

RESEARCH ARTICLE

MEDICAL PHYSICS

Deep learning-based fast denoising of Monte Carlo dose calculation in carbon ion radiotherapy

Xinyang Zhang^{1,2,3,4} | Hui Zhang^{1,2,3,5} | Jian Wang^{1,2,3,4} | Yuanyuan Ma^{1,2,3,5} |
Xinguo Liu^{1,2,3,5} | Zhongying Dai^{1,2,3,5} | Rui He^{1,6} | Pengbo He^{1,2,3,5} |
Qiang Li^{1,2,3,4,5}

¹Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou, China

²Key Laboratory of Heavy Ion Radiation Biology and Medicine of Chinese Academy of Sciences, Lanzhou, China

³Key Laboratory of Basic Research on Heavy Ion Radiation Application in Medicine, Gansu Province, Lanzhou, China

⁴University of Chinese Academy of Sciences, Beijing, China

⁵Putian Lanhai Nuclear Medicine Research Center, Putian, China

⁶School of Nuclear Science and Technology, Lanzhou University, Lanzhou, China

Correspondence

Qiang Li and Pengbo He, Institute of Modern Physics, Chinese Academy of Sciences, 509 Nanchang Road, Lanzhou 730000, China.
Email: liqiang@impcas.ac.cn and hepengbo@impcas.ac.cn

Funding information

National Key Research and Development Program of China, Grant/Award Number: 2022YFC2401503; National Natural Science Foundation of China, Grant/Award Number: 12005271; Natural Science Foundation of Gansu Province, Grant/Award Numbers: 22JR5RA125, 23JRR577; Special Project of Science and Technology Cooperation between Hubei Province and Chinese Academy of Sciences, Grant/Award Number: 42000021817T300000050

Abstract

Background: Plan verification is one of the important steps of quality assurance (QA) in carbon ion radiotherapy. Conventional methods of plan verification are based on phantom measurement, which is labor-intensive and time-consuming. Although the plan verification method based on Monte Carlo (MC) simulation provides a more accurate modeling of the physics, it is also time-consuming when simulating with a large number of particles. Therefore, how to ensure the accuracy of simulation results while reducing simulation time is the current difficulty and focus.

Purpose: The purpose of this work was to evaluate the feasibility of using deep learning-based MC denoising method to accelerate carbon-ion radiotherapy plan verification.

Methods: Three models, including CycleGAN, 3DUNet and GhostUNet with Ghost module, were used to denoise the 1×10^6 carbon ions-based MC dose distribution to the accuracy of 1×10^8 carbon ions-based dose distribution. The CycleGAN's generator, 3DUNet and GhostUNet were all derived from the 3DUNet network. A total of 59 cases including 29 patients with head-and-neck cancers and 30 patients with lung cancers were collected, and 48 cases were randomly selected as the training set of the CycleGAN network and six cases as the test set. For the 3DUNet and GhostUNet models, the numbers of training set, validation set, and test set were 47, 6, and 6, respectively. Finally, the three models were evaluated qualitatively and quantitatively using RMSE and three-dimensional gamma analysis (3 mm, 3%).

Results: The three end-to-end trained models could be used for denoising the 1×10^6 carbon ions-based dose distribution, and their generalization was proved. The GhostUNet obtained the lowest RMSE value of 0.075, indicating the smallest difference between its denoised and 1×10^8 carbon ions-based dose distributions. The average gamma passing rate (GPR) between the GhostUNet denoising-based versus 1×10^8 carbon ions-based dose distributions was 99.1%, higher than that of the CycleGAN at 94.3% and the 3DUNet at 96.2%. Among the three models, the GhostUNet model had the fewest parameters (4.27 million) and the shortest training time (99 s per epoch) but achieved the best denoising results.

Conclusion: The end-to-end deep network GhostUNet outperforms the CycleGAN, 3DUNet models in denoising MC dose distributions for carbon ion radiotherapy. The network requires less than 5 s to denoise a sample of MC

simulation with few particles to obtain a qualitative and quantitative result comparable to the dose distribution simulated by MC with relatively large number particles, offering a significant reduction in computation time.

KEYWORDS

carbon ion radiotherapy, deep learning, gamma passing rate, Monte Carlo simulation, plan verification

1 | INTRODUCTION

Patient's anatomy often changes during the different fractions of carbon-ion radiotherapy and thus this geometric variation has a dosimetric impact on the organ at risks (OARs) as well as the target volume. The goal of adaptive carbon ion therapy is to measure and calculate the dynamic changes in anatomy of the patient during the treatment course, so that the final doses delivered are the same as the optimal planned ones.^{1–3}

The main technical components of the adaptation workflow include image guidance, replanning and plan verification.² Plan verification is the last and most important step before the implementation of adaptive carbon ion therapy, as its results determine whether the adaptive treatment plan can be executed.⁴ However, the conventional plan verification with phantom-based measurement is not suitable for online adaptive carbon-ion radiotherapy due to the consumption of time and human resource.

In recent years, some studies have proposed the use of independent Monte Carlo (MC) simulation based on anatomy of the patient for online adaptive proton therapy (APT) to solve the problem of online plan verification.^{5–9} The MC approach allows end-to-end calculations based on the patient's anatomy without transplanting the treatment plan to a phantom for measurement. Moreover, MC-based radiation transport simulations are acknowledged as the gold standard for accurately calculating the absorbed dose distribution.^{10–12} Compared with analytical algorithms, MC radiation transport techniques can accurately simulate nuclear interactions and Coulomb scatterings, providing the more accurate modeling of the physics.¹³ Due to these advantages, the MC method can serve as a benchmark reference in the verification of carbon-ion treatment plans, replacing conventional measurement methods as the plan verification scheme in online adaptive carbon-ion radiotherapy.

However, the simulated dose distributions can be strongly influenced by statistical uncertainty due to the randomness of the MC algorithms.^{14,15} Some residual statistical noises affect pixel intensities because the number of simulated particles is limited and the simulation process is completely random indeed. In MC simulation, the number of particles required for simulation increases with the improvement of the desired

level of precision. MC simulations of millions of particles is usually time-consuming to achieve a certain level of acceptable statistical uncertainty. Therefore, shortening the Monte Carlo simulation time and reducing statistical uncertainty while not sacrificing the accuracy of dose distribution are currently a research focus.

