

Fast and accurate dose predictions for novel radiotherapy treatments in heterogeneous phantoms using conditional 3D-UNet generative adversarial networks

Florian Mentzel¹ | Kevin Kröninger¹ | Michael Lerch^{2,3} | Olaf Nackenhorst¹ | Jason Paino^{2,3} | Anatoly Rosenfeld^{2,3} | Ayu Saraswati⁴ | Ah Chung Tsoi⁴ | Jens Weingarten¹ | Markus Hagenbuchner⁴ | Susanna Guatelli^{3,4}

¹Department of Physics, TU Dortmund University, Dortmund, Germany

²Centre for Medical Radiation Physics, University of Wollongong, Wollongong, New South Wales, Australia

³Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong, New South Wales, Australia

⁴School of Computing and Information Technology, University of Wollongong, Wollongong, New South Wales, Australia

Correspondence

Florian Mentzel, Otto-Hahn-Strasse 4a, 44227 Dortmund, Germany
Email: florian.mentzel@tu-dortmund.de

Funding information

Large-Scale Equipment Initiative by the German Research Foundation (DFG), Grant/Award Number: 271512359

Abstract

Purpose: Novel radiotherapy techniques like synchrotron X-ray microbeam radiation therapy (MRT) require fast dose distribution predictions that are accurate at the sub-mm level, especially close to tissue/bone/air interfaces. Monte Carlo (MC) physics simulations are recognized to be one of the most accurate tools to predict the dose delivered in a target tissue but can be very time consuming and therefore prohibitive for treatment planning. Faster dose prediction algorithms are usually developed for clinically deployed treatments only. In this work, we explore a new approach for fast and accurate dose estimations suitable for novel treatments using digital phantoms used in preclinical development and modern machine learning techniques. We develop a generative adversarial network (GAN) model, which is able to emulate the equivalent Geant4 MC simulation with adequate accuracy and use it to predict the radiation dose delivered by a broad synchrotron beam to various phantoms.

Methods: The energy depositions used for the training of the GAN are obtained using full Geant4 MC simulations of a synchrotron radiation broad beam passing through the phantoms. The energy deposition is scored and predicted in voxel matrices of size $140 \times 18 \times 18$ with a voxel edge length of 1 mm. The GAN model consists of two competing 3D convolutional neural networks, which are conditioned on the photon beam and phantom properties. The generator network has a U-Net structure and is designed to predict the energy depositions of the photon beam inside three phantoms of variable geometry with increasing complexity. The critic network is a relatively simple convolutional network, which is trained to distinguish energy depositions predicted by the generator from the ones obtained with the full MC simulation.

Results: The energy deposition predictions inside all phantom geometries under investigation show deviations of less than 3% of the maximum deposited energy from the simulation for roughly 99% of the voxels in the field of the beam. Inside the most realistic phantom, a simple pediatric head, the model predictions deviate by less than 1% of the maximal energy deposition from the simulations in more than 96% of the in-field voxels. For all three phantoms, the model generalizes the energy deposition predictions well to phantom geometries, which have not been used for training the model but are interpolations of the training data in

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine.

multiple dimensions. The computing time for a single prediction is reduced from several hundred hours using Geant4 simulation to less than a second using the GAN model.

Conclusions: The proposed GAN model predicts dose distributions inside unknown phantoms with only small deviations from the full MC simulation with computations times of less than a second. It demonstrates good interpolation ability to unseen but similar phantom geometries and is flexible enough to be trained on data with different radiation scenarios without the need for optimization of the model parameter. This proof-of-concept encourages to apply and further develop the model for the use in MRT treatment planning, which requires fast and accurate predictions with sub-mm resolutions.

KEYWORDS

deep learning, dose prediction, generative adversarial networks, novel treatments, synchrotron radiation therapy

1 | INTRODUCTION

A variety of highly conformal treatment techniques such as intensity-modulated radiation therapy¹ (IMRT) and volumetric modulated arc therapy² (VMAT) have been developed in recent decades.³ These techniques have helped to reduce the healthy tissue dose during the irradiation of tumor cells and have led to strong improvements in the treatment of many types of cancer.^{4,5} However, for certain types of cancer, such as glioblastoma, the overall clinical outcome has seen little improvement.⁶ One potential treatment technique that has shown promising results in preclinical studies to treat these aggressive types of cancer, is microbeam radiation therapy (MRT), which is based on polarized high-intensity, spatially fractionated synchrotron X-ray beams.^{7–9} The treatment planning of such beams requires dose calculations with a spatial resolution of a few micrometers and are currently performed using time-consuming Monte Carlo (MC) simulations.⁸ Accurate predictions of the dose distributions inside a phantom can take days using such MC simulations.¹⁰

Several faster approximations for MRT dose distribution computations have been investigated in recent years. The fastest reported computation method comprises the integration of MRT in the Varian Eclipse treatment planning software.¹¹ The approach calculates energy depositions based on the radiological depth and yields short computation times but cannot provide accurate dose predictions in heterogeneous phantoms (overestimation of approximately 20% of the dose in lung tissue inside water) and at high resolution (only macroscopic treatment of the dose calculation). Another approach uses a point kernel algorithm, which results in an execution time of approximately 5 min for the prediction of the dose inside a phantom but is found to lead to large deviations compared to the full MC simulations, which are especially present at interfaces between tissues.¹² The agreement to full MC simulations is improved using a hybrid model, which

combines a simplified MC simulation with an analytical dose kernel approach.¹⁰ This method is reported to take approximately 30 min to compute the dose distribution inside a phantom, which is still relatively long for a clinical application of a treatment planning system.

The goal of our work is to develop an algorithm, which is fast enough for real-world treatment planning systems and accurate enough for high-resolution radiation therapy such as MRT. In the presented study, we take a first step in that direction by developing a general and flexible model for the fast and accurate prediction of dose distributions inside phantoms using modern machine learning techniques.

In recent years, machine learning has been used in an increasing number of publications for fast treatment planning of commercially available therapies.¹³ Most of these studies have in common, that neural networks are used to generate a full treatment plan based on a database of already delivered treatments.¹⁴ For novel treatments, such as MRT, existing treatment plans are usually not available. More recent developments include the computation of dose distributions using computed tomography (CT) scans of a large number of patients as training data for neural networks.^{15–17} In the case of novel treatments like MRT, many different targets from slab phantoms to rats are usually subject to *in-silico* and experimental investigations (e.g., by Engels et al.⁸). To allow development work on the machine learning model prior to the acquisition of dedicated CT data, we present a fully MC simulation-based approach using Geant4¹⁸ and digital phantoms, which is possible for any kind of new therapy method, as long as a model for the MC simulation exists. The use of digital phantoms, rather than CT-based phantom geometries, in addition allows to generate accurate MC training data faster and to address eventual development problems more efficiently.

The machine learning algorithm presented in this paper is based on generative adversarial networks (GANs).¹⁹ GANs have been successfully applied to a

variety of complex data generation tasks²⁰ and have recently also been employed in the field of dose distribution computations for radiotherapy.^{21,22} The basic concept of GANs is to use two competing neural networks: A generator network that learns to generate new data samples similar to given training samples, and a discriminator (or critic) network, which is trained to distinguish those generated samples from the ones of the training data. Feedback from the discriminator enables the generator to learn to generate output, which is indistinguishable from the training data, while the discriminator improves its level of distinction by being presented with more and more similar data from the generator. A recent study²² found that, especially for the case of dose predictions in heterogeneous material resulting in high-dose gradients, GANs are more accurate in their predictions compared to regression networks due to their dynamically changing objective function, which can capture more volume-level information than a localized loss function. While the training of GANs is a complex process and often challenging and time-consuming, the prediction using only the generator is usually very fast, which makes them an ideal candidate for our purposes.

To our knowledge, there is no prior published study that includes every step of the development process: the creation of the required digital phantoms, the design of full MC simulations, which are adapted to a special irradiation modality and are experimentally validated, and finally, training of the studied ML model with detailed investigation of its performance. While there are many publications about dose predictions using deep learning models, only few publications that we know of aim at similar research goals. For example, Zang et al.²³ published a very notable study on dose predictions with a GAN in the case of proton irradiation, they did not compare their predictions though on a voxel-by-voxel base, especially around tissue borders. This makes it difficult to inspect the usability for high-gradient applications. In the work of Kearney et al.,²² GANs were used for dose prediction and results were compared with MC simulations as well, but no detailed comparison of depth–dose curves is shown as the focus remains on clinical parameters like dose–volume histograms. More detailed comparisons were conducted, for example, in the work of Kontaxis et al.,¹⁷ but as they use only prostate CT scans for all their data, it is difficult to understand how the training and test data really differ, even if taken from individual patients. Our approach using digital phantoms allows for a more straightforward understanding of differences between training and test.

As a proof-of-concept, we apply our GAN model to three simplified irradiation scenarios, which are incrementally increasing in their complexity: a water phantom with a rotated bone slab, a water phantom with a rotated bone slab of variable thickness, and a pediatric head phantom. For simplicity, we use a broad beam instead of a microbeam for the irradiation of the phantoms and

require a spatial resolution of 1 mm^3 in the dose prediction.

This paper is structured as follows: Section 2 describes the MC simulation using Geant4 and the developed machine learning model. In Section 3 the dose predictions of the GAN obtained for the three scenarios are presented and compared to the full MC simulation. Finally, in Section 4 the results are discussed and in Section 5 conclusions are made.

2 | MATERIALS AND METHODS

The development of the digital phantoms and the dose computation model follows an iterative-incremental approach, which is recognized to a successful software process, adopted also in the development of Geant4.^{24,25} Following this software process, the machine learning dose prediction model is trained on the simpler configurations described in this section. This allows to generate accurate MC data more rapidly compared to CT-based simulations and thereby to address eventual development problems more easily and to develop a more robust software product. As MRT is at preclinical stage and dosimetric studies are performed in phantoms and in pets, two digital phantoms that are commonly used in preclinical research are presented in this work: a slab and an ellipsoid head phantom.

