Iranian Journal of Medical Physics

ijmp.mums.ac.ir

DL-Dose Framework: Introducing a Fast Dose Calculation Engine for Radiotherapy-Based Deep Learning

Naser Mahdavi [1], Mojtaba Shamsaei Zafarghandi [1*], Saeed Setayeshi [1], Hassan Ali Nedaie[2]

1. Department of Physics and Energy Engineering, Amir Kabir University of Technology, Iran
2. Radiation Oncology Research Centre, Cancer Institute, Tehran University of Medical Sciences, Tehran, Iran

ARTICLE INFO

Article type:
Original Paper

Article history:
Received: Jan 02, 2025
Accepted: Jul. 04, 2025

Keywords:
Dose Prediction
Deep Learning
Heterogeneity
Radiotherapy
Monte Carlo Simulation

ABSTRACT

Introduction: This study aimed to develop a dose prediction framework based on deep learning that utilizes water dose distribution and the characteristics of heterogeneous phantoms as inputs.

Material and Methods: A dataset of two hundred heterogeneous phantoms with identical geometry and variable bone and lung layer thicknesses was generated using the DOSXYZnrc Monte Carlo (MC) code. The thickness of each inhomogeneity was randomly assigned between 1 and 5 cm at different positions along the z-axis of the phantom. A deep learning–based dose prediction framework was then developed to estimate three-dimensional dose distributions. The model used five input channels, including water dose distribution, mass density, CT number, voxel distance from the radiation field center, and a binary radiation field mask. The network's output was the predicted dose distribution for each heterogeneous phantom.

Results: The accuracy of the predicted results by the DL-Dose Framework was evaluated against those obtained through the Monte Carlo method, using the delta index in heterogeneous phantoms. In the water medium, before encountering heterogeneities, 100% of the dose distribution for voxels with deviations of less than 1% from the maximum dose, consistent with the results measured by the MC method. Furthermore, 94.2% of the dose distribution for voxels in lung heterogeneity areas and 98.1% for voxels in bone heterogeneity regions were comparable to the MC method results, with deviations of less than 1% from the maximum dose.

Conclusion: The developed DL-Dose Framework can accurately predict dose distribution in heterogeneous phantoms.

► *Please cite this article as:*
Mahdavi N, Shamsaei Zafarghandi M, Setayeshi S, Nedaie HA. DL-Dose Framework: Introducing a Fast Dose Calculation Engine for Radiotherapy-Based Deep Learning. Iran J Med Phys 2026; 23 (1): 26-37. 10.22038/ijmp.2025.84901.2493.

Introduction

Deep learning, which is defined as the application of multi-layer neural network architectures, has recently gained substantial prominence across various scientific fields [1-2]. In radiation therapy, both machine learning and deep learning techniques, particularly convolutional neural networks, have been increasingly utilized for treatment planning tasks [3]. Multiple studies have demonstrated that deep learning models can predict complete three-dimensional dose distributions. In these approaches, the model takes patient-specific and treatment-related information, such as prescription dose, target, and organs at risk (OAR) contours, and computed tomography (CT) images as input, and produces the corresponding predicted dose distribution as output [4].

The knowledge-based planning (KBP) paradigm, which has been the focus of much research on dose prediction in radiotherapy, uses data from previous patient plans to predict a patient's dose volume histogram (DVH) and dose constraints[5-8]. While KBP has seen significant advancements, it requires a considerable amount of time and effort to select specific elements, such as spatial data on organs at risk (OAR) and planning target volumes (PTV)[6-10]. Furthermore, it is limited by the inherent information present in the data [6]. Another challenge in dose prediction using deep learning methods is the loss of spatial consistency regarding neighboring anatomy, which can exhibit significant differences in CT numbers and steep gradients of tissue heterogeneity.

Additionally, the critical parameters of the radiation field and heterogeneous phantoms, which are effective in calculating the dose in heterogeneous environments, theoretically or through the Monte Carlo method, should be evaluated in dose prediction. The existing approaches, such as knowledge-based planning systems, do not incorporate all the primary physical parameters of heterogeneous environments and radiation fields, which can influence dose calculations.

Therefore, using physical parameters to estimate the heterogeneous environment dose distribution, independent of factors such as the contour of the organs at risk (OAR) or planning target volume, can help develop dose distribution prediction systems in

*Corresponding Author: Tel: +98-64545200; Fax: +98-66495432; Email: pysham@aut.ac.ir

radiotherapy. It is possible to create a system that accurately predicts dose distribution in heterogeneous environments with results comparable to those obtained using Monte Carlo methods, by these critical parameters in the design of dose distribution prediction systems.

In this study, we introduce the DL-Dose Framework, based on deep learning, for accurate and rapid dose computation in radiotherapy. Our model incorporated influence parameters related to heterogeneous environments and the radiation field to predict dose distribution. The primary objective was to create a deep learning-based model that can predict dose distribution faster than the Monte Carlo approach.

We simulated heterogeneous phantoms containing voxels filled with specific materials, including water, lung, and bone, at defined depths using the DOSXYZnrc code. The DL-Dose Framework predicted the percentage depth dose (PDD) distribution and dose maps for heterogeneous phantoms. We assessed the performance of the developed framework on a test dataset by comparing the PDD curves and 2D dose maps with the corresponding Monte Carlo dose distributions.

Materials and Methods

Beam Model

Initially, the water phantom was simulated using the DOSXYZnrc code with dimensions of (16cm × 16 cm × 32cm), subdivided into 32 × 32 × 64 voxels with a uniform size of 5 mm in the X, Y, and Z directions, respectively. In this study, we utilized a rectangular simulated source, as it allowed us to easily account for both the primary dose and scattered dose components when predicting dose distribution. The DOSXYZnrc code manual lists eleven different types of sources, and the rectangular beam characterization model (Isource=0) was selected from those sources. Therefore, X1 = 2.5 cm, X2 = 2.5 cm, Y1 = 2.5 cm, and Y2 = 2.5 cm on the phantom surface define the parallel rectangular beam field (isource = 0). Beam directions were specified relative to the positive x-axis, the positive y-axis, and the negative z-axis [11].

The primary source used in the calculations was the 6 MV photon beam spectrum, taken from the Mohan6 spectrum file in the EGS4 Spectra Library[12]. Monte Carlo calculations were performed using the Simorgh supercomputer at Amir Kabir University. Each simulation lasted approximately ten hours and was executed with 1.6 × 10^9 histories to achieve an overall statistical uncertainty of less than 0.3%.

