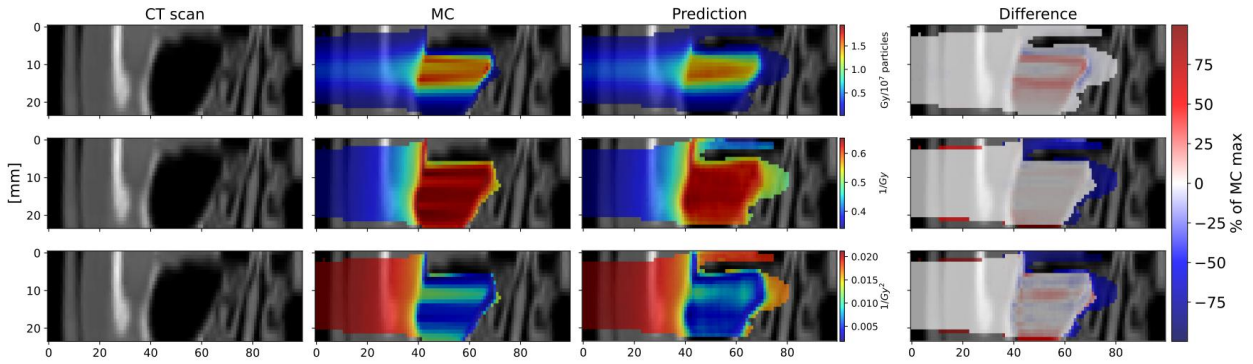

Figure S 3 Worst physical dose GPR result. The top figure only AI AI-selected voxels, bottom-original thresholds. GPR 1%/1 mm (top row d:92.06%, middle row alpha:89.16%, bottom rowbeta:90.62%)

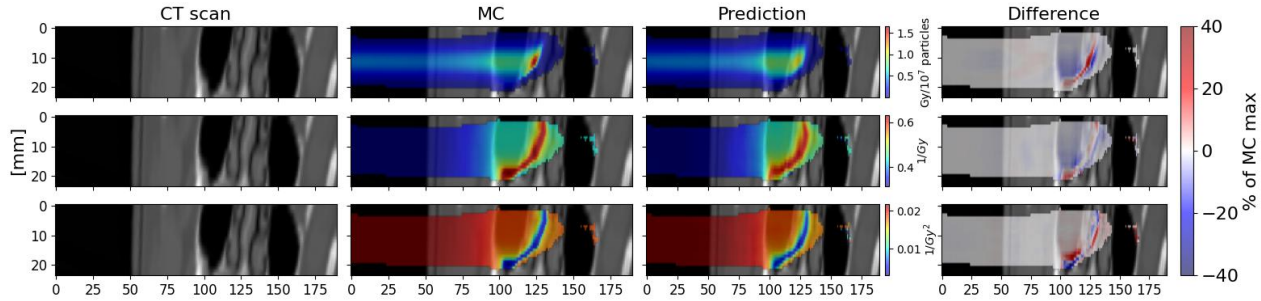

Figure S 4 Worst result GPR for alpha. The GPR with a 1%/1 mm criterion was top row absorbed dose (D): 95.28%, middle row alpha (α): 86.36%, bottom row beta (β): 87.76%. On the left, a slice of the BEV CT is shown. In the centre, the MC ground truth is shown on the left and the C-DoTA predictions on the right. The far right row of the plot shows the difference (MC minus C-Dota). For better visibility, the maximum and minimum values of the colour scale of the difference plots were set to the maximum between the absolute values of the 99.9%-tile and 0.1 %-tile.