2.1 | MC simulation

The photons of the broad beam and their interactions with the material of the various phantoms are simulated using Geant4 10.6p01¹⁸ with option 4 of the standard electromagnetic physics constructor, which is adapted to model the interactions of the polarized photons with the target (Livermore Polarized Physics).²⁶ The beam is simulated using the description of the Imaging and Medical beam line (IMBL) at the Australian Synchrotron²⁷ and corresponds to a typical highly polarized synchrotron beam²⁸ as used for preclinical research for MRT,^{8,29} before passed through a multislit collimator to produce the microbeams. The energy spectrum of the X-ray beam is shown in Figure 1a. The initial beam traverses a tungsten mask in order to obtain a beam cross section of $8 \times 8\text{ mm}^2$ before it hits the various phantoms. The lateral beam profile together with the in-field and out-of-field definition is shown in Figure 1b. The phantoms are placed at a distance of 4 cm behind the mask to reduce back-scattering effects, which is shown exemplary in Figure 2a for a cubic water phantom with a size of $14 \times 14 \times 14\text{ cm}^3$. The tungsten mask (dark gray) and the water phantom (blue) are surrounded by a total simulation volume (light gray) of size $30 \times 30 \times 30\text{ cm}^3$. For all simulated phantoms, the energy depositions are calculated for each voxel of size $1 \times 1 \times 1\text{ mm}^3$ in

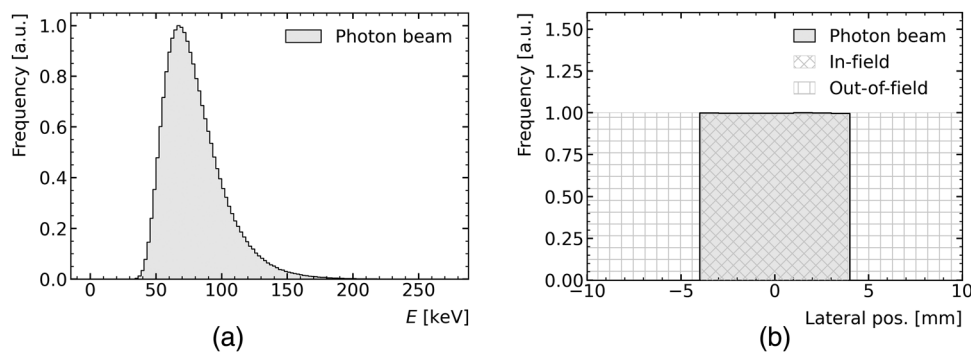


FIGURE 1 (a) Energy spectrum of X-ray photons in keV as used as input to the simulation and as derived from Dipuglia et al.²⁹ (b) Lateral beam profile after the tungsten mask indicating the in-field and out-of-field region of the beam

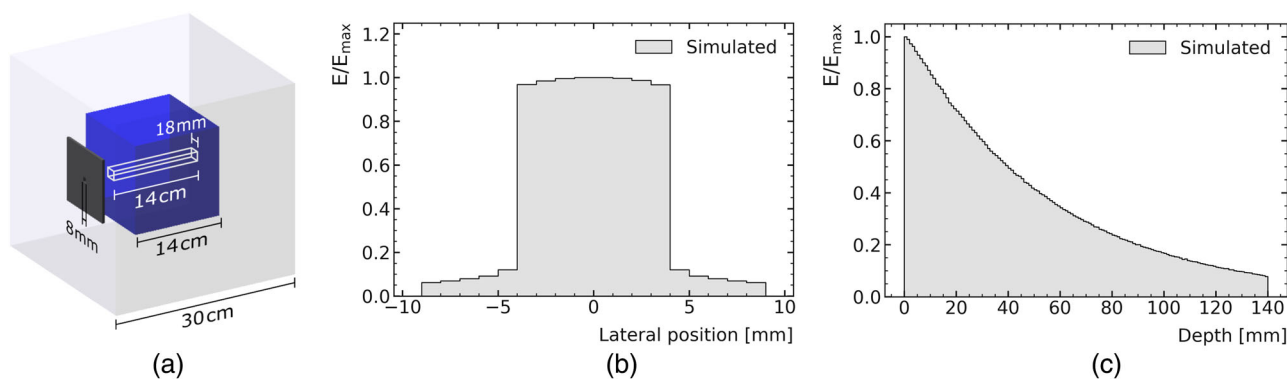


FIGURE 2 (a) Schematic of the simulation setup of a cubic water phantom including the world volume (light gray), the tungsten mask (dark gray), the water phantom (blue), and the scoring volume (outlined in white). (b) Normalized energy deposition inside the water phantom along the depth at the center of the beam. (c) Normalized lateral energy deposition inside the water phantom integrated along beam direction. The energy deposition is normalized to the maximum value E_{\max} of the in-field region of the beam

the scoring volume of $D \times W \times H = 140 \times 18 \times 18 \text{ mm}^3$, which starts at the surface of the phantom and is centered around the beam. The one-dimensional (1D) projection of the normalized energy depositions of the beam inside this water phantom along the beam (depth) and the profile perpendicular to the beam (lateral) are shown in Figure 2b and Figure 2c, respectively.

The Geant4-based data samples were simulated using the LiDO3 cluster at TU Dortmund University, which took approximately 2 weeks using on average 500 Intel Xeon cores simultaneously. All simulated data are divided into *training data*, which are presented to the neural network for weight optimization, *validation data*, which are used to identify the best model parameters and *test data*, which are exclusively used to evaluate the performances after the best machine learning model was chosen.

2.1.1 | Water phantom with rotated bone slab of constant thickness

The simplest phantom considered in this studies, consist of a water cube as described above with a centrally

inserted compact bone slab (ICRU, $\rho = 1.85 \text{ g/cm}^3$) with a constant thickness of $d = 2.5 \text{ mm}$, which is rotated continuously with angles of $\alpha \in [0, 87]^\circ$ around an axis perpendicular to the beam axis. A schematic of the phantom including the tungsten mask is shown in Figure 3a with an exemplary rotation angle of $\alpha = 45^\circ$. Figure 3b,c shows the normalized energy depositions of the beam inside the phantom along the beam axis and along the rotation axis integrated over the beam axis, respectively. Due to the rotation of the bone slab, the energy deposition peaks slightly before the center of the phantom (70 mm in depths), where the bone slab is inserted.

The simulated data of this phantom serve as a feasibility study to develop a flexible GAN model, which is general enough to be applied to more complex phantom geometries and predicts at the same time accurately the simulated energy depositions inside the phantom. The traversed distance of the beam within the bone material and, as a result, the energy depositions change more drastically at larger rotation angles. To increase the sampling density of the simulation at those larger rotation angles, the rotation angles of the simulation are determined by drawing random numbers from a $1/\cos(\alpha)$ distribution. The simulated data are split into training data

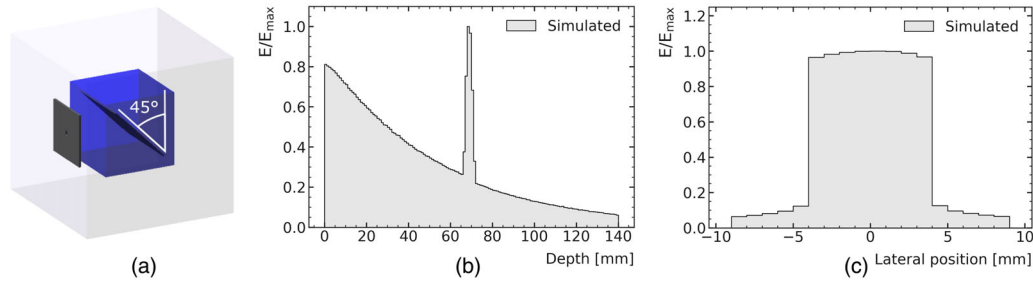


FIGURE 3 (a) Schematic of the simulation setup of a cubic water phantom with inserted bone slab including the world volume (light gray), the tungsten mask (dark gray), the water phantom (blue), and the bone slab (dark blue) rotated at 45° (white). (b) Normalized energy deposition inside the water phantom with the rotated bone slab along the depth at the center of the beam. (c) Normalized lateral energy deposition inside the water phantom with the rotated bone slab integrated along beam direction. The energy deposition is normalized to the maximum value E_{\max} of the in-field region of the beam

with rotation angles $\alpha \in \{[0, 3) \cup [7, 23) \cup [27, 43) \cup [47, 63) \cup [67, 78) \cup [82, 87]\}^\circ$, validation data with rotation angles of $\alpha \in \{[5, 7) \cup [25, 27) \cup [45, 47) \cup [65, 67) \cup [80, 82]\}^\circ$ and test data with rotation angles of $\alpha \in \{[3, 5) \cup [23, 25) \cup [43, 45) \cup [63, 65) \cup [78, 80)\}^\circ$. Here, the symbol “[” denotes an inclusive interval limit, “)” means an exclusive interval limit, and “ \cup ” the union of the individual intervals. The resulting number of simulated samples and the split in training, validation, and test set is shown in Figure 4. The purpose of this split is to have unbiased simulation data for the performance evaluation during the training and hyperparameter optimization (validation data) as well as for the final assessment of the performance (test data). At the same time this allows for investigating how well the GAN model can interpolate in one dimension.