Heterogeneous Phantom Modeling and Dataset Generation

All data used in this study were generated via a Monte Carlo simulation, and no data related to humans or animals were used. Heterogeneous phantoms with similar dimensions of 8 cm × 8 cm × 16 cm and 32 × 32 × 64 voxels in the x, y, and z directions were simulated using the DOSXYZnrc code. All voxels at specific depths in the phantoms were filled with different materials, such as water, lung, and bone, with densities ρ = 1000 kg/m³, ρ = 260 kg/m³, and ρ = 1850 kg/m³, respectively, to model heterogeneous phantoms.

The simulated square source with a 6 MV photon beam spectrum irradiated each heterogeneous phantom. Then, the Simorgh supercomputer, placed at AmirKabir University, was utilized to calculate the dose distribution in each simulated heterogeneous phantom. The specifications of the Windows cloud computing used to run simulations included 12 virtual processors with a 3GHz Processor base frequency, 48GB of Main memory capacity, and 200 GB of Hard disk capacity. The dose distributions for each heterogeneous phantom were calculated using the DOSXYZnrc code and then transformed into a matrix with dimensions of 32 × 32 × 64 in the x, y, and z directions.

As we know, the DOSXYZnrc code provides a dose distribution for a single photon. Based on the energy spectrum of 6 MV used in the simulations, we can estimate that 0.01 Gy would be delivered at the maximum depth of the water phantom with one monitor unit (MU). The calculated dose distribution in the water phantom should be multiplied by 1e15 to achieve a dose of 0.01Gy at the maximum depth. Therefore, the dose distribution results will be multiplied by 1e15. It is important to note that this adjustment was applied to both the input dose distributions for the water phantom and the output dose distribution for the heterogeneous phantoms. This factor of 1e15 can be used with other coefficients and will not affect the accuracy of the predicted dose results.

Network Architecture

In recent years, artificial intelligence (AI) and deep learning methods have made extraordinary developments, especially in computer vision and decision-making. In 2015, Ronneberger et al. introduced the U-Net architecture, a deep-learning approach for semantic segmentation[13]. This architecture effectively combines local and global features to make accurate pixel-wise predictions. However, it is worth noting that the computational cost increases as the dimensionality of the U-Net architecture increases, especially when transitioning from 2D to 3D. The increasing computational cost is evident in 3D dose distribution prediction. Therefore, future versions of the U-Net architecture should focus on reducing the number of filters in each convolutional or max pooling layer. One approach to addressing this issue in 3D U-Net architectures involves incorporating skip connections within each block, akin to those utilized in ResNet, which are more effective at propagating information[14].

Huang et al. published a Densely Connected Convolutional Neural Network, commonly known as DenseNet, in 2017 [15]. They introduced the innovative concept of densely connecting convolutional layers to enhance feature propagation and reuse, mitigate the

vanishing gradient problem, and reduce the required trainable parameters. Although this approach necessitates more memory, the authors demonstrated that DenseNet can achieve superior performance while utilizing significantly fewer parameters than traditional neural network architectures. Therefore, it can be concluded that DenseNet is considerably more efficient in feature computation than existing network architectures.

Although DenseNet employs dense inter-layer connections that enhance information flow, feature reuse, and gradient propagation, the absence of an encoder–decoder structure with pooling and subsequent upsampling—such as that used in U-Net—reduces its ability to capture multi-scale contextual features and broader global spatial information. To address this limitation, our work aims to combine the strong global-local representation capability of U-Net with the efficient feature propagation and reuse mechanism of DenseNet, thereby improving performance for three-dimensional dose prediction tasks. Building on this concept, Dan Nguyen and colleagues previously proposed a neural network framework that integrates the fundamental design principles of both architectures. Inspired by this approach, the architecture used in this study is referred to as the Hierarchically Densely Connected U-Net [15-16].

The standard Hierarchically Dense (HD) U-Net was utilized in this study to develop the DL-Dose Framework for dose prediction in heterogeneous phantoms [15-16]. The Structurally Hierarchically Dense (HD) U-Net is a hybrid model that combines features of DenseNet and U-Net. It employs densely connected layers at each hierarchical level of the U-Net architecture. Furthermore, the U-Net structure incorporates pooling and upsampling techniques, which facilitate the extraction of global features from the input data. The Dense-Unet architecture uses three procedures: dense convolutional layers, dense downsampling, and U-net upsampling. The dense convolutional layer applies a standard convolution with ReLU, followed by a concatenation of the previous feature set.

The dense downsampling operation was performed by a convolution of (3×3×3) with strides of two and ReLU activation to generate a new feature map with half the resolution. On the previous feature maps, a max-pooling operation (2 × 2 × 2 kernel size; voxel stride = 2) was implemented to reduce the size of feature maps. After that, the previous feature was concatenated with the new feature set.

The U-Net upsampling operation was performed by upsampling with a size of (2 ×2 ×2) and using strides of two, followed by (3×3×3) convolutions and ReLU activation. Finally, it is concatenated with the feature set on the other side of the "U" architecture. A growth rate for each dense operation can be determined by calculating the number of new features during the convolution step. The single output channel was constructed by convolving (1 ×1 ×1) with a linear activation function in the final layer.

The Dense-Unet architecture used to design the DL-Dose Framework included five multiscale hierarchical levels of 3D dense convolutional layers. Finally, the DL-Dose Framework had five three-dimensional input channels and one three-dimensional output channel for dose prediction with dimensions of 32 × 32 × 64 in the x, y, and z directions, respectively.

Figure 1 shows an illustration of the DL-Dose Framework for dose prediction. The number above the boxes indicates the quantity of concatenated features, while those to the left indicate the size of the three-dimensional feature maps.

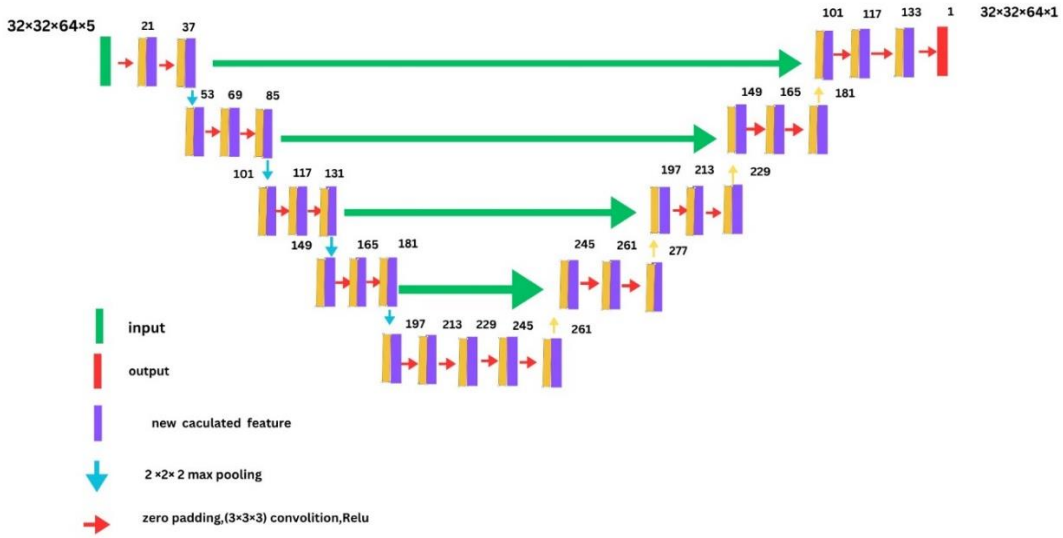

Figure 1. The image of the DL-Dose Framework for dose prediction is presented. The numbers above the boxes indicate the number of concatenated features.

