HHS Public Access

Author manuscript

Phys Med Biol. Author manuscript; available in PMC 2023 October 19.

Published in final edited form as:

Phys Med Biol. ; 67(21): . doi:10.1088/1361-6560/ac9663.

Validation of a Deep Learning-Based Material Estimation Model for Monte Carlo Dose Calculation in Proton Therapy

Chih-Wei Chang[1], **Shuang Zhou**[2], **Yuan Gao**[1], **Liyong Lin**[1], **Tian Liu**[1], **Jeffrey D. Bradley**[1], **Tiezhi Zhang**[2], **Jun Zhou**[1], **Xiaofeng Yang**[1,3]

[1]Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA 30308

[2]Department of Radiation Oncology, Physics Division, Washington University in St. Louis School of Medicine, St. Louis, MO 63110

[3]Department of Biomedical Informatics, Emory University, Atlanta, GA 30308

Abstract

Objective: Computed tomography (CT) to material property conversion dominates proton range uncertainty, impacting the quality of proton treatment planning. Physics-based and machine learning-based methods have been investigated to leverage dual-energy CT (DECT) to predict proton ranges. Recent development includes physics-informed deep learning (DL) for material property inference. This paper aims to develop a framework to validate Monte Carlo dose calculation (MCDC) using CT-based material characterization models.

Approach: The proposed framework includes two experiments to validate *in vivo* dose and water equivalent thickness (WET) distributions using anthropomorphic and porcine phantoms. Phantoms were irradiated using anteroposterior proton beams, and the exit doses and residual ranges were measured by MatriXX PT and multi-layer strip ionization chamber. Two pre-trained conventional and physics-informed residual networks (RN/PRN) were used for mass density inference from DECT. Additional two heuristic material conversion models using single-energy CT (SECT) and DECT were implemented for comparisons. The gamma index was used for dose comparisons with criteria of 3%/3mm (10% dose threshold).

Main results: The phantom study showed that MCDC with PRN achieved mean gamma passing rates of 95.9% and 97.8% for the anthropomorphic and porcine phantoms. The rates were 86.0% and 79.7% for MCDC with the empirical DECT model. WET analyses indicated that the mean WET variations between measurement and simulation were −1.66 mm, −2.48 mm, and −0.06 mm for MCDC using a Hounsfield look-up table with SECT and empirical and PRN models with DECT. Validation experiments indicated that MCDC with PRN achieved consistent dose and WET distributions with measurement.

Corresponding Author: Xiaofeng Yang, xiaofeng.yang@emory.edu.

Ethical Statement

Emory IRB review board approval was obtained (IRB #114349), and informed consent was not required for this Health Insurance Portability and Accountability Act (HIPAA) compliant retrospective analysis.

Significance: The proposed framework can be used to identify the optimal CT-based material characterization model for MCDC to improve proton range uncertainty. The framework can systematically verify the accuracy of proton treatment planning, and it can potentially be implemented in the treatment room to be instrumental in online adaptive treatment planning.

1 Introduction

Clinical evidence has shown that proton therapy can reduce side effects and unplanned hospitalizations for patients compared to conventional photon therapy. In contrast to a photon treatment modality, proton pencil beam scanning machines can deliver conformal doses to the target volumes while sparing the adjacent organs at risk (van de Water *et al.*, 2011; Knopf and Lomax, 2013). Such sharp dose characteristics require definite proton range prediction to ensure dosimetry consistency between the treatment planning system (TPS) and treatment delivery system (Baumann *et al.*, 2016). Modern Monte Carlo-based TPS (Paganetti *et al.*, 2008; Chang *et al.*, 2020) requires material mass density information of patients to perform dose calculations for treatment planning. This information is usually acquired from single-energy computed tomography (SECT) images via a Hounsfield look-up table (Schneider *et al.*, 2000). However, the SECT to material property conversion dominates proton range uncertainty, and a typical margin of 2.5% is reserved for this uncertainty (Paganetti, 2012).