2.1.2 | Water phantom with rotated bone slab of variable thickness

The second phantom is identical to the first phantom with the exception that the thickness of the bone slab is varied in addition to varying rotation angle. The pur-

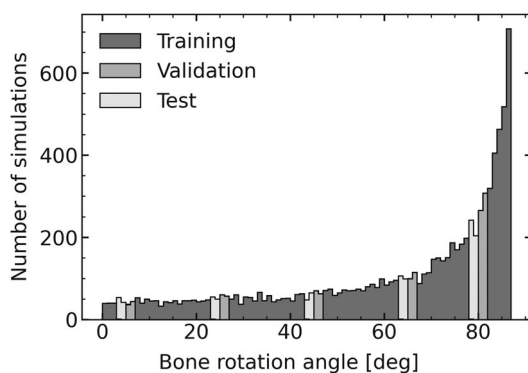


FIGURE 4 Number of simulated training (dark gray), validation (medium gray), and test samples (light gray) as a function of bone slab rotation angle

pose of the simulated data using this phantom is to demonstrate the interpolation capabilities of the GAN model in two dimensions. This is of particular importance, because for future applications in treatment planning, the model needs to be able to accurately predict energy depositions for unknown geometries, which have not been used for the training of the GAN. Multidimensional interpolation is in general not an easy task for a generative model, which is why we consider this as an important step before moving to more complex phantom geometries. For this purpose, the simulated data are split into training data with bone slab thicknesses of $d \in \{1, 2.5, 5, 10\}$ mm and rotation angles of $\alpha \in \{[4, 17) \cup [23, 37) \cup [43, 57) \cup [63, 84) \cup [86, 87]\}^\circ$. The simulated data of the same thicknesses but with rotation angles not used during the training and in the range of $[0, 87]^\circ$ are used as validation data. In order to investigate the ability of the GAN to predict the energy depositions inside unknown phantom geometries, the test data are simulated with bone slab thicknesses of $d \in \{1.75, 4, 7\}$ mm but continuously in the full rotation angle range of $[0, 87]^\circ$. The split of the simulated data is illustrated in Figure 5.

2.1.3 | Simple pediatric head phantom with translation

The third phantom is a simple model of a pediatric head. The skull of the head is constructed using two rotational ellipsoids (spheroids). The first spheroid has a length of 15 cm and a width/height of 13 cm. The second spheroid with a length of 14.25 cm and width/height of 12.6 cm is located inside the first one and is slightly shifted by 1.25 mm from the center to the front. The volume in between the two spheroids corresponds to the skull and consists of compact bone (ICRU, $\rho = 1.85 \text{ g/cm}^3$), resulting in bone thicknesses of 2.5 mm at the forehead, of 5 mm at the back of the head and of 2 mm at the sides. The obtained skull is filled with water in order to mimic the

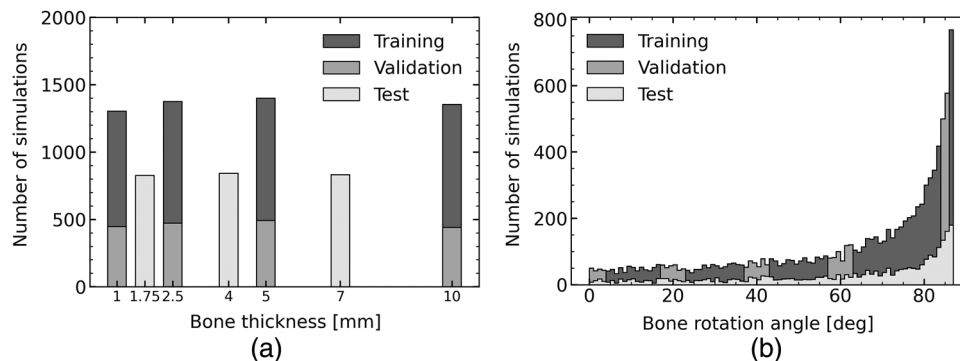


FIGURE 5 Number of simulated training (dark gray), validation (medium gray), and test samples (light gray) as a function of (a) bone slab rotation angle and (b) bone slab thickness

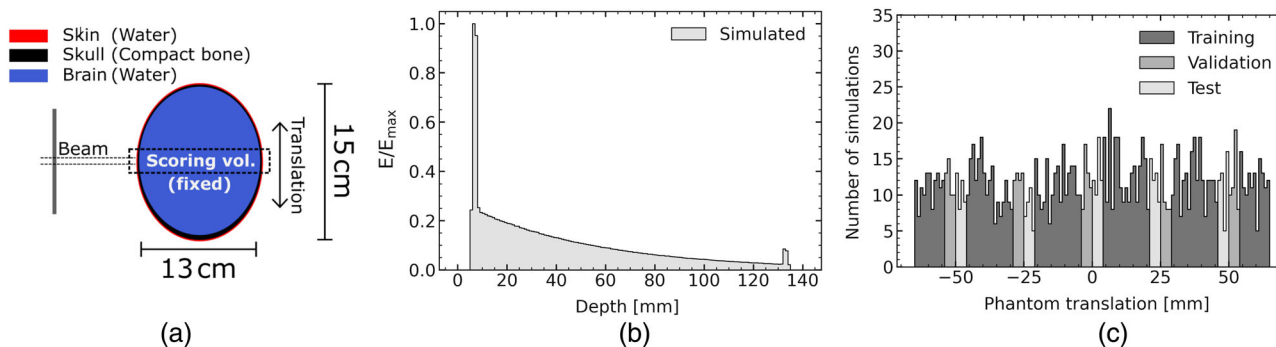


FIGURE 6 (a) Schematic of the simulation setup of a simple pediatric head phantom showing the tungsten mask (dark gray), the incoming beam, the skin (red), the skull (black), the brain (blue), and the scoring volume and the fixed scoring volume (outlined by dotted line). (b) Normalized energy deposition inside the water phantom along the depth at the center of the beam with translation $t = 0$ mm. The energy deposition is normalized to the maximum value E_{max} of the in-field region of the beam. (c) Number of simulated training (dark gray), validation (medium gray), and test samples (light gray) as a function of phantom translation. Normalized lateral energy deposition inside the water phantom integrated along the beam direction. The energy deposition is normalized to the maximum value E_{max} of the in-field region of the beam

brain tissue and surrounded by a thin layer (1 mm) of water in order to model the skin. The head is positioned orthogonal to the beam, such that the radiation would be performed centrally from the side. A schematic of the phantom and the tungsten mask is shown in Figure 6a. The simulation is performed with translations t of the head orthogonal to the beam in the range of $t \in [-65, 65]$ mm along the longitudinal axis of the spheroid in order to simulate any possible position of the beam along the equatorial ellipse. The number of simulated samples is uniformly distributed over the full translation range. Due to the translation of the head, the thickness, shape, angle, and position of the radiated bone changes at the same time. An exemplary normalized energy deposition inside the phantom at the center of the head with $t = 0$ mm is shown in Figure 6b. The two energy peaks result from the bones at the front and the back of the skull.

The purpose of this simulated data is to investigate the GANs ability to predict energy depositions in more complex geometries and to interpolate the predictions in dynamically changing geometries in multiple dimen-

sions. Although the head phantom has a relatively simple geometry compared to realistic heads, this study is an important milestone and a proof-of-concept for future applications aiming for real-life treatment planning. The simulated data are split in training data with $t \in \{-65, -54\} \cup [-46, -29] \cup [-21, -4] \cup [4, 21] \cup [29, 46] \cup [54, 65]$ mm, in validation data with $t \in \{-54, -50\} \cup [-29, -25] \cup [-4, 0] \cup [25, 29] \cup [50, 54]$ mm and in test data with $t \in \{-50, -46\} \cup [-25, -21] \cup [0, 4] \cup [21, 25] \cup [46, 50]$ mm as visualized in Figure 6c.

2.2 | The GAN model

In the following we describe and motivate our model, which was using the simulated data of the first phantom. We use a conditional GAN³⁰ with a Wasserstein loss function³¹ and a Gradient Penalty term³² to regularize and stabilize the weight updates, which are performed using the Adam optimizer³³ with an initial learning rate of $\alpha = 2 \cdot 10^{-5}$ and a batch size of 32. As the output of

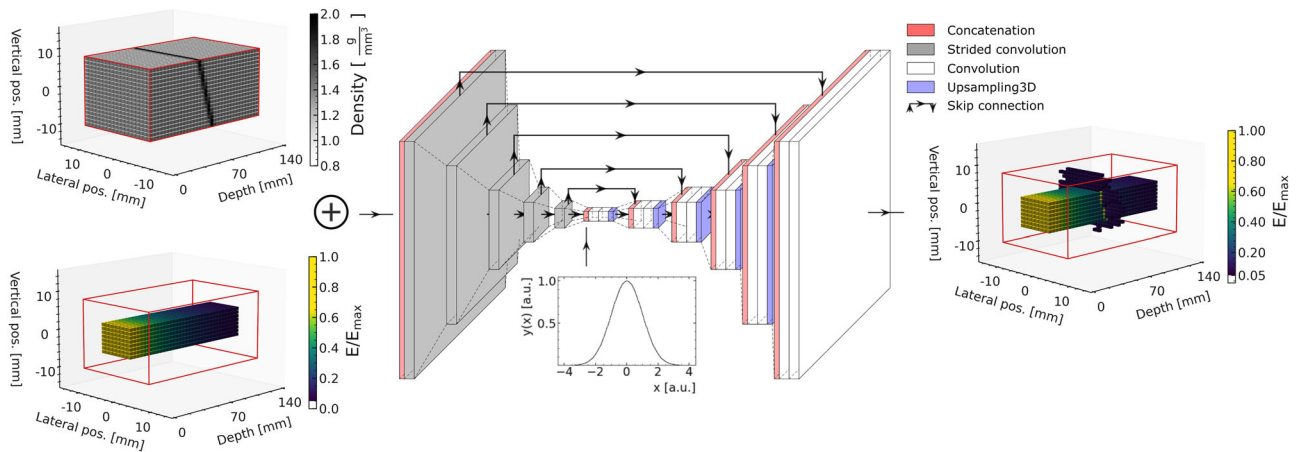


FIGURE 7 Schematic of the generator 3D U-Net structure with input and output data. Left: Conditional generator input consisting of a 3D representation of a material density matrix (top) and a 3D matrix of the energy deposition of the photon beam in a water phantom (bottom). Center: Convolutional encoder and decoder connected by skip connections. The normal distribution below indicates the injection of random noise in the bottleneck layer. Right: Generator output consisting of a 3D representation of energy depositions inside the phantom

Wasserstein GANs (WGANs) reflects not only a qualitative but also a quantitative assessment by using an approximation of the Earth-Mover's distance as loss, the discriminator is usually called critic instead. One batch of simulation data are used to update the weights of the critic five consecutive times, before they are used to update the weights of the generator a single time. This ensures that the critic is strong enough to provide valuable information to the training of the generator.