Input-Output

According to the goal of this study, five input channels were defined for the DL-Dose Framework. The dose distribution of the water phantom was considered the first input channel of the designed architecture. The selection of additional input channels in the architecture was based on physical parameters that influence dose calculation in a heterogeneous phantom.

Based on the American Association of Physicists in Medicine Report No. 85, the energy released by the cluster of photon rays in the tissue occurs in two stages. Initially, the photons interact in the medium, transferring kinetic energy to charged particles (TERMA), then the charged particles deposit their energy through ionization and excitation events along a finite track. In general, the quantity of TERMA is the product of the mass attenuation coefficient and energy fluence (ψ) at a point r (x, y, z) in a heterogeneous phantom. The energy fluence (ψ) is proportional to the inverse of the distance of each voxel from the central point of the radiation field. Therefore, the matrix containing the distance of each voxel from the reference point at the center of the radiation field was chosen as the second input channel. Also, due to the impact of the mass density of each voxel on estimating the TERMA in heterogeneous phantoms, the matrix of mass density of heterogeneous phantoms was selected as an input for the DL-Dose Framework. The CT numbers of each voxel in the heterogeneous phantom were used to correct inhomogeneity. Therefore, the matrix of the CT numbers of the heterogeneous phantoms was used as the fourth input for the DL-Dose Framework[17].

Additionally, the American Association of Physicists in Medicine Report No. 85 describes the total dose in a heterogeneous phantom as the sum of primary and scattered components. The primary dose is attributed to the primary particles, while the scattered dose is attributed to secondary particles. In this research, the primary and scattered dose components were produced separately due to the use of rectangular sources. The simulated rectangular source allowed for the fifth input of the Dl dose framework to be defined based on the direct and indirect paths of the primary radiation. In the 3D fifth input, a value of one was assigned to voxels in the direct path of the primary rectangular radiation field, and a value of zero was assigned to voxels outside the rectangular radiation fields. All matrices for input and output channels were constructed with dimensions of 32x32x64 in the x, y, and z directions, respectively.

Training

The DL-Dose Framework was implemented using Keras libraries in the Google Colab execution platform. Google Colaboratory, or Google Colab, is a free, cloud-based machine learning platform. The Google Colab service provides 358.27 GB of hard disk space and 12.72 GB of RAM in a single runtime[18].

The `train_test_split` function from the `sklearn.model_selection` package in Python was utilized to randomly divide 200 phantom datasets into three groups: training, validation, and testing data for the DL-Dose Framework [19]. Specifically, 120 heterogeneous phantoms were allocated for training, 60 for validation, and 20 for testing. The input and output matrices had dimensions of (32×32×64) with five channels and (32×32×64) with one channel, respectively.

The mean squared error (MSE) was used as the loss function for a batch size consisting of four patches. The ADAM optimizer, a stochastic gradient suitable for problems with large amounts of data or parameters in the Keras library, was utilized to enhance the training of the deep learning network. The optimizer accepted the learning rate, beta1, and beta2 as variable input values, set to 1e-4, 0.9, and 0.999, respectively[19].

Early stopping, a regularization strategy, was employed to reduce overfitting once the performance of the developed deep learning architecture on the validation dataset began to degrade. The model was evaluated on an independent validation dataset after each epoch, rather than predefining a fixed number of training epochs. Because of this evaluation, the training process was terminated when no further improvement was observed. In this study, early stopping was therefore applied to control model generalization and avoid excessive fitting to the training data. Furthermore, the validation loss was monitored to identify and retain the set of model parameters corresponding to the optimal performance, a commonly adopted procedure for selecting the best-performing weights during training [20-21].

Statistical Data Analysis

The deviation for the dose values predicted with the DL-Dose Framework and calculated by the Monte Carlo method was calculated using Equation (1). In Equation (1), $D_{z,MC}$ represent the reference dose at a depth of Z, and $D_{z,(DL)}$ represent the dose predicted by the DL-Dose Framework at a depth of Z.

$$D(\%) = 100 * \frac{(D_{z(DL)} - D_{z,MC})}{D_{z,MC}} \quad (1)$$

Additionally, we utilized the delta index, inspired by the global gamma index, to calculate a performance metric for the prediction quality of the DL-Dose Framework[22-23]. The Delta index is defined as in Equation (2), where D_{DL} represents the dose predicted by the DL-Dose Framework, D_{sim} represents the dose calculated by the DOSXYZnrc simulation code, and D_{sim}^{max} represents the maximum dose deposition in the phantom as calculated in the MC simulation.

$$\delta = \frac{D_{DL} - D_{sim}}{D_{sim}^{max}} \quad (2)$$

Also, we used passing rates of 1%, 2%, and 3% for the Delta index to assess the accuracy of the DL-Dose Framework. The percentage of voxels at various depths in heterogeneous phantoms with a Delta index below 1%, 2%, and 3% was also documented.

Results

Figure 2 illustrates the training and validation losses per epoch for the DL-Dose Framework. Generally, both the training and validation losses decrease as the number of epochs increases. As depicted in Figure 2, the mean squared errors for both training loss and validation loss converged toward zero after approximately forty epochs.

PDD curves and dose distributions were plotted using the Matplotlib library in Python based on the provided test data. The following section describes PDD curves and two-dimensional dose distributions for two examples of the test data.

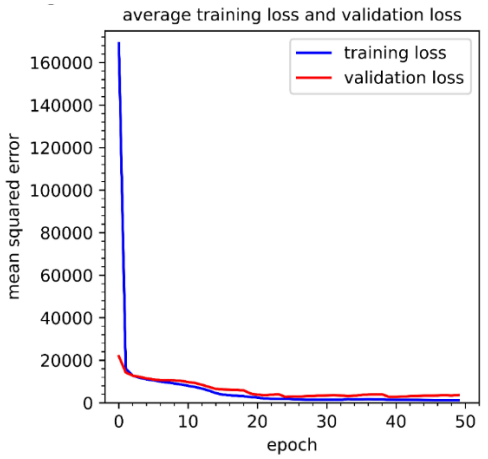

Figure 2. The training and validation losses per epoch for the DL-Dose Framework.