During the last decade, dual-energy computed tomography (DECT) has been explored to improve proton dose calculation (Bourque *et al.*, 2014; van Elmpt *et al.*, 2016; Wohlfahrt *et al.*, 2017). A simulation-based retrospective patient study has shown that DECT can be used to refine a conventional Hounsfield look-up table to improve proton range uncertainty (Wohlfahrt *et al.*, 2020). Meanwhile, data-driven modeling methods such as machine learning (ML) and deep learning (DL) have been investigated to derive accurate proton relative stopping power (RSP) (Su *et al.*, 2018) and material mass density (Chang *et al.*, 2022a). However, the performance of data-driven models may be compromised by measurement noises due to the ill-posed nature of inverse modeling (Arridge *et al.*, 2019). The issue can potentially be resolved by using physics-informed machine learning to regularize the models with physics insights (Chang and Dinh, 2019; Karniadakis *et al.*, 2021). Chang *et al.* (2022b) proposed a physics-informed DL framework to supervise the learning with a physics model. This approach can increase the predictive capability of material mass density mapping using DECT images. Although various data-driven methods have been investigated for DECT to material property conversion, there is still a lack of direct proton measurement to quantify the uncertainty for ML/DL methods regarding dose and range distributions.

Various methods have been explored to minimize proton range uncertainty, such as magnetic resonance imaging, range probe, and proton radiography (Knopf and Lomax, 2013). The proton imaging using multi-layer ionization chambers (MLIC) features real-time measurement without additional implants. The method is promising for clinical applications regarding *in vivo* range verification (Rinaldi *et al.*, 2014; Deffet *et al.*, 2020). Proton range measurement using a commercial MLIC typically requires 20 minutes to cover an area of 180×180 mm^2 by multiple proton fields with a size of 45×45 mm^2 and 1296 total spots

(Farace *et al.*, 2016). At the same time, dosimetry verification is essential to understand the limits of DL-based material conversion methods. Conventional two-dimensional (2D) ion chambers for patient-specific dose measurement require changes in experiment setups for different measured depths (Arjomandy *et al.*, 2010; Karger *et al.*, 2010).

This study proposes a validation framework to evaluate the feasibility of Monte Carlo dose calculation (MCDC) using CT-based material characterization methods, including mechanistic and physics-informed DL models. For the first time, the framework leverages the state-of-the-art multi-layer strip ionization chamber (MLSIC) detector (Zhou *et al.*, 2022), which features spatiotemporal signal identification. The device can effectively capture the dosimetric characteristics of a proton treatment beam, including position, energy, profile, and dose spot-by-spot. In contrast, commercial MLIC devices such as Zebra (IBA Dosimetry, Germany) can only measure a spot's depth dose information. The newly designed MLSIC can measure a single proton field with a field size of 160×160 mm^2 and 1681 spots within 4 minutes, and this measured time is dominated by the beam delivery time. The device can be used to efficiently quantify *in vivo* proton range uncertainty within patients. We also emulate the proton dosimetry measurement at various depths using multiple beam energies to efficiently and systematically verify dose accuracy without changing experiment setups.

To demonstrate the usability of the proposed validation framework, we investigate in which CT-based material characterization models can work robustly, effectively, and accurately with proton MCDC. While the previous work (Chang *et al.*, 2022b) only focused on image quality assurance, the originality of the present work can be summarized in three aspects to conduct the feasibility study for clinical implementation:

- The proposed validation framework, for the first time, leverages the 4D MLSIC to systematically and efficiently validate CT-based material characterization models for MCDC within the measurement time of four minutes.
- The proposed validation framework can conserve the proton physics and quantify the uncertainty for CT-based material characterization methods caused by photon physics.
- The proposed validation framework can holistically evaluate in vivo proton range distributions and identify the sources of uncertainty (i.e., lung, soft, or bone tissues) to enable divide-and-conquer strategies to further improve range inaccuracy in proton therapy.