2.2.1 | The generator model

A schematic of the generator model is shown in Figure 7. The goal of the generator is to predict the energy depositions of the photon beam used for the radiation therapy inside a three-dimensional (3D) body, for example, a phantom or a patient. In our case, the output of the generator is thus a 3D matrix of energy depositions of size $140 \times 18 \times 18$ corresponding to the voxelized scoring volume of the simulation (right). The prediction is dependent on the beam properties as well as the localisation and properties of the material inside the body, for example, the localization and density of bone, skin, or water. Hence, the generator network of the GAN is conditioned on this information.³⁰ The first conditional information of the generator are the 3D energy depositions of a beam inside a pure water phantom (lower left), which describes the relevant properties of the beam (energy spectrum, beam profile, polarization, etc.), and at the same time gives the network information about how the beam interacts with a default material (water) in the same format ($140 \times 18 \times 18$) as the prediction is made. Introducing this conditional information considerably accelerates and stabilizes the training and yields more accurate predictions. The second conditional infor-

mation is the 3D material density matrix of the phantom model (upper left) of dimension $140 \times 18 \times 18$ as used for the simulation. This information is of course not available for a real-life application but would be replaced by the 3D medical imaging information of the patient, such as the Hounsfield units of a CT scan. As the conditional input of the generator consists of two high-dimensional matrices ($140 \times 18 \times 18 \times 2$), the dimensionality of this information is reduced by encoding the 3D images with multiple strided 3D convolutions into low-dimensional representations ($3 \times 3 \times 3$) of the data combining the geometrical properties of the phantom with the information obtained from the energy depositions of the beam in water (center). The compressed features are then concatenated with 100 values of random noise drawn from a Gaussian distribution to statistically predict the energy depositions inside the phantom by decoding the images again to dimension $140 \times 18 \times 18$ using 3D up-samplings with subsequent convolutions. Each convolutional layers of the generator consists of 64 filters of variable size, is activated using the Swish function,³⁴ stabilized using batch normalization,³⁵ and regularized using dropout³⁶ with a rate of 15%. Because not only the low-dimensional features contain important information for the prediction but also the intermediately obtained features of the encoder, we introduce skip connections to pass information from the same level of the encoder to the same level of the decoder by concatenating the two layers and effectively doubling the number of the filters in the decoder. This obtained structure using a convolutional encoder-decoder with skip connection is usually referred to as U-Net structure, which was first introduced for fast 2D biomedical image segmentation.³⁷ A similar structure of a 3D U-Net³⁸ has recently been applied for dose distribution estimation in IMRT.¹⁷

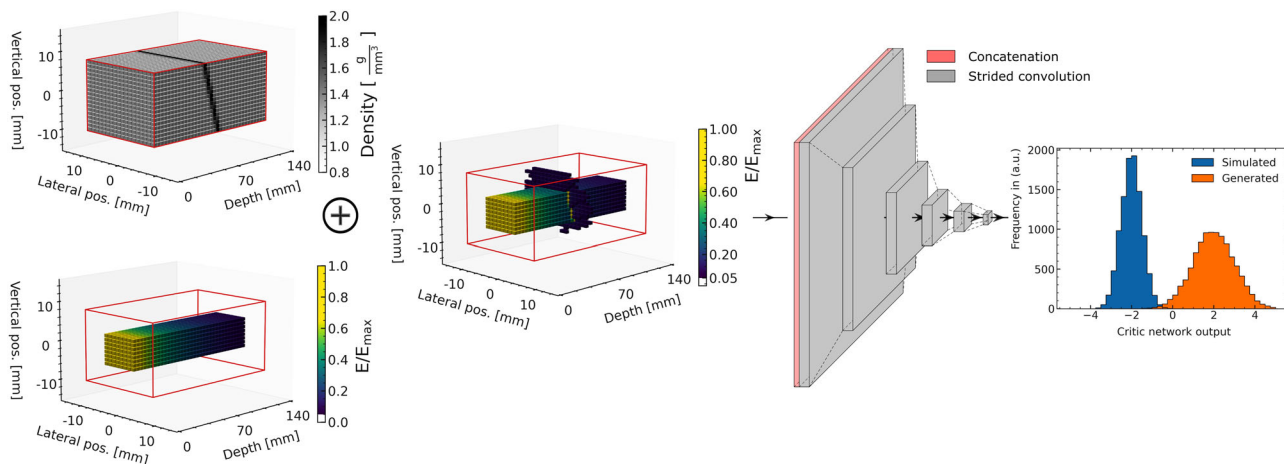


FIGURE 8 Schematic of the critic network with input and output data. Left: Conditional generator input consisting of a 3D representation of a material density matrix (top left) and a 3D matrix of the energy deposition of the photon beam in a water phantom (bottom left) as well as the simulated or generated 3D energy depositions inside the phantom (right). Center: The critic architecture with six subsequent convolutional layers. Right: The output is a Wasserstein loss rating the input energy depositions as simulated or generated

2.2.2 | The critic model

The critic of the GAN is built as a relatively simple 3D convolutional network. A schematic of the critic model is shown in Figure 8. In addition to the same conditional information as given to the generator, the critic receives either the simulated or generated energy depositions inside the phantom as input in order to classify these two types of data (left). The three input matrices of dimension $140 \times 18 \times 18$ are concatenated and passed through six consecutive transposed 3D convolutions of decreasing filter size (center). Each convolutional layer of the critic consists of 128 filters, is activated using the Swish function and regularized using a dropout rate of 15%. The last layer of the critic consists of a single linear unit resulting in a continuous and quantitative rating distinguishing simulated from generated samples (right).

2.2.3 | Performance measure

In many machine learning applications, the development of the loss over the training process is used as an indication whether the optimization of the weights is converged. In the training of GANs, the loss is usually not a good measure to assess the convergence of the two networks to a stable equilibrium or to judge about the performance of the generator. Therefore, we define the delta index, which is inspired by the global gamma index,³⁹ from which we can derive a performance measure for the prediction quality of the generator. The delta index is defined as

$$\delta = \frac{D_{\text{gen}} - D_{\text{sim}}}{D_{\text{sim}}^{\text{max}}}, \quad (1)$$

where D_{gen} is the dose predicted by the generator network of the GAN, D_{sim} is the dose obtained by the Geant4 simulation, and $D_{\text{sim}}^{\text{max}}$ is the maximum dose deposition in the phantom as calculated in the MC simulation. The delta index is calculated for each voxel of the in-field area of the beam inside the scoring volume. As global performance measure we use the 1% (3%) *passing rate*, which is the fraction of the voxels with a delta index smaller than 1% (3%). The model of the generator is chosen for the final performance evaluation, which maximizes the 1% passing rate on the validation data during the training.

3 | RESULTS

In this section we compare the predictions of the GAN model for the three different phantoms with the MC simulation and use the passing rate in order to summarize the prediction accuracy. The machine learning models are implemented using the Keras⁴⁰ interface to TensorFlow 2.2⁴¹ on a Nvidia GeForce GTX 1080i graphical processing unit (GPU). Training took several days for each model.

3.1 | Water phantom with rotated bone slab of constant thickness

The development of the 1% (3%) passing rate averaged over the bone slab rotation angles during the training process using the data of the first phantom is shown in Figure 9a for the training and validation data, where one epoch means, that the full training data have been used for updating the weights of the GAN model a single

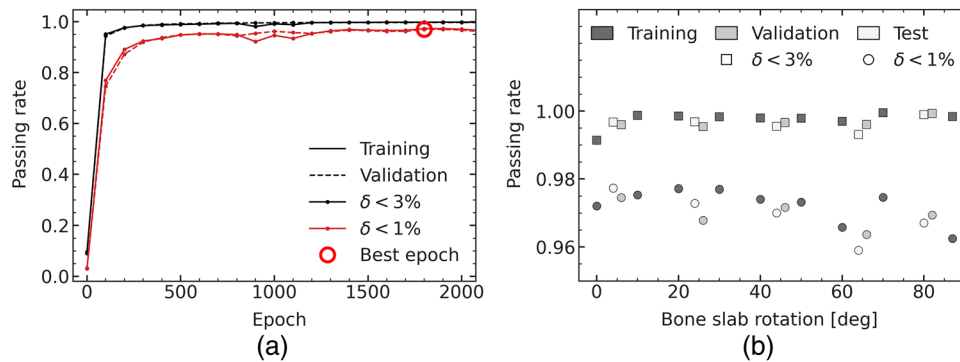


FIGURE 9 (a) Global 1% (red) and 3% (black) passing rates for the training (solid) and validation (dashed) data as a function of training epochs. The best model on the validation data is highlighted with a red circle. (b) Passing rates as a function of the bone slab rotation angle for training (dark gray), validation (medium gray), and test (light gray) data

TABLE 1 Delta index passing rates with standard deviation for training, validation, and test data for the water phantom with the rotated bone slab

	Passing rate (%)		
	Training	Validation	Test
$\delta < 1\%$	97.2 ± 0.5	96.9 ± 0.4	96.9 ± 0.7
$\delta < 3\%$	99.8 ± 0.2	99.7 ± 0.2	99.6 ± 0.2

time. Relatively high passing rates are already obtained after a few hundred epochs, the highest 1% passing rate of $(96.9 \pm 0.4)\%$ is achieved on the validation data after 2800 training epochs. The model from this epoch is chosen as the best model for evaluating the performance of the GAN. Table 1 summarizes the passing rates of the training, validation, and test data averaged over bone slab rotation angles of this model. The passing rates are all at a very high level and agree within their standard deviations, which indicates that the model is not overtrained and is capable of generalizing well to the unseen bone slab rotation angles of the validation and test data. The passing rates of the selected model dependent on the bone slab rotation angle are shown in Figure 9a for the training, validation, and test data. Over the full rotation angle range more than 99% of the voxel deviate by less than 3% of the maximal simulated dose. Although slightly smaller passing rates are obtained for the validation and test data compared to the training data the performance is relatively stable over the full rotation angle range and the GAN shows good ability to interpolate to geometries of unknown rotation angles. Figure 10 shows comparisons of normalized simulated and generated energy depositions of the test data as a depth profile in-field (a-c) and out-of-field (d-f) of the beam with bone slab rotation angles of $\alpha = [4, 64, 80]^\circ$ (a+d, b+e, c+f). In the in-field region of the beam, the relative energy deposition deviations $\Delta E_{\text{rel}} = E_{\text{gen}}/E_{\text{sim}} - 1$ are mostly well within 5% for all rotation angles except for the boundary regions of the bone, where the deviations in individual voxels are larger, and the end-region

of the phantom, where the absolute energy depositions are relatively small. Considering, that the energy depositions change most drastically in the transition regions of water and bone in particular for larger rotation angles and the model could not learn from data around the chosen angles, the agreement over the full depths of the phantom is remarkable and demonstrates a good interpolation capability of the model.