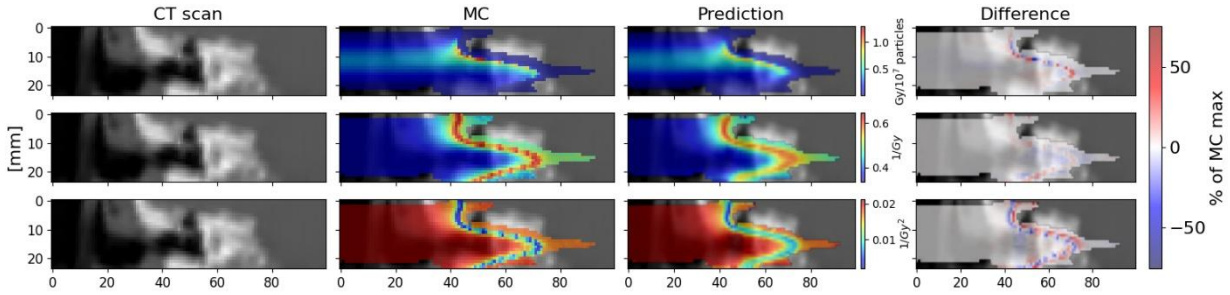

Figure S 5 Worst result GPR for beta. The GPR with a 1%/1 mm criterion was top row absorbed dose (D): 94.53%, middle row alpha (α): 88.73%, bottom row beta (β): 85.26%. On the left, a slice of the BEV CT is shown. In the centre, the MC ground truth is shown on the left and the C-DoTA predictions on the right. The far right row of the plot shows the difference (MC minus C-Dota). For better visibility, the maximum and minimum values of the colour scale of the difference plots were set to the maximum between the absolute values of the 99.9%-tile and 0.1 %-tile.

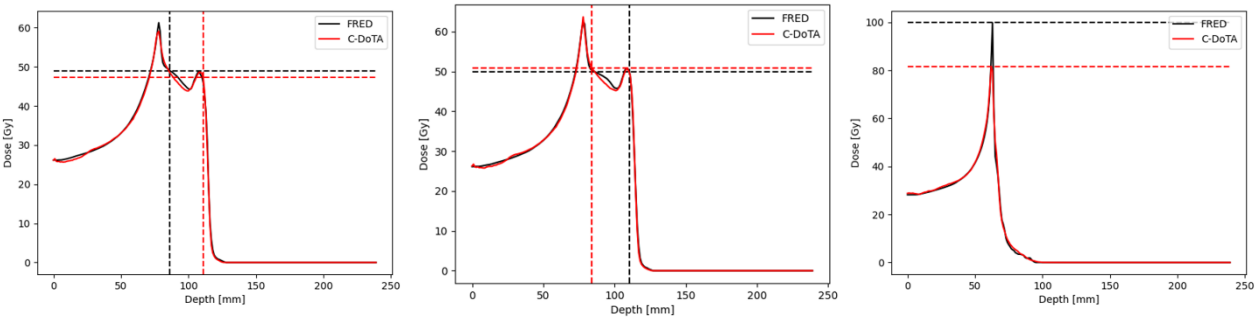

Figure S 6 IDD Curve comparison. The left figure corresponds to ΔR80 min, the middle one ΔR80 max and the right one to the maximum of ΔDmax