Image filtering methods have been proposed in some studies to smoothen the dose distribution.¹⁶ However, it is difficult to find a balance between the remaining noise and the bias introduced by smoothening. Some studies have shown that parallel computation based on graphics processing units (GPUs) can accelerate MC simulation from hours to minutes.^{17,18} This approach can greatly improve simulation speed but is unfriendly to researchers and organizations with limited hardware resources. In recent years, deep learning (DL) has achieved state-of-the-art performance in different research fields such as classification, detection and segmentation,^{19–22} and it has also been used to reduce the noise in low-dose computed tomography (CT) and MC-based dose distributions with promising results.^{23–25} Bai et al. designed a lightweight DL-based convolutional neural network (CNN) for fast MC dose denoising in photon radiotherapy.²⁶ Javaid et al. and Zhang et al. have denoised the MC simulated dose distribution for proton radiotherapy using dilated U-net and ResNet networks to reduce the computation time.^{4,24} Most of the MC simulation platforms used in the above study are MCsquare,²⁷ which was specifically designed for proton radiotherapy. It simplifies some physical processes to accelerate the simulation of proton transport and achieves results close to those of GATE,²⁸ which is an advanced open-source software for Monte Carlo simulations and mainly used for medical imaging and radiotherapy.^{29–33} To our knowledge, there is currently no research based on DL for denoising the dose distribution of carbon ions.

In this study, we proposed GhostUNet, a novel DL model to reduce statistical uncertainty of MC dose distributions in carbon ion radiotherapy and quantitatively compared it with 3D UNet and CycleGAN.³⁴ GhostUNet is a lightweight and efficient model that integrates the Ghost module based on the 3D UNet structure, which can improve the denoising ability of the network while reducing the model parameters and computation. Unlike the MCsquare-based previous studies where

the physical process was simplified, we developed a denoising framework based on the precise carbon-ion transport simulation of the GATE simulation platform, allowing for more accurate results. Furthermore, the proposed model combined with the MC method is expected to replace the conventional measurement-based plan verification to achieve online adaptive carbon-ion therapy.

2 | MATERIALS AND METHODS

2.1 | Image data and preprocessing

A database of 59 patients (lung cancer: 30, and head and neck cancer: 29) was used to conduct this study. All the patient CT images were obtained from Gansu Provincial Cancer Hospital and Gansu Provincial Hospital with the approval by the Institutional Review Board (IRB) of the hospital. This study was carried out under the approval by the Academic Committee of the Institute of Modern Physics, Chinese Academy of Sciences. The in-plane spatial dimension of all CT images was 512×512 , with the number of slices varying from 56 to 289 for different patients. To obtain MC dose distributions for training the network models, an open-source research treatment planning system (TPS) matRad³⁵ was used to devise the radiation therapy schemes, including intensity modulated particle therapy (IMPT) and single field uniform dose (SFUD) plans, based on the databases of the spot-scanning beam line for the Heavy Ion Medical Machine (HIMM) in Lanzhou, China.³⁶ All treatment plans were optimized on the basis of the physical dose, and comprised the total prescribed dose of 30 Gy regimen delivered in 30 fractions with one or two carbon-ion fields. Then the plans were recalculated in the GATEv8.2/Geant4v10.05³⁷ Monte Carlo simulation platform, adopting the same spot positions, energies and weights as calculated in matRad. The same conversion factors between Hounsfield value and stopping power were used in GATE. The particle (including gamma, electron and positron) production thresholds were set to 1 m, 1 mm and 0.1 mm in the world, beam line components and target, respectively, from the point of view of simulation time and calculation accuracy. The QGSP_BERT_HP_EMY physics list was used in the simulation, which contained the command of hadronic physics processes and EM processes for carbon ion therapy. The order of 1×10^8 particles were tracked under 100 parallel calculations on the Intel® Xeon(R) Platinum 8173 M CPU@2.00 Ghz. One of the results (1×10^6 particles) was used as the input to the network, while the ground truth (GT) was the total result (1×10^8 particles). The MC simulation times for the 59 cases with 1×10^6 and 1×10^8 carbon ions ranged from 16 to 72 min and from 248 to 6616 min, respectively.

Data were pre-processed before being fed into the network. First, all dose distribution samples were resized to $512 \times 512 \times 96$ due to their different depths in the z axis. Additionally, following the standard protocol for data preprocessing, we scaled and normalized the 3D dose distributions, where the pixel intensities were adjusted to different intervals for model training: [0, 255] for CycleGAN, and [0, 1] for 3D UNet and GhostUNet.

2.2 | Network architecture

In this study, three network models were used to denoise the dose distributions: the classical 3D UNet as a benchmark, CycleGAN and GhostUNet with the Ghost model.

2.2.1 | 3D UNet

Ronneberger et al. designed and proposed UNet³⁸ using encoder-decoder architecture with long skip connections at the two-dimensional (2D) level, which has been widely used for medical image segmentation. 3D medical image data are input into UNet as 2D slices, which cannot utilize their spatial structure information. To analyze the volumetric data of medical images, 3D UNet was raised, where 2D operations were replaced by corresponding 3D operations. We first denoised the MC dose distributions using the 3D UNet as the baseline model. The encoder network consisted of 5 levels for each with two convolutional blocks, which included a 3×3 convolutional layer with a stride of 1 followed by a batch normalization layer and leaky rectified linear unit (ReLU). Downsampling was implemented by a $2 \times 2 \times 2$ max pooling layer with a stride of 2 after each level. The decoding stage followed the similar design to the encoding path, but with a transposed convolutional layer between each level for upsampling operations.