Out-of-field, the relative deviations of the generated energies from the simulated energies are larger but mostly within 25% except for the transition regions. However, the deposited energy is more than one order of magnitude smaller in the out-of-field regions, which means deviations of this magnitude will result in relatively small penalties during the training. Considering this, the energy depositions in the out-of-field region are qualitatively and quantitatively well predicted. In order to get an idea of the accuracy of the prediction in two dimensions, Figure 11 shows the relative dose deviations of voxels as 2D slices of the phantom with a rotation angle of $\alpha = 4^\circ$ (a) and $\alpha = 79^\circ$ (b), which are part of the test data. Mostly voxel in the transition region of water and bone deviate by more than 1% and only few voxels exactly at the edge of the bone deviate by more than 3% in their predictions from the simulation.

3.2 | Water phantom with rotated bone slab of variable thickness

The development of the passing rates averaged over bone slab rotation angles and thicknesses during the training process using the data of the second phantom are shown in Figure 12a for the training and validation data. The training takes significantly longer until a plateau is reached and is less stable. The highest 1% passing rate on the validation data is achieved at epoch 3300, which is chosen as the best model. Table 2 summarizes the passing rates of the training, validation, and test data averaged over bone slab rotation angles

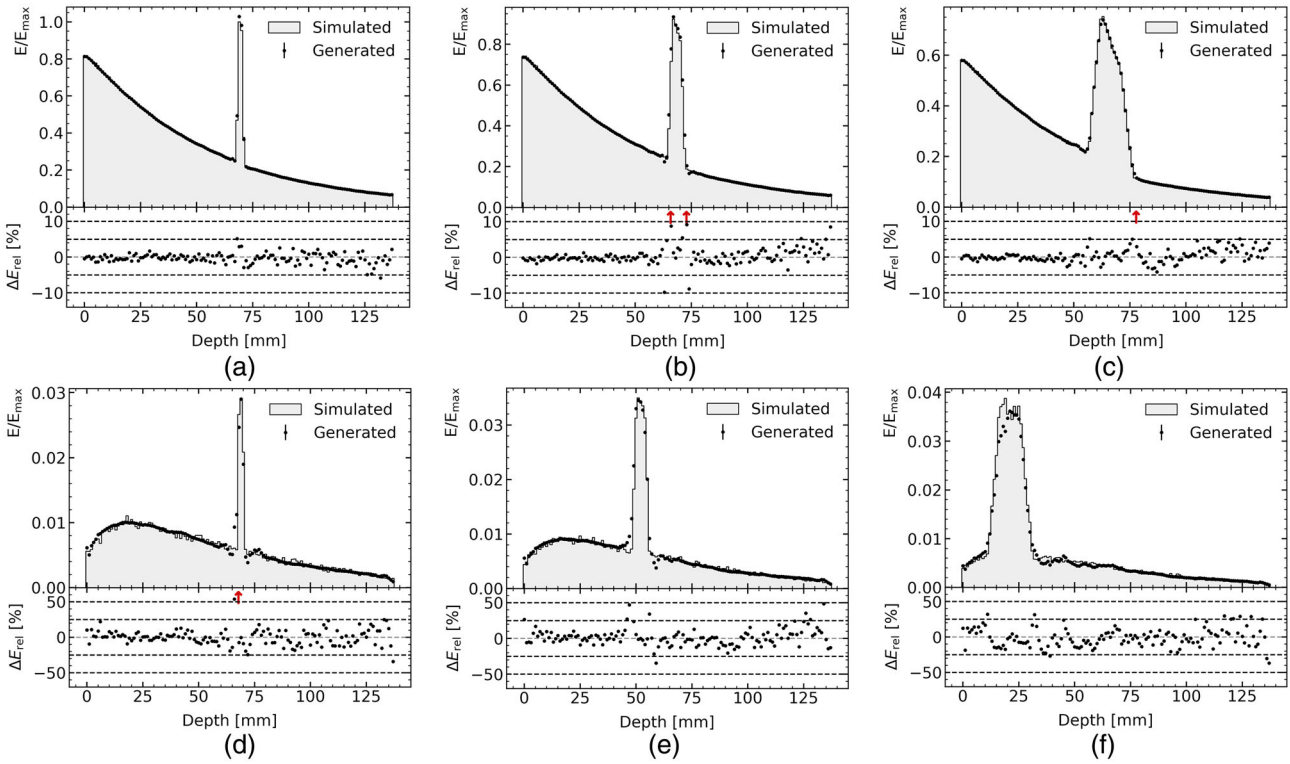


FIGURE 10 Comparisons of normalized simulated (gray) and generated (black) energy depositions of the test data as a depth profile in-field (a-c) and out-of-field (d-f) of the beam with bone slab rotation angles of $\alpha = [4, 64, 80]^\circ$ (a+d, b+e, c+f). The lower part of the plots shows the relative energy deposition deviation $\Delta E_{rel} = E_{gen}/E_{sim} - 1$ in percent. Outliers are indicated with a red arrow

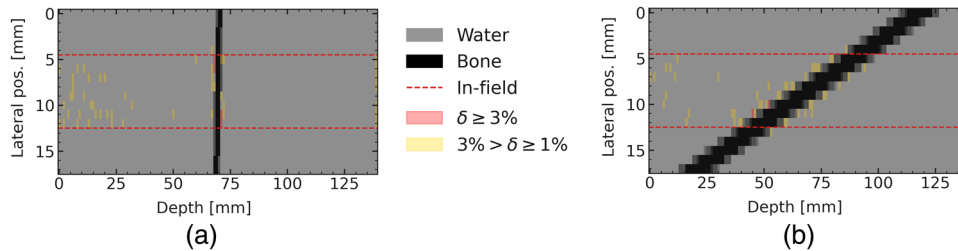


FIGURE 11 2D slices of the phantom with a rotation angle of (a) $\alpha = 4^\circ$ and (b) $\alpha = 79^\circ$ as part of the test data. Voxels with deviations in their dose predictions of larger than 3% (1%) are shown in red (yellow). Whenever the deviation is smaller than 1%, the gray-scale color of the material density matrix is shown. The in-field part of the beam is located between the red dashed lines

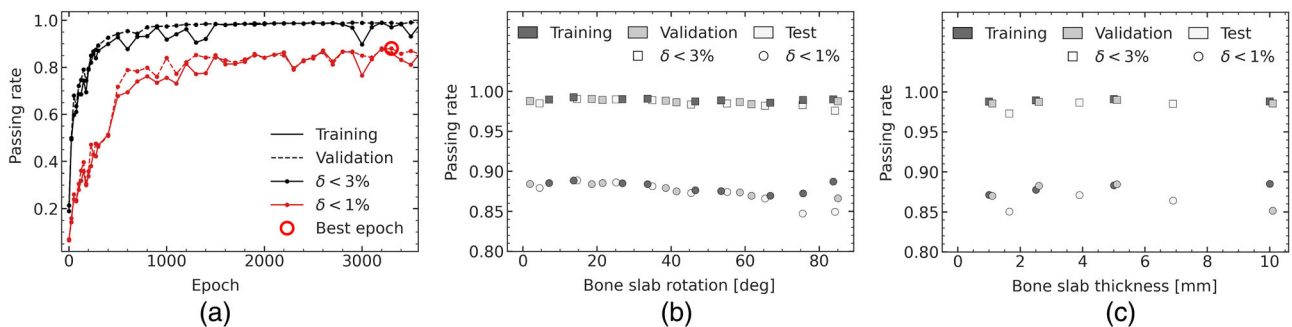


FIGURE 12 (a) Global 1% (red) and 3% (black) passing rates for the training (solid) and validation (dashed) data as a function of training epochs. The best model on the validation data is highlighted with a red circle. (b) Passing rates as a function of the bone slab rotation angle for training (dark gray), validation (medium gray), and test (light gray) data. (c) Passing rates as a function of the bone slab thickness for training (dark gray), validation (medium gray), and test (light gray) data