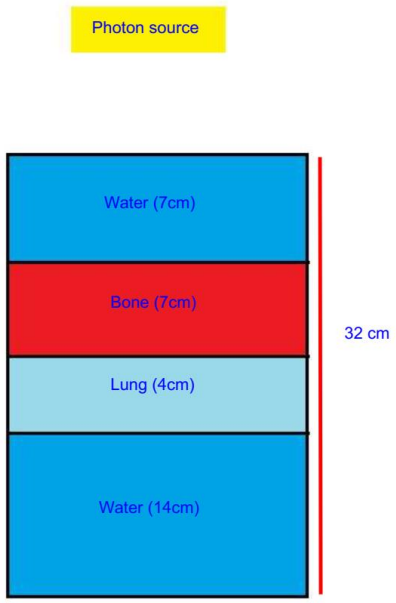

Figure 3. The illustration depicts the first Heterogeneous Phantom

PDD Curve and Dose Profiles for the First Heterogeneous Phantom

Figure 3 illustrates the first heterogeneous phantom used to study PDD curves and dose distributions, which incorporates lung and bone heterogeneities. This heterogeneous phantom includes four distinct layers. The first layer is composed of water, measuring 7 cm in thickness, while the second layer is composed of bone, also 7 cm thick. The third layer is composed of lung tissue, with a thickness of 4 cm, and the fourth layer is composed of water, measuring 14 cm in thickness.

Figure 4A displays PDD curves calculated using Monte Carlo simulation and predicted by the DL-Dose Framework for the first heterogeneous phantom. In Figure 4B, a bar chart compares the dose values obtained from the Monte Carlo method with those predicted by the DL-Dose Framework. The red bars in Figure 4B highlight the differences in dose values between the two methods for all voxels along the central beam axis. Additionally, Figure 4C presents a bar chart displaying the percentage dose deviation. The results show that the deviation generally remains within ±2%, with a slight increase in the water environment due to the presence of heterogeneities. The maximum dose calculated by the Monte Carlo method was 74.12 Gy, while the DL-Dose Framework predicted a maximum dose of 74.33 Gy. The deviation in maximum dose values between the two methods was only 0.28% in this instance.

Figures 5A and 5B display two-dimensional (x, z) images of the dose distribution and isodose curves obtained using the Monte Carlo and DL-Dose Framework methods. Additionally, Figure 5C illustrates the difference between the dose distribution calculated by the Monte Carlo method and that predicted by the DL-Dose Framework. Isodose curves corresponding to 90%, 80%, 70%, 60%, 50%, and 30% of the maximum dose are shown in Figure 5. Furthermore, Figure 5D compares the isodose contours of the dose calculated using the Monte Carlo method with the dose predicted by the DL-Dose Framework. The isodose curves demonstrate good agreement in the water region before heterogeneities and in the bone area. In the lung heterogeneity region, the comparison of isodose curves and dose distribution differences shows a relatively good match.

Figure 6 shows dose profiles for the first heterogeneous phantom at various depths calculated using the Monte Carlo method (red) and predicted by the DL-Dose Framework (blue). Figures 6A and 6B depict dose profiles at a depth of 7 cm at the water-bone interface and 10 cm within the bone layer, respectively. Also, Figures 6C and 6D show dose profiles at a depth of 14 cm at the bone-lung interface and 16.5 cm in the lung heterogeneity layer. Finally, Figures 6E and 6F show dose profiles at 18 cm at the lung-water interface and 25 cm within the fourth water layer.

The fraction of voxels with a delta index smaller than 0.01, 0.02, and 0.03 was determined using Equation 2 at various depths in the first heterogeneous phantom. Table 1 lists and summarizes the percentage of voxels with a delta index smaller than 0.01, 0.02, and 0.03 for the first heterogeneous phantom.

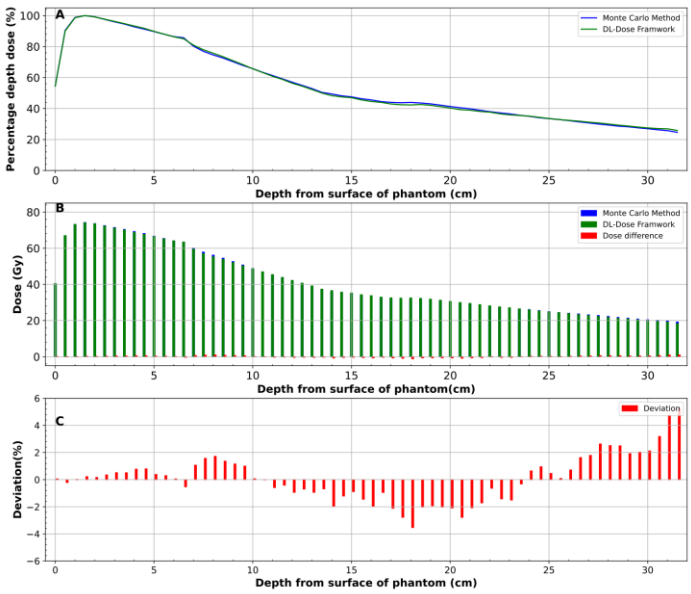

Figure 4. The illustration presents PDD curves (A), the dose values for voxels along the central beam axis (B), and the deviation (C) for the first heterogeneous phantom.

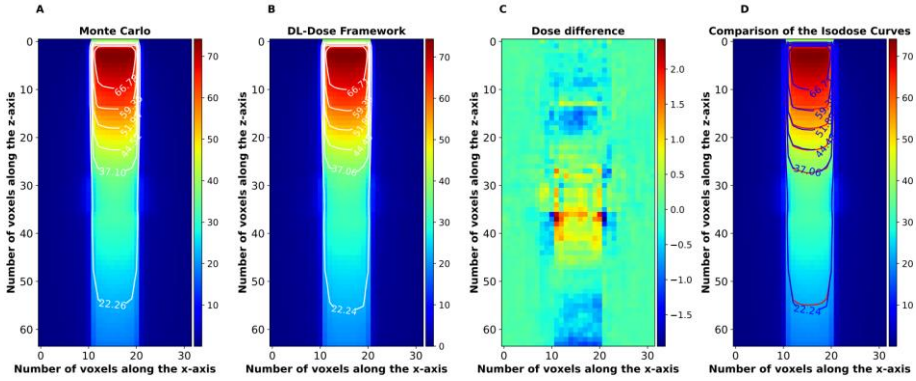

Figure 5. The illustration presents the two-dimensional (x, z) images of the dose distribution calculated by the Monte Carlo method (A), predicted by the Dl-Dose Framework (B), the dose difference between the two methods (C), and a comparison of the isodose curves (D) for the first heterogeneous phantom.