2.2.2 | CycleGAN

As another commonly used model for biomedical image research, CycleGAN was also adopted for comparison. The structure of the CycleGAN model is shown in Figure 1 and included double generators and discriminators compared to the conventional GAN³⁹ model, which worked in opposition until they converge to an equilibrium point. We employed the CycleGAN to synthesize the precise dose distributions from the rough ones so as to achieve the goal of denoising. Unet-custom was chosen as the generator for the CycleGAN, which had a similar structure to the 3D UNet above but only one convolutional layer was reserved for each level due to memory limitations. Meanwhile, the discriminator

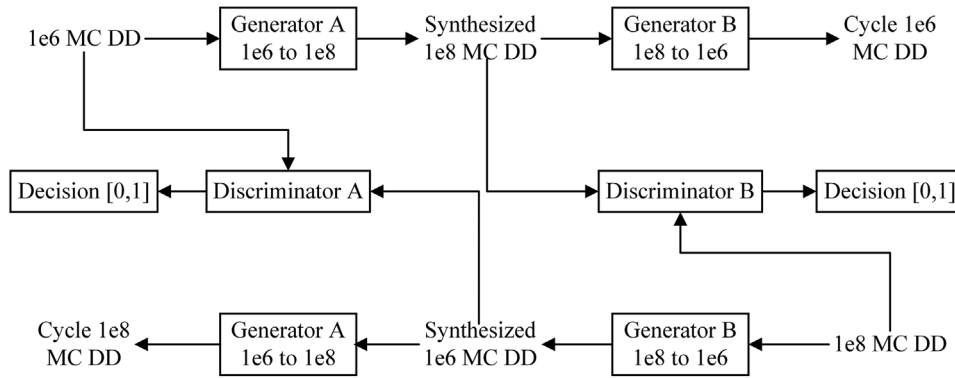


FIGURE 1 Structure of the CycleGAN network (DD means dose distribution).

was a Patch GAN, thus it penalized volume only at the scale of the dose distribution patches. The role of this discriminator was to determine whether each patch in the dose distribution was real or fake. Eventually, the convolution responses of all the patches across the dose distribution were averaged to offer the output of the discriminator. The loss function is illustrated below. The generator G_A converted the 1×106 MC dose distribution to generate a synthetic 1×108 MC dose distribution as close as possible to the real 1×108 MC dose distribution, and the discriminator D_B distinguished the synthetic 1×108 MC dose distribution from the real 1×108 MC dose distribution, which constituted an adversarial loss:

$$\mathcal{L}_{GAN}(G, D_B, A, B) = \mathbb{E}_B[\log D_B(X_B)] + \mathbb{E}_A[\log(1 - D_B(G_A(X_A)))] \quad (1)$$

Then, G_B was employed to generate the cycle 1×106 MC dose distribution from the synthetic 1×108 MC dose distribution, where the structure of 1×106 MC dose distribution was maintained by the cycle-consistent loss. For the two cycles:

$$\mathcal{L}_{cyc}(G, A, B) = \mathbb{E}_B|G_B(G_A(X_A)) - X_A|_{l1} + \mathbb{E}_A|G_A(G_B(X_B)) - X_B|_{l1} \quad (2)$$

To ensure that the synthetic 1×108 MC dose distribution is close to the real 1×108 MC dose distribution, identity loss was also introduced.

$$\mathcal{L}_{identity}(G, A, B) = \mathbb{E}_B|G_B(X_A) - X_A|_{l1} + \mathbb{E}_A|G_A(X_B) - X_B|_{l1} \quad (3)$$

Finally, the full loss function is expressed as Equation (4):

$$\mathcal{L}_{CycleGAN} = \mathcal{L}_{GAN}(G, D_B, A, B) + \mathcal{L}_{GAN}(G, D_A, A, B) + \lambda_1 \mathcal{L}_{cyc}(G, A, B) + \lambda_2 \mathcal{L}_{identity}(G, A, B) \quad (4)$$

In order to improve the performance of denoising, we proposed a model called GhostUNet in this study. The GhostUNet can be considered as a variant of 3D UNet, since it possesses a architecture similar to the 3D UNet. The Ghost module was introduced to address the problem of improving model performance with limited computational resources.

Conventional convolutional neural networks usually require a large number of parameters and floating point operations to achieve satisfactory accuracy.⁴⁰ However, many improvements to convolutional neural networks insert other modules into the network such as attention mechanisms that increase the parameters and computational effort of the model, which is unfriendly to many institutions and researchers with insufficient hardware. Therefore, how to design simple and efficient network architectures with acceptable or better performance on mobile devices is the recent trend. The redundancy in feature maps is an important feature of those excellent convolutional neural networks, but how to balance the redundancy of feature maps with the reduction of model computation has rarely investigated in network architecture design. In recent years, many methods have been proposed to study the compression of deep neural networks, such as network pruning,⁴¹ low-bit quantization⁴² and knowledge distillation.⁴³ However, pre-trained deep neural networks usually limit the performance of these methods.

Han et al. proposed that abundant or even redundant information in the feature maps of well-trained deep neural networks usually ensures a comprehensive understanding of the input data, and redundancy in the feature maps could be an important characteristic of excellent deep neural networks.⁴⁰ They built the Ghost module,⁴⁰ which can generate more features through cheap operations. Specifically, half the number of convolution kernels were first used to compress the input feature dimensions by ordinary convolution to obtain the intrinsic feature maps. Then, the depthwise separable convolution that can increase the receptive field was employed to enlarge the compressed feature maps to

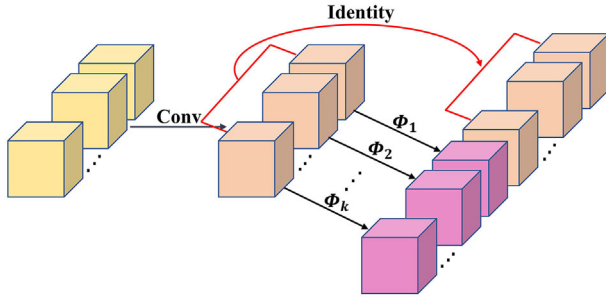


FIGURE 2 The 3D Ghost module. Φ and Identity represent the cheap operation and identity mapping, respectively.

obtain ghost features. Finally, the intrinsic feature maps obtained by ordinary convolutions were concatenated with the ghost feature maps as the output.