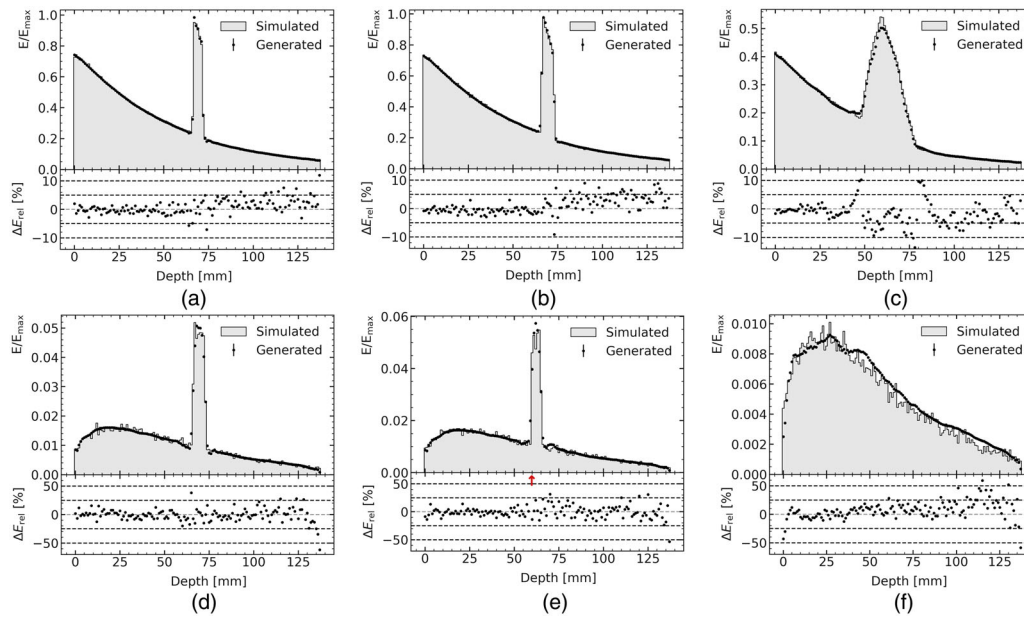


FIGURE 13 Comparisons of normalized simulated and generated energy depositions inside the phantom along the beam using the test data for the in-field (a-c) and out-of-field (d-f) regions of the beam with bone slab rotation angles of $\alpha = [0, 40, 85]^\circ$ (a+d, b+e, c+f) and thicknesses of $d = [7, 4, 1.75]$ mm (a+d, b+e, c+f). The lower part of the plots shows the relative energy deposition deviation ΔE_{rel} in percent. Outliers are indicated with a red arrow

TABLE 2 Delta index passing rates with standard deviation for training, validation, and test data for the rotated bone slab phantom with different thicknesses

	Passing rate (%)		
	Training	Validation	Test
$\delta < 1\%$	88.1 ± 0.7	87.7 ± 0.2	87.2 ± 1.4
$\delta < 3\%$	98.9 ± 0.2	98.8 ± 0.2	98.5 ± 0.4

and thicknesses of this model. The passing rates agree within the statistical uncertainties, hence there is no indication of overfitting and the model seems to generalize well to the unseen bone slab rotation angles and thicknesses of the validation and test data. The passing rates in dependence of the bone slab rotation angle but averaged over the bone thickness is shown in Figure 12b for the training, validation, and test data. While the 3% passing rate is still constantly above 98% for all rotation angles, the 1% passing rate is considerably lower at values between 85% and 90% with a small decrease at higher rotation angles. The model generalizes well to the rotation angles and thicknesses of the validation and test data. In Figure 12c the passing rates are shown in dependence of the bone slab thickness but averaged over the rotation angle. The performance is stable for all bone thicknesses of the training and validation data set. However, the obtained passing rates are slightly lower for the interpolated bone thicknesses of the test data, in particular for thin bones with large rotation angles. These geometries result in dose distributions, which are particularly challenging to predict. Figure 13 shows exem-

plary normalized simulated (gray) and generated (black) energy depositions inside the phantom along the beam using the test data for the in-field (a-c) and out-of-field (d-f) region of the beam with bone slab rotation angles of $\alpha = [0, 40, 85]^\circ$ (a+d, b+e, c+f) and thicknesses of $d = [7, 4, 1.75]$ mm (a+d, b+e, c+f). While the energies are generally well predicted in front of the inserted bone, the relative deviations ΔE_{rel} are larger at the bone slab transition region and behind the bone slab with a small trend to overpredict the deposited energies behind the bone slab in the in-field region. In particular the thinnest bone slab of 1.75-mm thickness with the largest rotation angle of 85° results in larger deviations at the bone slab of up to 10% but represents also the most challenging case of the interpolation. In general, the model shows a good ability to interpolate to unknown geometries.

3.3 | Simple pediatric head phantom with translation

During the training of the model using the data of the simple pediatric head phantom, the highest 1% passing rate on the validation data is achieved at 6900 training epochs as shown in Figure 14a. As the dose in the bone of the skull is much higher compared to the dose in the water, which is mimicking the brain tissue, any dose deviations in the brain tissue would have negligible impact on the delta index, which is normalized to the total maximal simulated dose. Therefore, we calculate an additional delta index δ_{brain} that is normalized to the

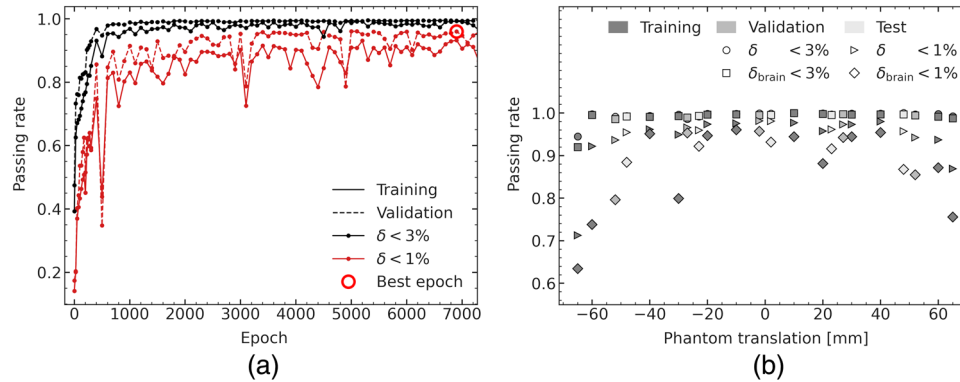


FIGURE 14 (a) Global 1% (red) and 3% (black) passing rates for the training (solid) and validation (dashed) data as a function of training epochs. The best model on the validation data is highlighted with a red circle. (b) Passing rates as a function of the phantom translation for training (dark gray), validation (medium gray), and test (light gray) data

TABLE 3 Delta index passing rates with standard deviation for training, validation, and test data for the simple pediatric head phantom

	Passing rate (%)		
	Training	Validation	Test
$\delta < 1\%$	93 ± 7	96.0 ± 1.7	96.3 ± 0.9
$\delta < 3\%$	99.3 ± 1.5	99.4 ± 0.5	99.6 ± 0.1
$\delta_{\text{brain}} < 1\%$	87 ± 10	90 ± 6	90.4 ± 2.4
$\delta_{\text{brain}} < 3\%$	98.9 ± 2.1	99.2 ± 0.4	99.5 ± 0.2

maximal simulated dose inside the brain and the corresponding passing rates are only taking the voxels of the brain tissue into account. The global passing rates of the best model averaged over the translation of the head are summarized in Table 3 for the training, validation, and test data. The generated doses of more than 96% (99%) of the voxels deviate by less than 1% (3%) of the maximal dose from the simulated dose using the test data if the whole phantom is considered. Although the maximal dose inside the brain is much smaller, the 3% passing rate of the brain is very similar, while the 1% passing rate is reduced to 90%. In Figure 14b the passing rates are shown as a function of the translation of the head for the training, validation, and test data. The 3% passing rate is very stable around 99% for the whole phantom as well as for the brain except for the extreme cases of the maximum translation of $t = \pm 65$ mm. These cases are in particular challenging to predict, because part of the beam traverses only bone and the angle and thickness of the bone change most drastically, which results in steep local gradients of the dose deposition. The 1% passing rate shows a stronger dependence on the head position with a decrease to higher absolute translation distances. The model generalizes well to the unseen translation distances of the test and validation data with minimal differences in performance to the training data. Figure 15 shows exemplary the relative dose deviations as a 2D slices of the head for a

central case of the test data with $t = 2$ mm (a) and the worst performing case of the training data with $t = -65$ mm (b). For the central case (Figure 15a) only few voxels mostly at the first bone and behind show deviations of larger than 1%. For the worst case (Figure 15b), where the beam is passing through the edge of the head, many voxels inside the bone and close to it show deviations of more than 1% and several voxels in particular at thicker bone structures deviate by more than 3% from the simulation. Figure 16 shows the comparison of the normalized simulated and generated energy depositions inside the phantom along the beam for the in-field (a-c) and out-of-field (d-f) region of the beam with head translations of $t = [2, 48, -65]$ mm (a+d, b+e, c+f). The translations of $t = [2, 48]$ mm correspond to test data, while the translation of $t = -65$ mm represents training data but shows the worst performing case. The relative deviations between simulated and generated energy depositions of the test data (a,b,d,e) are mostly within 5% (25%) for the in-field (out-of-field) region and the model shows a good ability to interpolate to unknown phantom geometries. The deviations of the worst performing case (c,f) are only at the transition regions of the bones larger than 5% (25%) in the in-field (out-of-field) region and the energies are generally even for this case well predicted by the GAN.

4 | DISCUSSION

We developed, trained, and evaluated a GAN model for dose predictions of a photon beam inside three variable phantom geometries with increasing complexity.

On the first phantom, a water cube with a bone slab of fixed thickness inserted at various rotation angles, our model achieved on the test data a global 1% and 3% passing rate of $(96.9 \pm 0.7)\%$ and $(99.6 \pm 0.2)\%$, respectively, and demonstrated stable performance over the full range of the varied rotation angles. In the

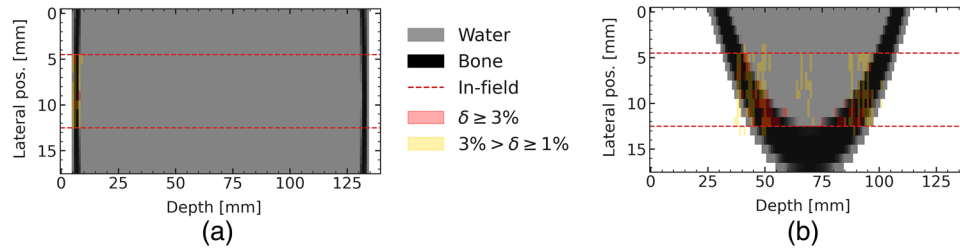


FIGURE 15 2D slices of the phantom with a translation of (a) $t = 2$ mm as part of the test data and (b) $t = 65$ mm as the worst performing case with training data. Voxels with deviations in their dose predictions of larger than 3% (1%) are shown in red (yellow). Whenever the deviation is smaller than 1%, the gray-scale color of the material density matrix is shown. The in-field part of the beam is located between the red dashed lines. Voxels with more than $3\%D_{\max}$ dose prediction deviation (red) and voxels with less than $3\%D_{\max}$ but more than $1\%D_{\max}$ dose deviation (yellow). The gray-scale colors show the densities of the voxels, the red dashed lines indicate the beam edges