2 Materials and methods

Monte Carlo dose calculation (MCDC) algorithms have been implemented into proton TPS to achieve accurate and robust treatment planning. The accuracy of MCDC is impacted by CT material characterization and proton range uncertainty. Mechanistic and semi-analytical models (Schneider *et al.*, 2000; Meyer *et al.*, 2010; Bourque *et al.*, 2014) have been developed to characterize materials from CT for proton therapy. A recent investigation involves using conventional and physics-informed DL methods (Su *et al.*, 2018; Chang *et al.*, 2022c; Chang *et al.*, 2022d) to improve the uncertainty from CT to material conversion.

However, there is still a lack of proton experiment data to validate those newly proposed DL models. This work investigates the feasibility of utilizing physics-informed DL-based CT material characterization for proton therapy using the proposed validation framework and state-of-the-art detectors. The proposed framework systematically validates the physics-informed DL model regarding proton dosimetric and range accuracies.

2.1 CT material characterization using physics-informed deep learning

Occam's razor is preferred by conventional data-driven methods because the simplest model is usually interpretable and generalizable (Blumer *et al.*, 1987; Domingos, 1999). However, the model performance may be saturated when dealing with a substantial amount of data (Champion *et al.*, 2019). DL with hierarchical model structures has been proven as a universal approximator that can discover complex patterns from data (Hornik *et al.*, 1989; LeCun *et al.*, 2015). The so-called physics-informed DL uses physics insights to regularize conventional DL models (pure data driven) to increase the predictive capability especially when the data are insufficient to allow DL models to capture the underlying physics (Chang and Dinh, 2019). Chang *et al.* (2022b) have introduced a physics-informed loss function to regularize the training of DL models for material property inference from DECT images.

Eq. (1) gives a loss function ($\mathscr{L}$), applied to supervise the learning of physics-informed DL models to find the optimal mass density where $\mathscr{L}_{MSE}$ and $\mathscr{L}_{physics}$ are the conventional mean square error (MSE) and physics-informed loss functions. The conventional DL models are trained merely using $\mathscr{L}_{MSE}$. Eq. (2) defines a physics loss where $y_{physics}$ insight and y_{meas} denote a physics-based model and measured data of CT Hounsfield units.

$$\mathscr{L} = \mathscr{L}_{MSE} + \mathscr{L}_{physics} \tag{1}$$

$$\mathscr{L}_{physics} = \sum \| y_{physics\ insight} - y_{meas} \|_2^2 \tag{2}$$

Eq. (3) defines a heuristic model for $y_{physics\ insight}$ (Chang *et al.*, 2022b) where $\rho_{m,DL}$ presents the mass density, queried from DL models during each training iteration. The $\rho_{e,w}$ is the electron molar density for water (0.56 mol/cm^3). The $\sum_i \omega_i {Z_i}/{A_i}$ presents the electron molar density for calibration materials where ω, Z, and A are the elemental weight percent, atomic number, and atomic mass number. The $\tilde{z}$ and $\hat{z}$ are parameters that can be derived from the elemental composition of phantom materials using the power law additivity rule (Mayneord, 1937; Spiers, 1946; Bourque *et al.*, 2014). The k_{ph}, k_{coh}, and k_{incoh} are fitting parameters (Jackson and Hawkes, 1981; Schneider *et al.*, 1996), depending on a CT energy spectrum. Based on our institutional data, the estimated values are 1.4×10^{-5}, 1.6×10^{-3}, and 0.9, respectively.

$$y_{physics\ insight} \equiv 1000 \left(\frac{\rho_{m,\,DL} \sum_i \omega_i \frac{Z_i}{A_i}}{\rho_{e,\,w}} \cdot \frac{k_{ph}\tilde{z}^{3.62} + k_{coh}\hat{z}^{1.86} + k_{incoh}}{k_{ph}\tilde{z}_w^{3.62} + k_{coh}\hat{z}_w^{1.86} + k_{incoh}} - 1 \right) \tag{3}$$

2.2 Validation framework for Monte Carlo dose calculation

Figure 1 depicts the framework to validate proton MCDC using CT-based material mass density characterization models including a Hounsfield look-up method, empirical correlation, and conventional and physics-informed DL-based models. The workflow includes three steps: a) computed tomography (CT) numbers to material mass density conversion; b) proton treatment planning for beam delivery; c) validation experiments to measure the dose and WET distributions using proton pencil beams. The performance of each model was evaluated, and the optimal method should allow MCDC to be consistent with measured dose and WET distributions.