Inspired by the Ghost model in two-dimensional (2D) convolutional neural networks, as shown in Figure 2, we extended the Ghost module to 3D in this study. Conventionally, given the 3D input $X \in \mathbb{R}^{c \times d \times h \times w}$, where c is the number of input channels and $d \times h \times w$ is the depth, height and width of the 3D input, respectively, the output $Y \in \mathbb{R}^{c' \times d' \times h' \times w'}$ can be obtained by the operation of an arbitrary convolutional layer for generating w' feature maps expressed as:

$$Y = X * f + b, \quad (5)$$

where b is the bias term and $f \in \mathbb{R}^{c \times k \times k \times k \times n}$ is the convolution filter in this layer.

The normal convolution process described above generates a large number of similar feature map pairs, and the extended 3D Ghost module can generate more feature maps from intrinsic feature maps by cheap linear operations. In detail, the intrinsic feature maps $Y' \in \mathbb{R}^{m \times d' \times h' \times w'}$ are generated from Equation (6) by a primary convolution:

$$Y' = X * f' + b, \quad (6)$$

where $f' \in \mathbb{R}^{c \times k \times k \times k \times m}$ is the filters used and $m \leq n$. To maintain the same spatial size of the output feature maps, the filter size, stride and padding are set to be the same as the convolution above (Equation 5). $s - 1$ redundant feature maps Y'_i are generated from m intrinsic feature map $Y'_j \in \mathbb{R}^{d' \times h' \times w'}$ by:

$$Y'_{ij} = \Phi_{ij}(Y'_j), \quad \forall i = 1, \dots, m, j = 1, \dots, s - 1, \quad (7)$$

By using Equation (7), we can generate $n = m \times s$ feature maps $Y = [Y_{11}, Y_{12}, \dots, Y_{ms}]$ as shown by the 3D Ghost module. The cheap linear operation Φ uses 3D convolution with kernel of $3 \times 3 \times 3$ and the number of operations can be compressed by factor s .

The architecture of GhostUNet network in detail is described in Figure 3. We constructed the GhostUNet by replacing the second 3D convolutional layer in each of the five levels of the downsampling process using the Ghost model on top of the 3D UNet structure.

2.3 | Implementation details

The loss function of the CycleGAN has been described in detail above and we defined the loss function of the 3DUNet and GhostUNet as the mean absolute error (MAE) between the denoised and precise dose distributions. For the CycleGAN experiment, the training data were acquired from 48 patients (lung: 24 and head and neck: 24) and the test data were consisted of 6 patients (lung: 3 and head and neck: 3). Moreover, a total of 47 patients (lung: 24 and head and neck: 23) were used to train the 3DUNet and GhostUNet models, whereas 6 patients (lung: 3 and head and neck: 3) constituted the validation set. To facilitate testing the performance of the three models, we selected the remaining 6 patients (lung: 3 and head and neck: 3) for testing, which were the same as the test patients in the CycleGAN. We trained the CycleGAN, 3DUNet and GhostUNet using one NVIDIA A6000 graphics processing unit for 500, 200 and 200 epochs, respectively.

2.4 | Evaluation metrics

There was a perfect spatial alignment between the denoised dose distribution and the precise simulated dose distribution, and the following metrics to assess the similarity between the two dose distributions were adopted.

Root mean squared error (RMSE) was used first to calculate the pixel intensity difference between the denoised and reference dose distributions. Lower values indicate that the differences between the dose distributions under comparison are relatively small. It is compared by Equation (8):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (D_i - \hat{D}_i)^2}, \quad (8)$$

where i stands for pixel i and n refers to the number of pixels, D_i and \hat{D}_i denote the reference and denoised dose distributions, respectively.

We also evaluated the similarity between the denoised and reference dose distributions using 3D gamma analysis. Gamma analysis can produce gamma indices in each voxel involved, and higher gamma passing rates indicate better similarity between the two dose distributions under comparison.

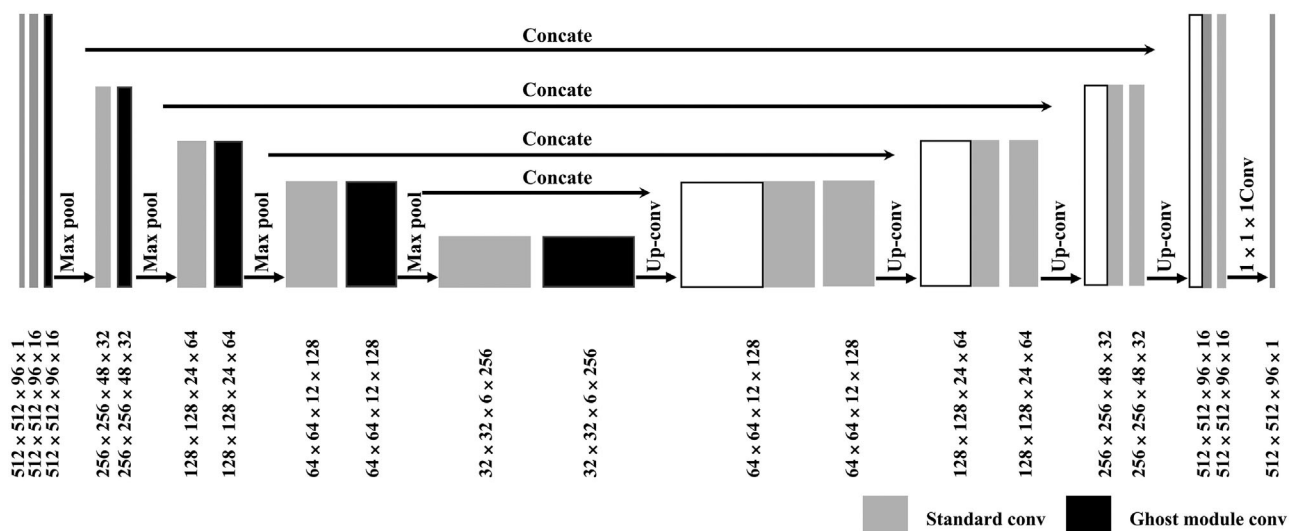


FIGURE 3 The detailed architecture of the GhostUNet network.