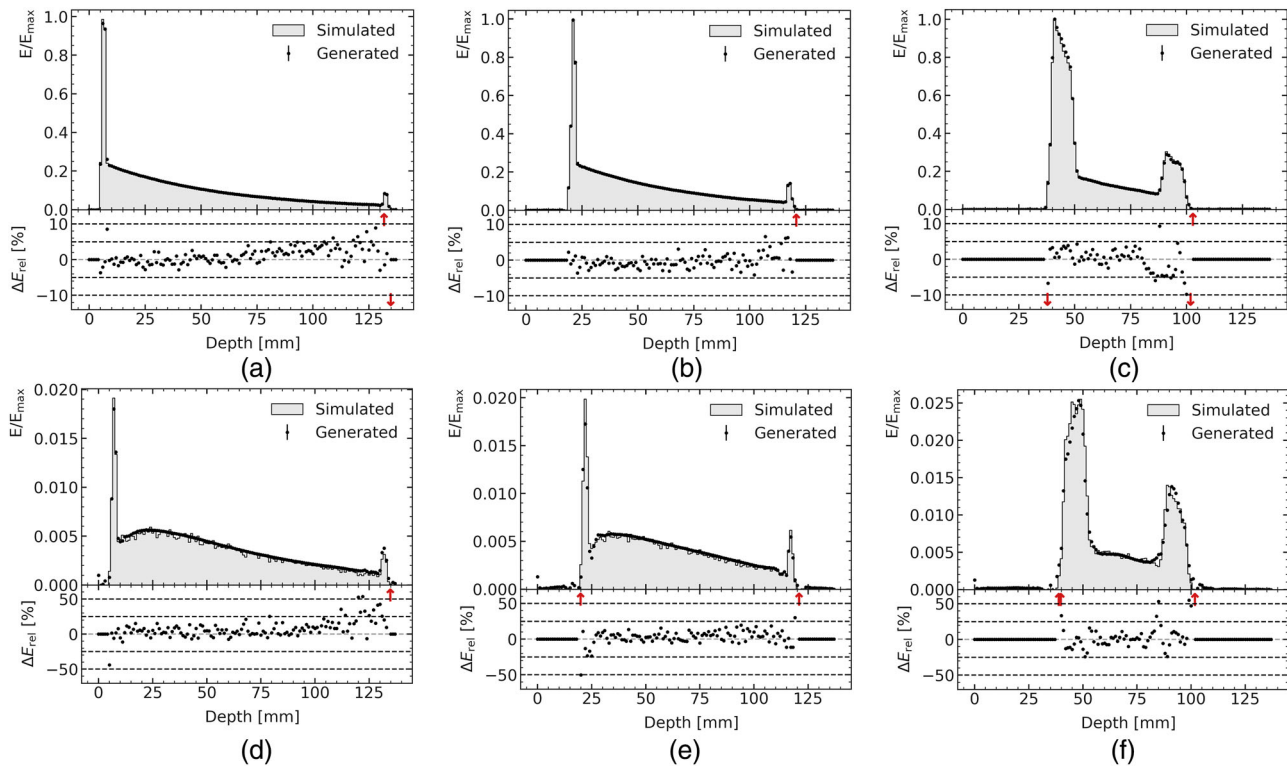


FIGURE 16 Comparisons of normalized simulated and generated energy depositions inside the phantom along the beam using the test data for the in-field (a-c) and out-of-field (d-f) regions of the beam with phantom translations of $t = [2, 48, -65]$ mm (a+d, b+e, c+f). The lower part of the plots shows the relative energy deviation ΔE_{rel} in percent. Outliers are indicated with a red arrow

in-field region of the beam the relative voxel-by-voxel deviations of the predicted energy depositions were below 5%, except for voxels at the interface of water and bone. Out-of-field, the relative deviations were mostly within 25%, however, the absolute deposited energies were also approximately two orders of magnitude smaller and had hence less impact on the loss function during the training. The model was able to interpolate the learnt energy depositions of the training data to energy depositions inside bone slabs with unlearned rotation angles of the validation and test data with little loss in performance.

In the second phantom, the bone thickness of the slab was varied in addition to the rotation angle. This increase of complexity reduced the global 1% and 3% passing rate to $(87.2 \pm 1.4)\%$ and $(98.5 \pm 0.4)\%$, respectively, on the test data, but both rates were relatively stable over the variation of bone slab rotation angle and thickness with small decrease for bone thicknesses that were only part of the test data. The deposited energies were accurately predicted on a voxel-by-voxel basis mostly within 5% (25%) of relative deviation to the simulated energy deposition in the in-field (out-of-field) region of the beam. We found that the model was able to predict

TABLE 4 Execution times of the GAN and the Geant4 MC simulations compared. CPU refers to an Intel Xeon E5-2630 v4 @ 2.2 GHz, GPU to a Nvidia GeForce GTX 1080i

Model	Time per prediction (s)	Rel. speed
Geant4 (1 CPU)	$9.5 \cdot 10^5$	1
GAN (1 CPU)	0.6	$\sim 1.6 \cdot 10^6$
GAN (1 GPU)	0.1	$\sim 9.5 \cdot 10$

the deposited energy with comparable accuracy for geometries (bone slab thickness and rotation angle) that were not part of the training data. This demonstrates that the model is able to generalize the learnt energy deposition prediction to unknown but similar phantom geometries. The model showed a small trend to systematically overestimate the energy depositions behind the bone slab insert, in particular for thicker bones, which indicates that the model did not fully learn the attenuation effect of the bone slab on the beam.

The last phantom, a simple pediatric head, was shifted orthogonal to the beam in order to emulate dynamically changing phantom geometries in multiple dimensions. The obtained 1% and 3% passing rates on the test data of $(96.3 \pm 0.9)\%$ and $(99.6 \pm 0.1)\%$, respectively, were at a high level and stable over the full range of phantom translation except for the extreme case of $t = -65$ mm, for which the beam passes through the edge of the head. If only the voxels of the brain tissue were considered, the 1% passing rate was reduced by roughly 6% and showed stronger dependency on the translation, while the 3% passing rate was comparable to the one of the whole phantom. The relative deviations of the generated energies from the simulations were mostly below 5% (25%) for most voxels of the in-field (out-of-field) region of the beam. Larger deviations were found at the interfaces of the bones. No significant difference was found in the passing rates or in the energy deposition deviations between those obtained from training data and from validation or test data. This establishes that the presented model is able to generalize very well to dynamically changing geometries, which are unknown to the model.

Table 4 shows a comparison of the execution times to obtain the deposited energies inside a phantom either using full Geant4 MC simulation or the GAN model on a CPU (Intel Xeon E5-2630 v4 @ 2.2 GHz) and a GPU (Nvidia GeForce GTX 1080i). While the Geant4 MC simulation takes roughly 264 computing hours, the GAN is able to make the predictions in only 0.6 s on the same CPU, which corresponds to a speed-up factor of roughly $1.6 \cdot 10^6$. Using a GPU instead for the GAN prediction, the deposited energies are obtained in 0.1 s, which is almost 10 million times faster than the Geant4 MC simulation and also significantly faster than the 30 min as reported by a previously devel-

oped method.¹⁰ Compared to the fastest previously published dose computation method,¹¹ the presented model shows significantly better spatial resolution (the fastest previous model does not perform voxel by voxel predictions at all) and more accurate predictions at material interfaces and generally in the case of heterogeneities in the phantom at similar execution speeds. Compared to generalist approximations such as GPUMCD⁴² and the hybrid approaches for MRT,¹⁰ the presented model does not rely on approximations like averaged attenuation coefficients introduced before generating simulation data. As a consequence, any physical process can be taken into account for the model prediction as long as it can be simulated accurately, for example, by MC methods for radiation physics. As the training data are generated using full MC simulations, it can be adapted closely to realistic treatment scenarios. Fast dose prediction approximations, which are currently used in treatment planning programs, have been found to require greater attention and corrections when it comes to commissioning of the system for clinical application (discussed, e.g., in the studies by Montenari et al.⁴³ or Jia et al.⁴⁴). The constructed MC simulation application used in the presented study was validated against experimental measurements in a previous publication²⁹ and it was successfully used for dose predictions in an in vivo experiment with rat irradiations.⁸ These are important research steps, which pose a solid foundation for an eventual, future commissioning of the proposed TPS engine. To fully commission the proposed TPS engine, the pathway would be to commissioning the MC simulations used for the ML training, against dosimetric experimental measurements.

The high accuracy of the predictions, the ability to interpolate and generalize to unknown phantom geometries, and the fast execution times encourage to further develop the model and adapt it for the use in MRT. The model demonstrated flexibility and scalability due to the conditioning on the beam properties and on the geometry of the phantom. At this stage, the presented model is trained only at one fixed resolution. To use different voxel and prediction volume sizes will be subject of the next stage of the project, where we adapt the developed ML algorithm to microbeams.

5 | CONCLUSION

We present a proof-of-concept study to use a 3D-UNet GAN to predict the dose or energy depositions of a photon beam inside a variety of phantoms. The GAN is conditioned on the beam and phantom properties and is trained using a Geant4 MC simulation of the full radiation transport. The presented model is successfully trained on three different phantoms with changing geometry without further need for optimization. The model predicts with adequate accuracy the

energy depositions and deviates by less than 1% of the maximal deposited energy from the simulation in more than 96% of the in-field voxels for the most realistic phantom—a simple pediatric head. The model demonstrates good ability to generalize the predictions to unknown but similar phantom geometries.

Our approach differs from previously published studies on dose estimations with machine learning methods by relying purely on MC simulations for radiation physics as training data. This allows the development of a more general approach that can be deployed to novel treatment methods, for which there are no large number of available patient CT scans or existing treatment plans already, such as mini⁵ or MRT.