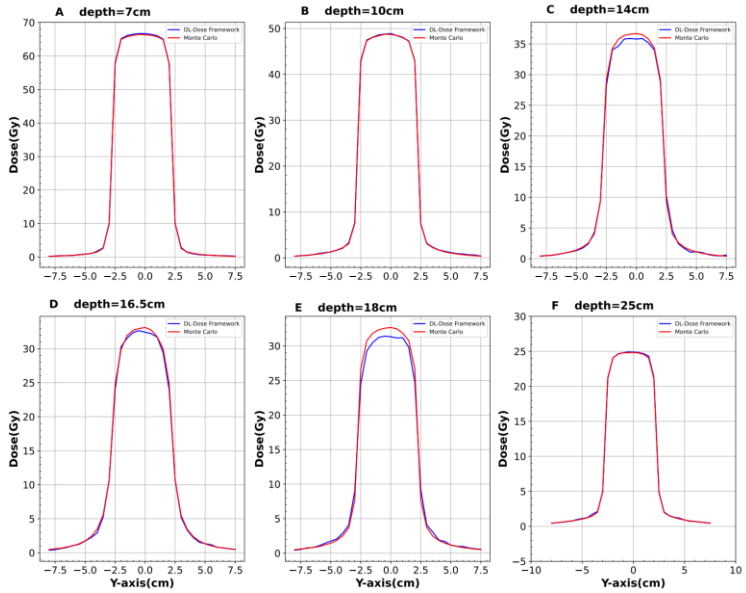

Figure 6. The illustration depicts dose profiles for the first heterogeneous phantom at various depths, calculated using the Monte Carlo method (red) and predicted by the DL-dose Framework (blue).

Table 1. The percentage of voxels with a delta index smaller than 0.01, 0.02, and 0.03 for the first heterogeneous phantom.

First heterogeneous phantom	depth (cm)	fraction of the voxels in percentage with a delta index smaller than $\delta < 0.01$	fraction of the voxels in percentage with a delta index smaller than $\delta < 0.02$	fraction of the voxels in percentage with a delta index smaller than $\delta < 0.03$
In the water medium before heterogeneity	5	100	100	100
Interface of water and bone	7	99.6	100	100
Within the bone heterogeneity	10	100	100	100
Interface of bone and lung	14	95.2	99.9	100
Within the lung heterogeneity	16.5	98.1	100	100
The interface of the lung and water	18	87.9	94.7	98.8
In the water medium, beneath the heterogeneity	25	100	100	100

PDD Curve and Dose Profiles for The Second Heterogeneous Phantom

The second heterogeneous phantom shown in Figure 7 was also used to study PDD curves and dose distribution maps, incorporating lung and bone heterogeneities. This heterogeneous phantom is composed of four distinct layers. The first layer is water, with a thickness of 10.5 cm; the second layer is bone, with a thickness of 6 cm; the third layer is lung, with a thickness of 5 cm; and the fourth layer is again water, with a thickness of 10.5 cm.

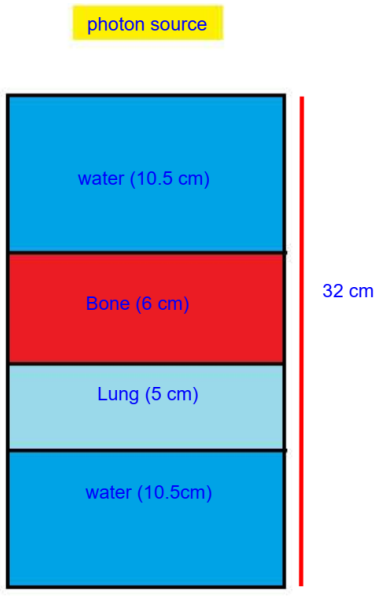

Figure 7. The illustration depicts the Second Heterogeneous Phantom

Figure 8A depicts PDD curves calculated using the Monte Carlo method and predicted by the DL-Dose Framework for the second heterogeneous phantom. In addition, Figure 8B presents a bar chart comparing the dose values obtained by the two methods. The differences in dose values between the two methods are illustrated by the red bars in Figure 8B for all voxels along the central beam axis. Furthermore, Figure 8C displays the percentage dose deviation on a bar chart. As shown in Figure 8C, the deviation generally remains within ±2%, with slight increases at the interface between the bone and lung areas. The maximum dose was calculated to be 74.19 Gy using the Monte Carlo method, while the DL-Dose Framework predicted a maximum dose of 74.25 Gy. The deviation in maximum dose values between the two methods was only 0.08% in this case.

Figures 9A and 9B display two-dimensional (x, z) images of the dose distribution and isodose curves obtained using the Monte Carlo method and the DL-Dose Framework. Figure 9C highlights the difference between the dose distribution calculated by the Monte Carlo method and the dose predicted by the DL-Dose Framework. Isodose curves corresponding to 90%, 80%, 70%, 60%, 50%, and 30% of the maximum dose obtained by the two methods are shown in Figure 9. Furthermore, Figure 9D compares the isodose curves derived from the Monte Carlo calculation and the DL-Dose Framework predictions. The isodose curves demonstrate a high level of agreement in the water region before heterogeneities and in the bone area. In the lung region, there is minimal variation in the dose distribution.

Figure 10 displays dose profiles for the second heterogeneous phantom at various depths calculated using the Monte Carlo method and predicted by the DL-Dose Framework. Figures 10A and 10B show dose profiles at depths of 10 cm in the water medium and 12.5 cm in the bone medium, respectively. Also, Figures 10C and 10D illustrate dose profiles at a depth of 16.5 cm at the interface between the bone and lung layers, and 20 cm within the lung layer. Finally, Figures 10E and 10F show dose profiles at a depth of 21.5 cm at the interface between the lung and water layer, and 25 cm in the fourth water layer.

The fraction of voxels with a delta index smaller than 0.01, 0.02, and 0.03 was determined using Equation 2 at various depths in the second heterogeneous phantom. Table 2 lists and summarizes the percentage of voxels with a delta index smaller than 0.01, 0.02, and 0.03 for the second heterogeneous phantom.