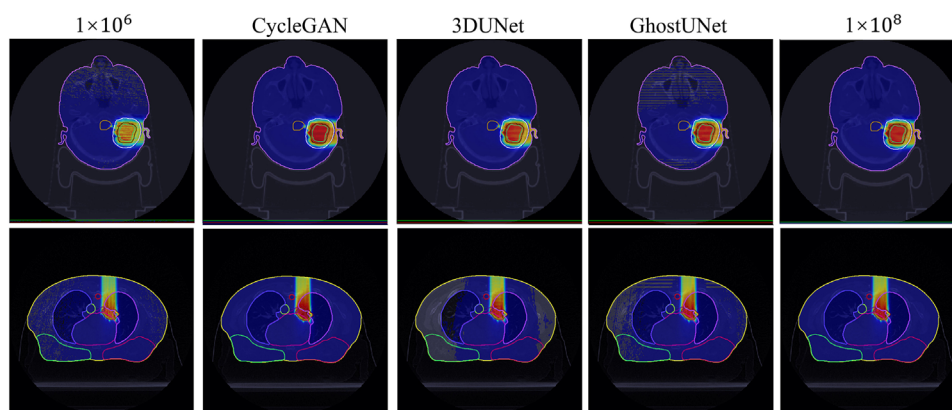


FIGURE 4 Dose distributions of the two cases in the corresponding slice of 106 carbon ions-based, CycleGAN, 3D UNet, GhostUNet and 108 carbon ions-based results. (a) a head-and-neck cancer case and (b) a lung cancer case.

3 | RESULTS

In this section, the denoising results of the three models were first confirmed on the dose images and dose volume histograms (DVHs), respectively, through comparing the denoising results of the three models qualitatively with the 1 × 10⁶ and 1 × 10⁸ carbon ions-based MC dose distributions. The differences in performance of the three models were then compared quantitatively using RMSE and 3D GPR. Finally, the complexity of the models such as the number of parameters, the computational effort and the runtime was evaluated.

Figure 4 shows the dose distribution maps of a layer in the axial view of a head-and-neck patient and a lung cancer patient in the test set after denoising using the CycleGAN, 3DUNet and GhostUNet. The leftmost, rightmost and middle three columns display the results for the 1 × 10⁶ carbon ions-based, 1 × 10⁸

carbon ions-based and denoised dose distributions of the three models, respectively. Clearly, the 1 × 10⁶ carbon ions-based dose distribution was underdosed and accompanied by a large amount of noise. Meanwhile, the three models showed a significant improvement in denoising the 1 × 10⁶ carbon ions-based dose distribution.

The DVHs corresponding to the two cases in Figure 4 is shown in Figure 5. For comparison, we plotted the results of the CycleGAN (dot-dash line), 3DUNet (dashed line), GhostUNet (dotted line) and 1 × 10⁸ particles (solid line) in DVH. For the head-and-neck patient case, it was evident that the GhostUNet-based denoising result was comparable to the 1 × 10⁸ carbon ions-based dose distribution. However, the results of the CycleGAN and 3DUNet were low in most areas of the target volume compared to the 1 × 10⁸ carbon ions-based dose distribution. The GhostUNet denoising results for the lung patients were similar to those for the

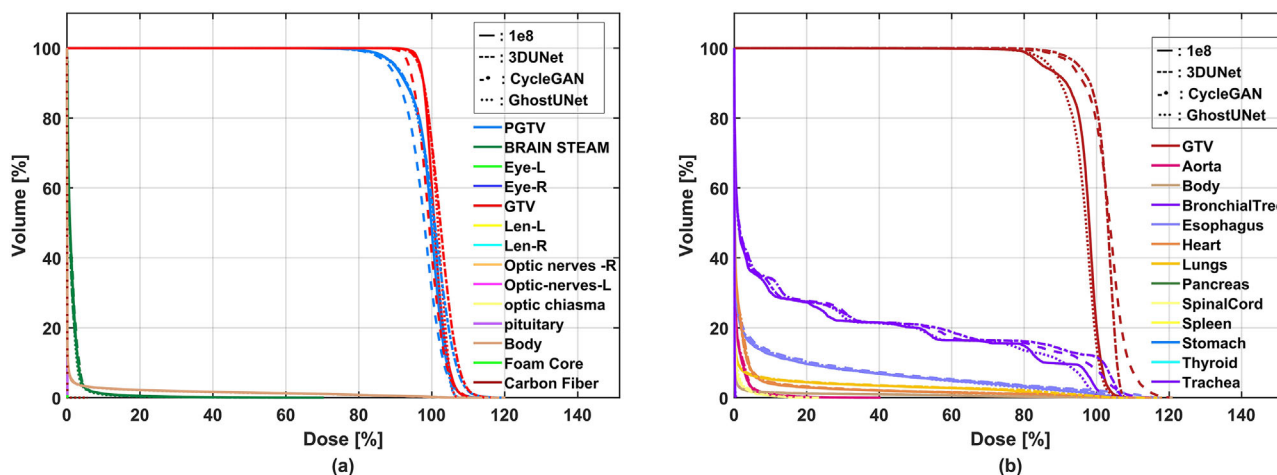


FIGURE 5 DVHs of the three model denoising results in two patient cases. (a) a head-and-neck cancer case and (b) a lung cancer case.

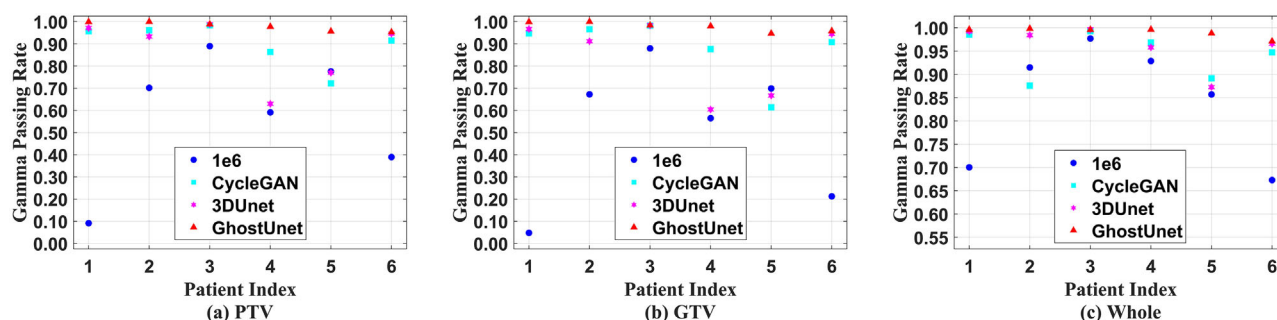


FIGURE 6 The scatter plots of gamma passing rate between the dose distributions based on 1×10^6 carbon ions, the three models after denoising and 1×10^8 carbon ions. (a) PTV gamma passing rate, (b) GTv gamma passing rate and (c) Whole gamma passing rate.

head and neck patients, but the CycleGAN and 3DUNet denoising results had much larger doses over the target volume than the 1×10^8 carbon ions-based dose distribution.