For an application in treatment planning systems, dose simulations with satisfactory statistical accuracy using MC techniques are usually computationally too demanding, however, approximations of the physics processes translate to large dose calculation inaccuracies in heterogeneous phantoms, especially at material interfaces. In the adopted dosimetric calculation scenario, the presented GAN model is able to achieve almost MC simulation precision in only a fraction of a second for a variety of radiation scenarios, which makes it an ideal candidate for future treatment planning systems. Encouraged by these results, future studies will evaluate the presented machine learning approach using simulations with more realistic patient geometries, deriving, for example, from the ICRP110 phantom,⁴⁵ digital phantoms, which allow customization via parameterization,^{46,47} and finally patient data for the application in treatment planning.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the computing time provided on the Linux HPC cluster at TU Dortmund University (LiDO3), partially funded in the course of the Large-Scale Equipment Initiative by the German Research Foundation (DFG) as project 271512359.

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

REFERENCES

- Brahme A, Roos JE, Lax I. Solution of an integral equation encountered in rotation therapy. *Phys Med Biol*. 1982;27(10):1221-1229.
- Otto K. Volumetric modulated arc therapy: IMRT in a single gantry arc. *Med Phys*. 2008;35(1):310-317.
- Bucci MK, Bevan A, Roach M. Advances in radiation therapy: conventional to 3D, to IMRT, to 4D, and beyond. *CA Cancer J Clin*. 2005;55(2):117-134.
- Jani AB, Su A, Milano MT. Intensity-modulated versus conventional pelvic radiotherapy for prostate cancer: analysis of acute toxicity. *Urology*. 2006;67(1):147-151.
- Deman P, Vautrin M, Edouard M, et al. Monochromatic minibeam radiotherapy: from healthy tissue-sparing effect studies toward first experimental glioma bearing rats therapy. *Int J Radiat Oncol Biol Phys*. 2012;82(4):e693-700.
- Von Neubeck C, Seidlitz A, Kitzler HH, Beuthien-Baumann B, Krause M. Glioblastoma multiforme: emerging treatments and stratification markers beyond new drugs. *Br J Radiol*. 2015;88(1053):20150354.
- Bräuer-Krisch E, Adam JF, Alagoz E, et al. Medical physics aspects of the synchrotron radiation therapies: microbeam radiation therapy (MRT) and synchrotron stereotactic radiotherapy (SSRT). *Physica Med*. 2015;31(6):568-583.
- Engels E, Li N, Davis J, et al. Toward personalized synchrotron microbeam radiation therapy. *Sci Rep*. 2020;10(1):1-13.
- Grotzer MA, Schültke E, Bräuer-Krisch E, Laissue JA. Microbeam radiation therapy: clinical perspectives. *Physica Med*. 2015;31(6):564-567.
- Donzelli M, Bräuer-Krisch E, Oelfke U, Wilkens JJ, Bartzsch S. Hybrid dose calculation: a dose calculation algorithm for microbeam radiation therapy. *Phys Med Biol*. 2018;63(4):45013.
- Poole CM, Day LRJ, Rogers PAW, Crosbie JC. Synchrotron microbeam radiotherapy in a commercially available treatment planning system. *Biomed Phys Eng Express*. 2017;3(2):025001.
- Debus C, Oelfke U, Bartzsch S. A point kernel algorithm for microbeam radiation therapy. *Phys Med Biol*. 2017;62(21):8341-8359.
- Wang M, Zhang Q, Lam S, Cai J, Yang R. A review on application of deep learning algorithms in external beam radiotherapy automated treatment planning. *Front Oncol*. 2020;10:580919.
- Shiraishi S, Moore KL. Knowledge-based prediction of three-dimensional dose distributions for external beam radiotherapy. *Med Phys*. 2016;43(1):378-387.
- Peng Z, Shan H, Liu T, Pei X, Zhou J, Wang G & Xu XG. Deep learning for accelerating Monte Carlo radiation transport simulation in intensity-modulated radiation therapy. arxiv: Medical Physics, 2019.
- Barragán-Montero AM, Nguyen D, Lu W, et al. Three-dimensional dose prediction for lung IMRT patients with deep neural networks: robust learning from heterogeneous beam configurations. *Med Phys*. 2019;46(8):3679-3691.
- Kontaxis C, Bol GH, Lagendijk JJ, Raaymakers BW. DeepDose: towards a fast dose calculation engine for radiation therapy using deep learning. *Phys Med Biol*. 2020;65(7):075013.
- Agostinelli S, Allison J, Amako K, et al. GEANT4—a simulation toolkit. *Nucl Instrum Methods Phys Res A*. 2003;506(3):250-303.
- Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Adv Neural Inf Process Syst*. 2014;3:2672-2680.
- Lan L, You L, Zhang Z, et al. Generative adversarial networks and its applications in biomedical informatics. *Front Public Health*. 2020;8:164.
- Sarrut D, Krahn N, L'etang JM. Generative adversarial networks (GAN) for compact beam source modelling in Monte Carlo simulations. *Phys Med Biol*. 2019;64(21):215004.
- Kearney V, Chan JW, Wang T, et al. DoseGAN: a generative adversarial network for synthetic dose prediction using attention-gated discrimination and generation. *Sci Rep*. 2020;10(1):11073.
- 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(5):2646-2660.
- Allison J, Amako K, Apostolakis J, et al. Recent developments in Geant4. *Nucl Instrum Methods Phys Res A*. 2016;835:186-225.
- Cosmo G. Software Process in Geant4 <https://www.researchgate.net/publication/41588801> Software Process in Geant4 2001. Accessed: February 28, 2022
- Geant4 Collaboration. Physics Reference Manual Documentation. <https://geant4-userdoc.web.cern.ch/UsersGuides/>

- [PhysicsReferenceManual/html/index.html](#). Accessed May 18, 2021.
27. Stevenson AW, Crosbie JC, Hall CJ, Häussermann D, Livingstone J, Lye JE. Quantitative characterization of the X-ray beam at the Australian Synchrotron Imaging and Medical Beamline (IMBL). *J Synchrotron Radiat*. 2017;24:110-141, International Union of Crystallography.
 28. Spiga J, Prezado Y, Bräuer-Krisch E, Fanti V, Randaccio P & Bravin A. The effect of beam polarization in Microbeam Radiation Therapy (MRT): Monte Carlo simulations using Geant4. IEEE Nuclear Science Symposium Conference Record, 2009;2170-2173. Orlando, Florida, USA.
 29. Dipuglia A, Cameron M, Davis JA, et al. Validation of a Monte Carlo simulation for microbeam radiation therapy on the imaging and medical beamline at the Australian Synchrotron. *Sci Rep*. 2019;9(1):1-14.
 30. Mirza M & Osindero S. Conditional Generative Adversarial Nets. arXiv: Machine Learning. 2014.
 31. Arjovsky M, Chintala S & Bottou L. Wasserstein GAN, arXiv: Machine Learning, 2017
 32. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A. Improved training of Wasserstein GANs. In: von Luxburg Ulrike, Guyon Isabelle, Bengio Samy, Wallach Hanna, Fergus Rob, *Advances in Neural Information Processing Systems*, Red Hook, NY, USA: Curran Associates Inc.; 2017:5768-5778.
 33. Kingma DP & Ba JL. Adam: a method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, International Conference on Learning Representations, ICLR 2015. San Diego, CA, USA.
 34. Ramachandran P, Zoph B & Le QV. Searching for activation functions. arXiv: Neural and Evolutionary Computing, 2017
 35. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. 32nd International Conference on Machine Learning, ICML 2015;1:448-456, International Machine Learning Society (IMLS) 2015.
 36. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15:1929-1958.
 37. Ronneberger O, Fischer P & Brox T. U-net: Convolutional Networks for Biomedical Image Segmentation. arXiv: Computer Vision and Pattern Recognition, 2015.
 38. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T & Ronneberger O. 3D U-net: Learning dense volumetric segmentation from sparse annotation. In Lecture Notes in Computer Science (including sub-series Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Editors: Ourselin S, Wells WS, Sabuncu MR, Unal G, Joskowicz L; 9901 LNCS, 2016;424-432. Springer,
 39. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys*. 1998;25(5):656-661.
 40. Chollet F and others. Keras <https://keras.io> 2015. Accessed: February 28, 2022
 41. Abadi M, Agarwal A & Barham P et al. TensorFlow: large-scale machine learning on heterogeneous systems. 2015. Software available from tensorflow.org. Accessed: February 28, 2022.
 42. Hissouiny S, Ozell B, Bouchard H, Després P. GPUMCD: a new GPU-oriented Monte Carlo dose calculation platform. *Med Phys*. 2011;38(2):754-764.
 43. Montanari D, Scolari E, Silvestri C, et al. Comprehensive evaluations of cone-beam CT dose in image-guided radiation therapy via GPU-based Monte Carlo simulations. *Phys Med Biol*. 2014;59(5):1239-1253.
 44. Jia X, Yan H, Gu X, Jiang SB. Fast Monte Carlo simulation for patient-specific CT/CBCT imaging dose calculation. *Phys Med Biol*. 2012;57(3):577-590.
 45. HGM CC, P D. ICRP Publication 110. Realistic reference phantoms: an ICRP/ICRU joint effort. A report of adult reference computational phantoms. *Ann ICRP*. 2009;39(2):3-5.
 46. Arce P, Bolst D, Bordage MC, et al. Report on G4-Med, a Geant4 benchmarking system for medical physics applications developed by the Geant4 Medical Simulation Benchmarking Group. *Med Phys*. 2021;48(1):19-56.
 47. Large MJ, Malaroda A, Petasecca M, Rosenfeld AB, Guatelli S. Modelling ICRP110 adult reference voxel phantoms for dosimetric applications: development of a new Geant4 Advanced Example. *J Phys Conf Ser*. 2020;1662:012021, IOP Publishing Ltd 2020.

How to cite this article: Mentzel F, Kröniger K, Lerch M, et al. Fast and accurate dose predictions for novel radiotherapy treatments in heterogeneous phantoms using conditional 3D-UNet generative adversarial networks. *Med Phys*. 2022;49:3389–3404. <https://doi.org/10.1002/mp.15555>