2.2.1 Data acquisition—Figure 1(c4)–(c5) shows an adult anthropomorphic phantom, CIRS (Computerized Imaging Reference Systems, Inc., Norfolk, VA, USA) Atom M701 and a porcine tissue phantom (meatPhan). The phantoms were irradiated by anterior proton pencil beams to acquire dosimetry and WET data for validation. The meatPhan included air (lung surrogate), porcine adipose, muscle, rib, and femur to increase the heterogeneity of the phantom. The meatPhan was made with a 200×200×200 mm3 container. All Phantom images were acquired with a Siemens SOMATOM Definition Edge scanner. To reduce CT image noise, the reconstruction kernels of I41s/3 and Q30f/3 with sinogram-affirmed iterative reconstruction (SAFIRE) was used for SECT and DECT. DECT material decomposition was performed using Siemens Syngo.Via, and the acquired parametric maps included effective atomic number, electron density relative to water, and virtual monochromatic images. Table 1 summarizes the acquisition parameters for CT scans.

2.2.2 CT numbers to material mass density conversion—Multiple CT-based material characterization models were implemented to demonstrate the proposed validation framework. Two heuristic models, a Hounsfield look-up model (Schneider *et al.*, 2000) and empirical correlation (Hünemohr *et al.*, 2014), were implemented to characterize material properties using SECT and DECT images. Figure 1(a1)/(a2) shows that CT numbers are directly obtained from the CT machine. Siemens Syngo.Via was used to derive DECT parametric maps in Figure 1(a2), including the effective atomic number (Z_{eff}), relative electron density (ρ_e), and virtual monochromatic images (VMI) of 80 keV, 135 keV, and 190 keV. Figure 1(a4) shows the CT-density curve for the look-up method in SECT. Table A1 (Appendix A) gives the Hounsfield look-up table to convert CT numbers to material mass densities (Chang *et al.*, 2020). Figure 1(a5) gives the empirical model using Z_{eff} and ρ_e to derive relative stopping power (RSP) maps. Then the voxel RSP can be converted to mass densities in Table A1 (Appendix A). MATLAB R2021a was used to program heuristic models.

We adopted two pre-trained residual networks (ResNet) (He *et al.*, 2016; Wang *et al.*, 2018) for evaluation: conventional ResNet (RN) and physics-informed ResNet (PRN) in Figure 1(a6) from the previous work (Chang *et al.*, 2022b). Figure 2 depicts the model structure of ResNet, including four convolutional layers, two fully connected layers, and twenty residual boxes. Each residual block includes two convolutional and single residual layers. Rectified linear units (ReLU) (Nair and Hinton, 2010) were used as the activation function. Table B1 (Appendix B) summarizes the model parameters for the ResNet. RN and PRN share

the same model architecture but use different loss functions. RN was trained only using the MSE loss. In contrast, PRN was trained using the MSE and physics-informed loss functions given by Eq. (1). The training data included 79 CT slides (512×512×79 voxels) from an electron density phantom with known material properties (Chang *et al.*, 2022b). Training times for RN/PRN were approximately two hours using NVIDIA Quadro RTX A6000. The validation dataset included 512×512×651 CT voxels from the anthropomorphic phantom in Figure 1(c4). DL models were implemented using the PyTorch framework (Paszke *et al.*, 2019).