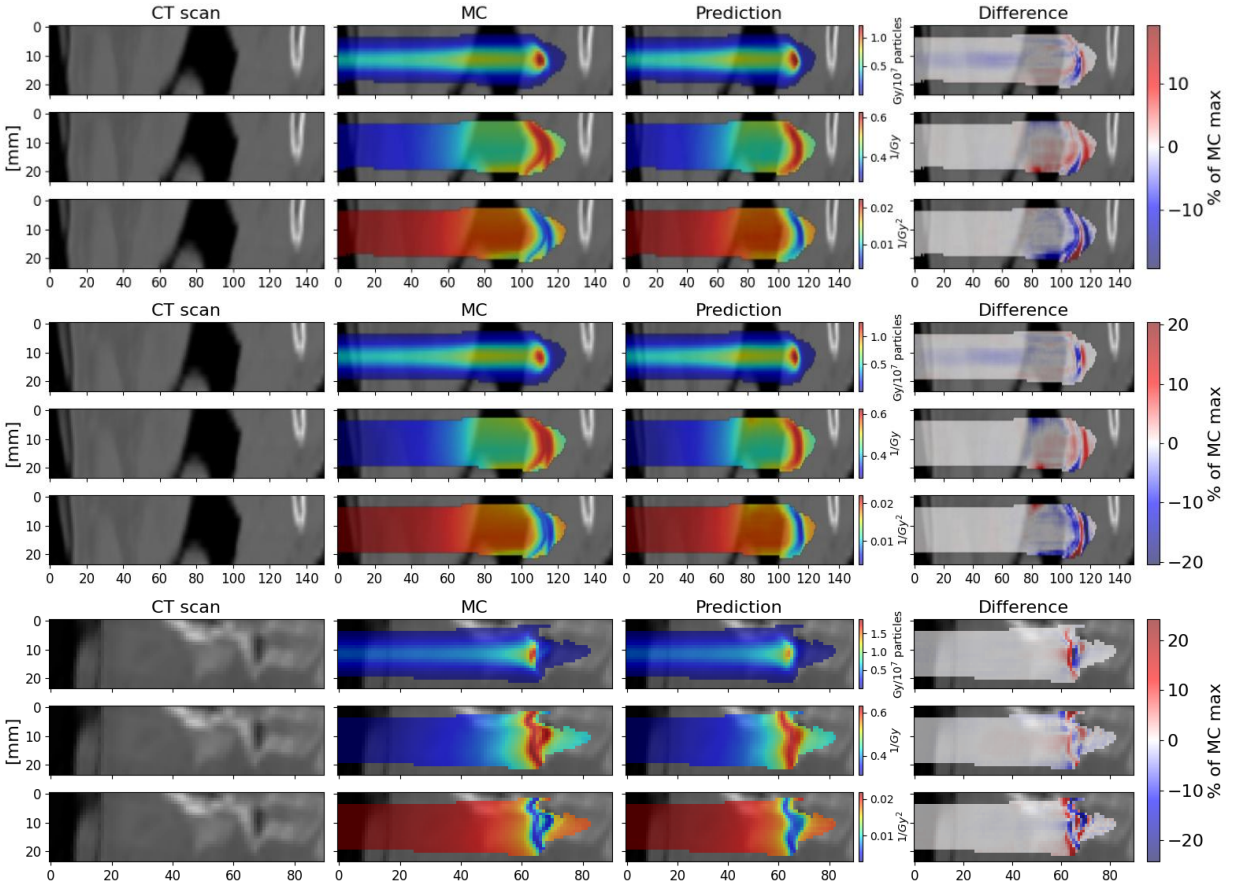

Figure S 7 Top corresponds to Figure S 6 left, middle to the middle and the bottom to the right one

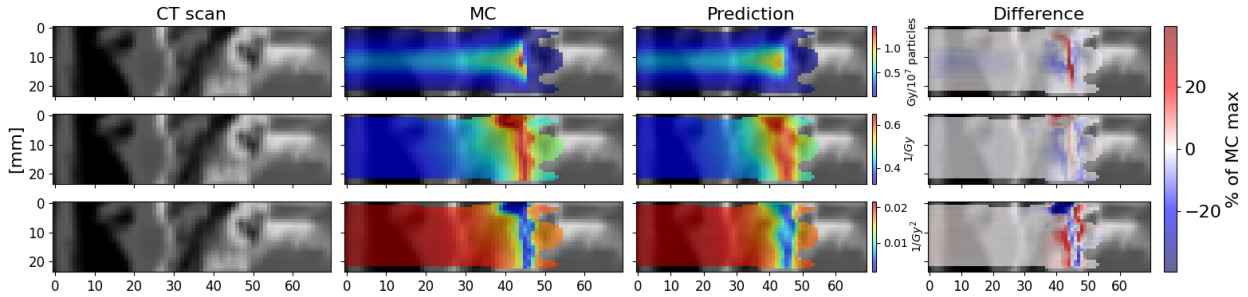

Figure S 8 An additional test of the models on inhomogenious structure. The GPR with a 1%/1 mm was top row absorbed dose (D): 94.60%, middle row alpha (α): 87.42%, bottom row beta (β): 80.26%. The top part: on the left, a slice of the BEV CT is shown; in the centre, the MC ground truth is shown followed by the C-DoTA predictions; the las column repersent the difference (MC minus C-Dota). For better visibility, the maximum and minimum values of the colour scale in the difference plots were set to the maximum between the absolute values of the 99.9%-tile and the 0.1%-tile.