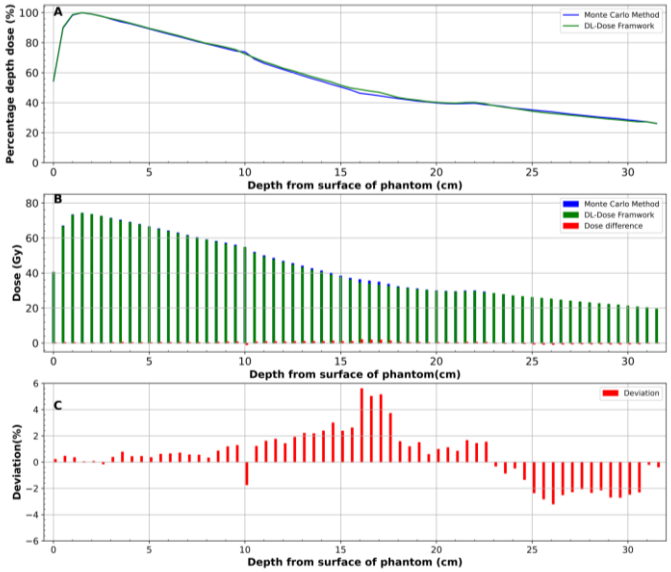

Figure 8. The illustration depicts PDD curves (A), the dose values for voxels along the central beam axis (B), and the deviation (C) for the second heterogeneous phantom.

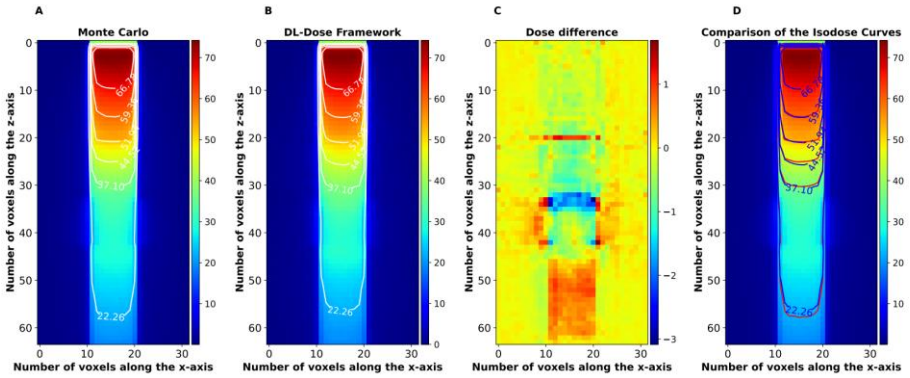

Figure 9. The illustration shows the two-dimensional (x, z) images of the dose distribution calculated by Monte Carlo (A), predicted by the Dl-Dose Framework (B), the dose difference between the two methods (C), and comparison of the isodose curves (D) for the second heterogeneous phantom.

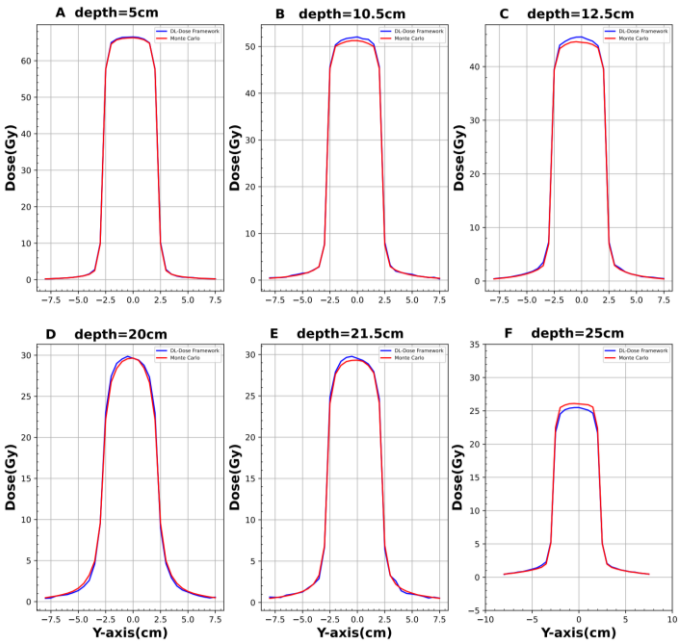

Figure 10. The illustration depicts dose profiles for the second heterogeneous phantom at various depths, calculated using the Monte Carlo method (red) and predicted by the DL-dose Framework (blue).

Table 2. The percentage of voxels with a delta index smaller than 0.01, 0.2, and 0.03 for the second heterogeneous phantom.

Second heterogeneous phantom	depth (cm)	fraction of the voxels in percentage with a delta index smaller than $\delta < 0.01$	fraction of the voxels in percentage with a delta index smaller than $\delta < 0.02$	fraction of the voxels in percentage with a delta index smaller than $\delta < 0.03$
In the water medium before heterogeneity	5	100	100	100
The interface of water and bone	10.5	97.6	100	100
Within the bone heterogeneity	12.5	98.1	100	100
The interface of bone and lung	16.5	86.9	89.8	95.7
Within the lung heterogeneity	20	94.2	99.4	100
The interface of the lung and water	21.5	98.3	100	100
In the water medium, beneath the heterogeneity	25	95.9	100	100

Discussion

As presented above, the thicknesses of the bone and lung inhomogeneity layers in the simulated heterogeneous phantoms were randomly varied between 1 cm and 5 cm. Additionally, the heterogeneous layers and the lung were randomly simulated at different locations along the Z-axis of the phantom. Consequently, various heterogeneous layers were incorporated with a high gradient in inhomogeneities from lung to bone to simulate these heterogeneous phantoms.

The mean squared error for both training and validation loss approaches zero after forty epochs for the respective datasets. This convergence of training and validation loss to zero indicates that the DL-Dose Framework can be effectively trained on heterogeneous phantoms, which exhibit significant variations in density, ranging from lung to bone and across different thicknesses of heterogeneities. It is important to note that employing a standard hierarchically dense (HD) U-Net in the design of the DL-Dose Framework helps reduce overfitting and facilitates the convergence of training and validation loss to zero, even when using only 120 heterogeneous phantoms as training data. Additionally, the developed DL-Dose Framework includes an independent input channel to account for the shape of the radiation field. This design allows each input channel to correspond to specific radiation fields. In other words, the distinct input channel in the framework can accept both the primary dose and the scattered dose, which are related to the shape of the radiation field.

Various methods can be used to compare the predicted dose with the reference dose. These methods can also serve as tests to evaluate the dose calculated by the treatment planning software against the results obtained from measurements using phantoms.[24-26]. One method involves comparing depth doses and beam profiles between the predicted and reference doses. In this context, Figures 4A and 8A illustrate the agreement between PDD curves predicted by the developed model and those observed using the Monte Carlo method, demonstrating a minimal discrepancy. Additionally, the bar graphs in Figures 4B and 8B show that the numerical values representing the differences in dose distribution obtained by both methods for the homogeneous and heterogeneous regions of two phantoms are less than 2 Gy. Furthermore, Figures 4C and 8C indicate that the deviation typically remains within ±2%.

Comparing isodose curves and creating graphical displays of dose differences by subtracting the reference dose distribution from the predicted dose distribution provides an effective method for evaluating predicted or calculated dose distributions [24-26]. As shown in Figures 5 and 9, the results of the isodose curves for values of 80%, 70%, 60%, 50%, and 30% of the maximum dose demonstrate good agreement with the isodose curves in the heterogeneous phantom. Additionally, the graphical representation of dose differences in Figures 5C and 9C illustrates minimal dose variation.