Figure 6 shows scatter plots for visual comparison of the three models for PTV, GTv, and overall dosimetric differences across all cases tested. Since the doses of 1×10^6 and 1×10^8 carbon ions-based MC simulations differed by two orders of magnitude, we normalized them before calculating the gamma passing rates for both. We could see that the gamma passing rates were poor between the 1×10^6 and 1×10^8 carbon ions-based MC dose distributions in different patients relative to the three models after denoising, especially in the PTV and GTv regions. Meanwhile, the GhostUNet demonstrated excellent denoising performance, with GPRs on overall, PTV and GTv significantly exceeding those of the other two models.

Table 1 presents the results of the quantitative evaluation of the three models, which are expressed as the mean \pm standard deviation (SD). On the test set, all three models improved the quality of the 1×10^6 carbon ions-based MC dose distribution to some extent. The denoising ability of the CycleGAN was better than

the 3D UNet on PTV and GTv, and slightly lower on the whole. It was worth noting that the proposed GhostUNet performed optimally in all regions, with an overall gamma passing rate of 99.1% and over 97.5% on GTv and PTV. In addition, the GhostUNet had the smallest RMSE value, indicating that the GhostUNet denoising-based dose distribution was least different from 1×10^8 carbon ions-based one. We also noted that the GhostUNet's denoising ability was relatively stable across all test cases, with minimal fluctuations relative to the CycleGAN and 3D UNet.

Although the network architectures of the CycleGAN generator, 3DUNet and GhostUNet are similar, the parameters and computational effort of the three models are different. The CycleGAN and 3D UNet, as our benchmark models, both had higher complexity than the GhostUNet proposed in this paper. Table 2 shows the number of parameters and the corresponding computation of the three models. Obviously, the GhostUNet network had the least number of parameters and computation, but brought the best denoising results. Most importantly, the introduction of the Ghost module not only increased the denoising performance of the model, but also reduced the model runtime.

TABLE 1 The gamma passing rate and RMSE over all test cases between the dose distributions based on 1×10^6 carbon ions, the three models after denoising and 1×10^8 carbon ions (mean \pm standard deviation).

	3D gamma passing rate (3 mm/3%)			RMSE
	Whole CT	GTV	PTV	
1×10^6	0.842 ± 0.126	0.513 ± 0.318	0.573 ± 0.291	0.148 ± 0.039
CycleGAN	0.943 ± 0.049	0.882 ± 0.137	0.900 ± 0.097	0.087 ± 0.022
3DUNet	0.962 ± 0.046	0.846 ± 0.166	0.873 ± 0.143	0.076 ± 0.020
GhostUNet	0.991 ± 0.010	0.977 ± 0.021	0.979 ± 0.021	0.075 ± 0.021

Abbreviations: CT, computed tomography; GTV, gross tumor volume; PTV, planning target volume; RMSE, root mean squared error.

TABLE 2 Details of the CycleGAN, 3D UNet and GhostUNet models.

Model	Parameters ($\times 10^6$)	Flops ($\times 10^{10}$)	Time per epoch(s)
CycleGAN	66.60	646.96	177
3DUNet	6.58	212.53	106
GhostUNet	4.27	180.17	99

4 | DISCUSSION

Plan verification is an important component of QA for carbon-ion radiotherapy, which assesses whether the treatment plan can be implemented subsequently. Conventional measurement-based plan verification methods are susceptible to detector sensitivity and time-consuming. A few studies have developed DL-based MC denoising methods and tried to replace conventional measurement methods.^{4,24,26} However, these studies have some drawbacks. First, the simulation software they used simplifies the physical processes during particle transport, which can affect the accuracy of the simulation results. In addition, these studies did not consider the introduction of lightweight and efficient networks. In this study, we first simulated the 1×10^6 and 1×10^8 carbon ions-based dose distributions for each case separately using the MC simulation software—GATE. Then, we further developed a lightweight and efficient DL network model to denoise the 1×10^6 carbon ions-based dose distribution to achieve the accuracy of 1×10^8 carbon ions-based one.

We explored the denoising ability of our proposed GhostUNet network and two baseline networks, the CycleGAN and 3D UNet networks, on full resolution (512×512) dose distributions. All networks were trained on the same batch of dose distributions and tested on the same test set independent of the training set. This could be considered as a reliable way to verify the performance of the network model because we forced the network to train its capabilities under the same data distribution, making the network evaluation robust.

As shown in Table 1, the three models had smaller RMSE and larger GPR compared with the 1×10^6 carbon ions-based dose distribution. Since there are rare

studies related to carbon-ion radiotherapy in such field, we compared our results with similar studies in proton radiotherapy.^{4,24} So far, when using DL techniques to denoise MC simulated dose distributions with a small number of particles, RMSE and GPR results on different models had achieved varying degrees of improvement. This is consistent with our results that DL techniques could be used to denoise and outperform other existing methods.

The qualitative and quantitative results of our study showed that the CycleGAN had the worst denoising ability among the three models. Taking Figure 5(b) as an example, the DVH curve of the CycleGAN near the target volume had the largest shift compared to the DVH curve of the 1×10^8 carbon ions-based dose distribution. This might be due to the fact that the CycleGAN's generator network was not deep enough. In the previous section, we have mentioned that the generator had only one layer of convolution at each depth due to memory limitations, which had an impact on the denoising performance of the network.