According to Figures 6 and 9, the cross-sectional profiles of the dose distribution generated by the developed model demonstrate exact symmetry with those obtained from the Monte Carlo method across various regions of the heterogeneous phantom, including bone, lung, and the interfaces between these inhomogeneous areas.

Furthermore, Tables 1 and 2 indicate that the percentage of voxels with a delta index less than 0.01 exceeded 100% for the water area before heterogeneities. As shown in these tables, 98.1% of the voxels exhibited a delta index smaller than 0.01 for inhomogeneities such as bone. Additionally, the minimum fraction of voxels with a delta index below 0.01 was 94.2% for lung inhomogeneities, demonstrating high accuracy in dose prediction for the heterogeneous phantom. As discussed, a set of heterogeneities consisting of the highest and lowest electron densities adjacent to one another was tested within a class of heterogeneous phantoms. The primary criteria recommended by Low et al. for Intensity-Modulated Radiation Therapy (IMRT) and Volumetric Modulated Arc Therapy (VMAT) treatment methods are 3%/3mm[22]. The American Association of Physicists in Medicine (AAPM) Task Group (TG-119) also considers 3%/3mm to be an acceptable criterion for the gamma passing rate[27]. Furthermore, the AAPM Task Group (TG-119) recommends a pass rate of over 95% at the evaluated point when applying the 3%/3mm

criteria[27-28]. Recent studies have indicated that the commonly used criteria (specifically, 3%/3 mm) are inadequate for identifying clinically significant errors in various measurement contexts[29]–[32]. Furthermore, the report from the American Association of Physicists in Medicine Task Group 218 (AAPM TG-218) recommends implementing a more stringent criterion of 3%/2 mm[33]. Based on these results, the DL-Dose Framework's capability to predict the dose distribution of heterogeneous phantoms, utilizing these comprehensive inputs, was well validated.

The following studies contribute to the field of dose prediction using deep learning techniques. Nguyen et al. utilized data from 88 clinical coplanar IMRT prostate cancer patients to predict dose distributions using a U-Net architecture [34]. Similarly, Kearney et al. developed the DoseNet architecture and employed data from 151 prostate cancer patients. Both studies demonstrated that deep learning methods can accurately predict dose distribution based on the contours of prostate patients[35]. A significant challenge in dose prediction using deep learning is maintaining spatial consistency concerning neighboring anatomical structures. The results indicate that the model can predict dose distribution in heterogeneous phantoms with an accuracy comparable to that of the Monte Carlo method, even in the presence of substantial variations in CT numbers and steep gradients of tissue heterogeneity, all without relying on patient contours.

The framework, including capabilities and limitations, can be evaluated and tested for clinical applications. The most critical inputs for the framework are the CT number and mass density of each voxel, which can be easily extracted from CT image data accurately.

The water dose distribution is used as a shared input for the framework during dose prediction. The water dose distribution can be obtained through direct measurements in a water phantom irradiated by an accelerator or by calculating the dose distribution from CT images of an equivalent water phantom using treatment planning software. Using water dose distribution as input within the framework is advantageous because it can be applied to all scenarios.

However, a limitation of the framework is its dependence on the beam shape of the radiation fields as input. In medical applications, optimal beam shapes are customized for each patient. Consequently, beam shape data, including field sizes and multi-leaf collimator (MLC) configurations, can be extracted from log files and utilized as input for the framework. Additionally, the number of input matrix modes associated with the shape of the radiation fields is significantly large because of the variety of modes present. Therefore, the computational requirements will increase with a much larger dataset.

The most significant advantage of the framework is its ability to predict dose distribution without relying on the contouring of sensitive organs or their positions. In other words, this framework operates differently from knowledge-based planning systems, which depend on clinical data, including organs at risk (OAR) or planning target volumes (PTVs). Consequently, it is advisable to expand the framework due to its comprehensive nature.

Additionally, heterogeneous phantom data with various inhomogeneous gradients were employed in the training, validation, and testing of the deep learning framework. Since most clinical or high-quality input data do not exhibit significant gradient heterogeneities, incorporating such data into the training, validation, and testing processes can lead to more accurate results. Consequently, future work should concentrate on enhancing the framework with high-quality clinical data and further developing it for clinical applications.

Conclusion

The DL-Dose Framework was evaluated to predict dose distribution in heterogeneous phantoms using dose distribution in homogeneous water phantoms and the physical properties of heterogeneous phantoms. The results showed good similarities between the expected dose distribution by the DL-Dose Framework and the dose calculation using the Monte Carlo method in the heterogeneous phantoms, demonstrating the effectiveness of the DL-Dose Framework for accurate dose prediction in heterogeneous phantoms. Also, the DL-Dose Framework could accurately predict dose distribution for mediums with high heterogeneity differences, such as bone and lung heterogeneities.

Also, the DL-Dose Framework was trained once, with a time proportional to the number of epochs. After that, the results for the testing data could be obtained within a few seconds. Compared to Monte Carlo methods, which consume a relatively long time per simulation, the DL-Dose Framework, based on deep learning methods, runs for a shorter duration.

Acknowledgment

This work was supported by the Amir Kabir University of Technology. We would like to thank all the participants in this study for their time and willingness to share their experiences. Also, all authors declare that they have no conflicts of interest.

References

1. Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013;35(8):1798–828. doi:10.1109/TPAMI.2013.50
2. Schmidhuber J. Deep Learning in neural networks: An overview. Neural Networks. 2015;61:85–117. doi:10.1016/j.neunet.2014.09.003
3. Willems S, Crijns W, Sterpin E, Haustermans K, Maes F. Feasibility of CT-only 3D dose prediction for VMAT prostate plans using deep learning. In. Cham: Springer International Publishing; 2019. p. 10–7. doi:10.1007/978-3-030-32486-5_2
4. Barragán-Montero AM, Nguyen D, Lu W, Lin MH, Norouzi-Kandalan R, Geets X, et al. Three-