The network depth of the 3D UNet was twice as deep as the CycleGAN, so its denoising ability was better than the CycleGAN. However, the average GPR of the 3D UNet on the tested cases was only 96.2%, and the denoising ability varied widely among cases. We analyzed the following two possible reasons. First, the limited patient dataset did not allow the 3D UNet to be fully trained and perform at its best. Moreover, due to GPU memory limitation, we reduced the number of filters in the network. That is, while the network had enough depth, the number of filters per depth was reduced, which could lead to learning saturation in the network. All these reasons could result in the limited performance of the 3D UNet.

As shown in Tables 1 and 2, the GhostUNet provided the best denoising ability with the least number of parameters and the shortest training time compared to the other two models. In addition, the GhostUNet had stable denoising capability in all regions on different patient data. It has been shown that the introduction of the GhostUNet into the structure of the 3D UNet was successful in improving the performance of the network and made it lightweight. Two key reasons were responsible for achieving this excellent denoising performance.

One was that the group convolution of cheap operations in the Ghost module increased the field of perception compared to conventional convolution, which ensured that more information was retained at different depths. Another point might be the reduced complexity of the model due to the introduction of the Ghost module, which made the model more difficult to be overfitted and thus improved the performance of the model.

There were also some limitations in our study. Firstly, only 59 cases of head-and-neck and lung cancer were included in this study. The lack of diversity in the data requires the addition of more tumor locations such as liver and prostate in future study. As more data are incorporated, the generalization ability and denoising performance of the model will be further improved. Secondly, to facilitate the conduct of the study, we simplified the treatment plan and did not adequately consider OARs prescription doses, but this had little impact on the exploration of noise reduction problem in this study. Treatment plans will be devised in strict accordance with clinical criteria in our further study. Thirdly, although the 1×108 carbon ions-based dose distribution could be quickly denoised from the 1×106 carbon ions-based dose distribution in this study ($< 5s$), it still took at least 16 min to simulate the 1×106 carbon ions-based dose distribution using GATE. However, it should be noted that the performance of the server seriously affected the simulation time. For example, when a case was simulated in parallel, the 22nd core started simulating only after the first core finished simulating, resulting in a single core simulation time of 3 min while the total simulation time took 18 min. Therefore, the minimum time for MC simulation based on 1×106 carbon ions can be reduced to about 5 min in practical applications when using high-performance servers, greatly improving the clinical applicability. Finally, although the proposed GhostUNet network alleviated the model complexity, the model still had a relatively large number of parameters. Thus, building the network model entirely using lightweight and efficient modules while ensuring the denoising performance of the model is the next working priority, which allows the model to be integrated on a clinical server.

5 | CONCLUSION

In this study, three models including CycleGAN, 3D UNet and GhostUNet were used for denoising the dose distributions obtained from 1×106 carbon ions-based MC simulations, and the following conclusions were drawn: (1) The 1×106 carbon ions-based dose distribution could be denoised using the three models; (2) The denoising performance of the GhostUNet network is significantly superior to that of the CycleGAN and 3D UNet, and the denoising results by the GhostUNet are stable in different test cases; (3) The proper introduction

of the Ghost module can reduce the model complexity and improve the model denoising performance. Thus, our study provides evidence for possibly implementing online plan verification in carbon-ion radiotherapy. Nevertheless, the current MC simulation speed of millions of carbon ions needs to be improved, and the parameter complexity of the deep learning model used for denoising remains to be further reduced.

ACKNOWLEDGMENTS

This work was jointly supported by the National Key Research and Development Program of China (grant number 2022YFC2401503), the National Natural Science Foundation of China (grant number 12005271), the Natural Science Foundation of Gansu Province (grant number 22JR5RA125 and 23JRRA577) and the Special Project of Science and Technology Cooperation between Hubei Province and Chinese Academy of Sciences (grant number 42000021817T300000050).

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

1. Yan D, Vicini F, Wong J, Martinez A. Adaptive radiation therapy. *Phys Med Biol.* 1997;42(1):123-132.
2. Wu QJ, Li TR, Wu QW, Yin FF. Adaptive radiation therapy technical components and clinical applications. *Cancer J.* 2011;17(3):182-189.
3. Albertini F, Matter M, Nenoff L, Zhang Y, Lomax A. Online daily adaptive proton therapy. *Br J Radiol.* 2020;93(1107):20190594.
4. Zhang GL, Chen XY, Dai JR, Men K. A plan verification platform for online adaptive proton therapy using deep learning-based Monte-Carlo denoising. *Phys Med.* 2022;103:18-25.
5. Meier G, Besson R, Nanz A, Safai S, Lomax AJ. Independent dose calculations for commissioning, quality assurance and dose reconstruction of PBS proton therapy. *Phys Med Biol.* 2015;60(7):2819-2836.
6. Matter M, Nenoff L, Meier G, Weber DC, Lomax AJ, Albertini F. Alternatives to patient specific verification measurements in proton therapy: a comparative experimental study with intentional errors. *Phys Med Biol.* 2018;63(20):205014.
7. Johnson JE, Beltran C, Tseung HWC, et al. Highly efficient and sensitive patient-specific quality assurance for spot-scanned proton therapy. *PLoS ONE.* 2019;14(2):0212412.
8. Mackin D, Li YP, Taylor MB, et al. Improving spot-scanning proton therapy patient specific quality assurance with HPlusQA, a second-check dose calculation engine. *Med Phys.* 2013;40(12):121708.
9. Zhu XR, Li YP, Mackin D, et al. Towards effective and efficient patient-specific quality assurance for spot scanning proton therapy. *Cancers.* 2015;7(2):631-647.
10. Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Phys Med Biol.* 2012;57(11):R99-R117.
11. Sorriaux J, Testa M, Paganetti H, et al. Experimental assessment of proton dose calculation accuracy in inhomogeneous media. *Phys Med.* 2017;38:10-15.

12. Yang J, Li J, Chen L, et al. Dosimetric verification of IMRT treatment planning using Monte Carlo simulations for prostate cancer. *Phys Med Biol*. 2005;50(5):869-878.
13. Bongrand A, Koumeir C, Vilhoing D, et al. A Monte Carlo determination of dose and range uncertainties for preclinical studies with a proton beam. *Cancers*. 2021;13(8):1889.
14. Chetty IJ, Curran B, Cygler JE, et al. Report of the AAPM Task Group No. 105: issues associated with clinical implementation of Monte Carlo-based photon and electron external beam treatment planning. *Med Phys*. 2007;34(12):4818-4853.
15. Buffa FM, Nahum AE. Monte Carlo dose calculations and radiobiological modelling: analysis of the effect of the statistical noise of the dose distribution on the probability of tumour control. *Phys Med Biol*. 2000;45(10):3009-3023.
16. El Naqa I, Kawrakow I, Fippel M, et al. A comparison of Monte Carlo dose calculation denoising techniques. *Phys Med Biol*. 2005;50(5):909-922.
17. Jia X, Gu XJ, Graves YJ, Folkerts M, Jiang SB. GPU-based fast Monte Carlo simulation for radiotherapy dose calculation. *Phys Med Biol*. 2011;56(22):7017-7031.
18. Jia X, Gu XJ, Sempau J, Choi D, Majumdar A, Jiang SB. Development of a GPU-based Monte Carlo dose calculation code for coupled electron-photon transport. *Phys Med Biol*. 2010;55(11):3077-3086.
19. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88.
20. Javai U, Dasnoy D, Lee JA. Multi-organ segmentation of chest CT images in radiation oncology: comparison of standard and dilated UNet. *Lect Notes Comput Sci*. 2018;11182:188-199.
21. Ma YY, Mao JF, Liu XG, et al. Deep learning-based internal gross target volume definition in 4D CT images of lung cancer patients. *Med Phys*. 2023;50(4):2303-2316.
22. Zhang XY, He PB, Li YZ, et al. DR-only carbon-ion radiotherapy treatment planning via deep learning. *Phys Med*. 2022;100:120-128.
23. Shan HM, Zhang Y, Yang QS, et al. 3D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2D trained network. *IEEE Trans Med Imaging*. 2018;37(12):2750-2750.
24. Javai U, Souris K, Dasnoy D, Huang S, Lee JA. Mitigating inherent noise in Monte Carlo dose distributions using dilated U-Net. *Med Phys*. 2019;46(12):5790-5798.
25. Neph R, Lyu QH, Huang YSB, Yang YM, Sheng K. DeepMC: a deep learning method for efficient Monte Carlo beamlet dose calculation by predictive denoising in magnetic resonance-guided radiotherapy. *Phys Med Biol*. 2021;66(3):35022.
26. Bai T, Wang BL, Nguyen D, Jiang S. Deep dose plugin: towards real-time Monte Carlo dose calculation through a deep learning-based denoising algorithm. *Mach Learn-Sci Technol*. 2021;2(2):25033.
27. Souris K, Lee JA, Sterpin E. Fast multipurpose Monte Carlo simulation for proton therapy using multi- and many-core CPU architectures. *Med Phys*. 2016;43(4):1700-1712.
28. Jan S, Santin G, Strul D, et al. GATE: a simulation toolkit for PET and SPECT. *Phys Med Biol*. 2004;49(19):4543-4561.
29. Sarrut D, Bardies M, Bousson N, et al. A review of the use and potential of the GATE Monte Carlo simulation code for radiation therapy and dosimetry applications. *Med Phys*. 2014;41(6):064301.
30. Jan S, Benoit D, Becheva E, et al. GATE V6: a major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy. *Phys Med Biol*. 2011;56(4):881-901.
31. Grevillot L, Bertrand D, Dessy F, Freud N, Sarrut D. A Monte Carlo pencil beam scanning model for proton treatment plan simulation using GATE/GEANT4. *Phys Med Biol*. 2011;56(16):5203-5219.
32. Soukup M, Fippel M, Alber M. A pencil beam algorithm for intensity modulated proton therapy derived from Monte Carlo simulations. *Phys Med Biol*. 2005;50(21):5089-5104.
33. Zhang H, Dai ZY, Liu XG, et al. A novel pencil beam model for carbon-ion dose calculation derived from Monte Carlo simulations. *Phys Medica*. 2018;55:15-24.
34. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017 *IEEE International Conference on Computer Vision (Iccv)*. 2017;2242-2251.
35. Bangert M, Wieser H, Stadler A, Jaekel O. MatRad—An open-source multi-modality treatment planning toolkit for educational purposes. *Med Phys*. 2017;44(6):3067-3068.
36. Zhang H, Li Q, Liu XG, et al. Validation and testing of a novel pencil-beam model derived from Monte Carlo simulations in carbon-ion treatment planning for different scenarios. *Phys Med*. 2022;99:1-9.
37. Agostinelli S, Allison J, Amako K, et al. GEANT4—a simulation toolkit. *Nucl Instrum Meth A*. 2003;506(3):250-303.
38. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention, Pt Iii*. 2015;9351:234-241.
39. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Adv Neur In*. 2014;27:2672-2680.
40. Han K, Wang YH, Tian Q, Guo JY, Xu CJ, Xu C. GhostNet: more features from cheap operations. 2020 *IEEE/Cvf Conference on Computer Vision and Pattern Recognition (Cvpr)*. 2020;1577-1586.
41. Luo JH, Wu JX, Lin WY. ThiNet: a filter level pruning method for deep neural network compression. 2017 *IEEE International Conference on Computer Vision (Iccv)*. 2017;5068-5076.
42. Rastegari M, Ordonez V, Redmon J, Farhadi A. XNOR-Net: imageNet classification using binary convolutional neural networks. *Computer Vision—Eccv*. 2016;9908:525-542. 2016, Pt Iv.
43. You S, Xu C, Xu C, Tao DC. Learning from Multiple Teacher Networks. *Kdd'17: Proceedings of the 23rd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*. 2017;1285-1294.

How to cite this article: Zhang X, Zhang H, Wang J, et al. Deep learning-based fast denoising of Monte Carlo dose calculation in carbon ion radiotherapy. *Med Phys*. 2023;50:7314–7323. <https://doi.org/10.1002/mp.16719>