dimensional dose prediction for lung IMRT patients with deep neural networks: robust learning from heterogeneous beam configurations. Medical Physics. 2019 Aug 1;46(8):3679–91. doi:10.1002/mp.13597 PubMed PMID: 31102554.
5. Nguyen D, Jia X, Sher D, Lin MH, Iqbal Z, Liu H, et al. 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture. Physics in Medicine and Biology. 2019;64(6). doi:10.1088/1361-6560/ab039b PubMed PMID: 30703760.
6. Shiraishi S, Moore KL. Knowledge-based prediction of three-dimensional dose distributions for external beam radiotherapy. Medical Physics. 2016 Jan 1;43(1):378–87. doi:10.1118/1.4938583 PubMed PMID: 26745931.
7. Tran A, Woods K, Nguyen D, Yu VY, Niu T, Cao M, et al. Predicting liver SBRT eligibility and plan quality for VMAT and 4Π plans. Radiation Oncology. 2017 Apr 24;12(1):70–9. doi:10.1186/s13014-017-0806-z PubMed PMID: 28438215.
8. Wu B, Ricchetti F, Sanguineti G, Kazhdan M, Simari P, Chuang M, et al. Patient geometry-driven information retrieval for IMRT treatment plan quality control. Medical Physics. 2009;36(12):5497–505. doi:10.1118/1.3253464 PubMed PMID: 20095262.
9. Wu B, Pang D, Simari P, Taylor R, Sanguineti G, McNutt T. Using overlap volume histogram and IMRT plan data to guide and automate VMAT planning: A head-and-neck case study. Medical Physics. 2013;40(2):021714. doi:10.1118/1.4788671 PubMed PMID: 23387737.
10. Zhu X, Ge Y, Li T, Thongphiew D, Yin FF, Wu QJ. A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. Medical Physics. 2011;38(2):719–26. doi:10.1118/1.3539749
11. Walters B, Kawrakow I, Rogers DWO. DOSXYZnrc Users Manual. Nrc Report Pirs. 2005;57–85.
12. Mohan R, Chui C, Lidofsky L. Differential pencil beam dose computation model for photons. Medical Physics. 1986;13(1):64–73. doi:10.1118/1.595924
13. 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 subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Athens, Greece: Springer; 2016. p. 424–32. doi:10.1007/978-3-319-46723-8_49
14. Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 Jun 15.
15. Nguyen D, Barkousaraie AS, Shen C, Jia X, Jiang S. Generating Pareto optimal dose distributions for radiation therapy treatment planning. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22. Springer; 2019. p. 59–67.
16. Nguyen D, Long T, Jia X, Lu W, Gu X, Iqbal Z, et al. A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning. Scientific Reports. 2019;9(1):1076–86.
17. Papanikolaou N and BJJ and BAL and KC and KE and MTR and SM and VDJ. AAPM report 85:Tissue Inhomogeneity Corrections for Megavoltage Photon Beams. Report of the AAPM radiation therapy committee task group 65. Delta. AAPM Report No. 85,; 2004. 136 p.
18. Kanani P, Padole M. Deep learning to detect skin cancer using google colab. International Journal of Engineering and Advanced Technology Regular Issue. 2019;8(6):2176–83. doi:10.35940/ijeat.F8587.088619
19. Kingma DP. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014;1–15.
20. Caruana R, Lawrence S. Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping. Neural Information Processing Systems. 2001;13:402–8.
21. Prechelt L. Automatic early stopping using cross validation: quantifying the criteria. Neural networks. 1998;11(4):761–7.
22. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. Medical physics. 1998 May;25(5):656–61.
23. Mentzel F, Paino J, Barnes M, Cameron M, Corde S, Engels E, et al. Accurate and Fast Deep Learning Dose Prediction for a Preclinical Microbeam Radiation Therapy Study Using Low-Statistics Monte Carlo Simulations. Cancers. 2023 Apr 1;15(7). doi:10.3390/cancers15072137
24. Fraass B, Doppke K, Hunt M, Kutcher G, Starkschall G, Stern R VDJ. American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning. Medical physics. 1998 Oct;25(10):1773–829.
25. Physics M, Guideline P, Smilowitz JB, Das IJ, Feygelman V, Fraass BA, et al. AAPM Medical Physics Practice Guideline 5 . a .: Commissioning and QA of Treatment Planning Dose Calculations — Megavoltage Photon and Electron Beams. Vol. 16. 2015;16(5):14–34.
26. International Atomic Energy Agency (IAEA). " Commissioning and quality assurance of computerized planning systems for radiation treatment of cancer", Technical Reports Series, TRS-430. International Atomic Energy Agency, Technical Report Series. 2004;(430):1–281.
27. Ezzell GA, Burmeister JW, Dogan N, Losasso TJ, Mechalakos JG, Mihailidis D, et al. IMRT commissioning: Multiple institution planning and dosimetry comparisons, a report from AAPM Task

Group 119. Medical Physics. 2009;36(11):5359–73. doi:10.1118/1.3238104 PubMed PMID: 19994544.
28. Low DA, Moran JM, Dempsey JF, Dong L, Oldham M. Dosimetry tools and techniques for IMRT. Medical Physics. 2011;38(3):1313–38. doi:10.1118/1.3514120 PubMed PMID: 21520843.
29. Abdulbaqi AM, Abdullah SS, Alabed HH, Alazawy NM, Al-Musawi MJ, Heydar AF. The Correlation of Total MU Number and Percentage Dosimetric Error in Step and Shoot IMRT with Gamma Passing Rate Using OCTAVIUS 4D-1500 Detector Phantom. Annals of Tropical Medicine and Public Health. 2020;23(19). doi:10.36295/asro.2020.232126
30. Templeton AK, Chu JCH, Turian J V. The sensitivity of ArcCHECK-based gamma analysis to manufactured errors in helical tomotherapy radiation delivery. Journal of Applied Clinical Medical Physics. 2015;16(1):32–9. doi:10.1120/jacmp.v16i1.4814 PubMed PMID: 25679155.
31. Nelms BE, Chan MF, Jarry G, Lemire M, Lowden J, Hampton C, et al. Evaluating IMRT and VMAT dose accuracy: Practical examples of failure to detect systematic errors when applying a commonly used metric and action levels. Medical Physics. 2013;40(11). doi:10.1118/1.4826166 PubMed PMID: 24320430.
32. Vieillevigne L, Molinier J, Brun T, Ferrand R. Gamma index comparison of three VMAT QA systems and evaluation of their sensitivity to delivery errors. Physica Medica. 2015 Nov 1;31(7):720–5. doi:10.1016/j.ejmp.2015.05.016 PubMed PMID: 26095758.
33. Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. Medical Physics. 2018;45(4):e53–83. doi:10.1002/mp.12810 PubMed PMID: 29443390.
34. Nguyen D, Long T, Jia X, Lu W, Gu X, Iqbal Z, et al. Dose Prediction with U-net: A Feasibility Study for Predicting Dose Distributions from Contours using Deep Learning on Prostate IMRT Patients. arXiv preprint arXiv:170909233. 2017;17.
35. Kearney V, Chan JW, Haaf S, Descovich M, Solberg TD. DoseNet: A volumetric dose prediction algorithm using 3D fully-convolutional neural networks. Physics in Medicine and Biology. 2018 Dec 1;63(23):235022. doi:10.1088/1361-6560/aaef74 PubMed PMID: 30511663.