2.2.3 Material composition assignment for proton Monte Carlo dose calculation—

We used the RayStation 10B (RayPhysics, 2020) to perform the Monte Carlo dose calculation. RayStation includes an internal table of 50 materials to correlate the CT voxel density to the proper material composition and mean ionization energy. The 50 pre-tabulated materials are interpolated from the eight basic materials given in Table 2 with the compositions from the literature (ICRP23, 1975; ICRU44, 1989; ICRU49, 1993). Eq. (4) gives the inverse rule of mixtures where ρ, w, Mix, U, and L denote the density, weight percent mass, mixture (interpolated material), basic material as the upper bound, and basic material as the lower bound. The selection of the upper and lower bounds is based on the interpolated material density that falls between which two basic materials. Eq. (4) can be rearranged, and we can derive w_L as density functions given by Eq. (5). Eq. (6) shows the solution for w_U since the summation of w_U and w_L is one. The densities of the interpolated 50 materials range from 0.001 g/cm^3 to 2.7 g/cm^3 with an incremental density of 0.055 g/cm^3. Then the elemental composition of each interpolated material can be obtained by Eq. (5)–(6) and the basic materials given in Table 1. RayStation will assign the pre-tabulated material composition to each CT voxel based on the material closest in mass density.

$$\frac{1}{\rho_{Mix}} = \frac{w_L}{\rho_L} + \frac{1 - w_L}{\rho_U} \tag{4}$$

$$w_L = \frac{\rho_L}{\rho_{Mix}} \cdot \frac{\rho_U - \rho_{Mix}}{\rho_U - \rho_L} \tag{5}$$

$$w_U = 1 - w_L \tag{6}$$

2.2.4 Proton planning for beam delivery—

The treatment planning system (TPS), RayStation 10B, was used to design two types of proton treatment plans for dosimetry and WET measurement, as shown in Figure 1(b1) and Figure 1(b2). Firstly, the proton plans for dosimetry included an anterior proton beam at a 0° gantry angle, and the exit dose was measured for the validation procedure. Each proton plan was designed with a single energy beam to investigate the impacts on anatomical heterogeneity from different phantoms and phantom sites. Multiple single-energy plans were used to emulate the proton dose distribution at different depths without the need to change the experiment setup. Secondly, proton plans for WET measurement used multiple proton beams with large spot spacing (40 mm) to obtain integrated depth doses for range calculation from each proton spot. Because

sparse spots were used for WET measurement plans, multiple proton beams were delivered to achieve a spatial resolution of 4 mm for WET maps. Table 3 gives the details of proton plans for each measurement and Monte Carlo dose calculation (MCDC). NVIDIA Quadro RTX 8000 was used for MCDC.

2.2.5 Validation experiment—Proton pencil beams were delivered by Varian ProBeam System (Varian Medical Systems, Palo Alto). To measure dose distribution, Figure 1(c1) depicts a MatriXX PT (IBA Dosimetry, Germany) with an active area and resolution of 244×244 mm^2 and 7.6 mm. A solid water phantom (2-cm thick) is placed on the MatriXX PT for dose buildup. For WET measurement, Figure 1(c2) shows a multi-layer strip ionization chamber (MLSIC) that is featured in the capability of spatiotemporal measurement with a large active area (Zhou *et al.*, 2022). The detector includes an active area, spatial, and time resolution of 256×256 mm^2, 2 mm, and 0.32 ms. Figure 1(c3) shows the experiment setup for dosimetry and WET measurement using an anterior proton beam.

To measure a WET map, MLSIC is first used to measure the residual range (R_{phantom}) of exit proton beams from the phantom. The total measured time is approximately four minutes, dominated by the machine preparation time of treatment delivery systems. The phantom WET is defined as the water thickness, which causes the same R_{phantom}. In actual experiment design, we can do the measurement twice, with and without the phantom, using the identical proton beam and experiment setup to obtain $R_{phantom}$ and the proton range without the phantom (R_{water}). Then the WET can be obtained by taking the difference between the two ranges ($R_{water} - R_{phantom}$) (Zhang *et al.*, 2010). The 80% distal range (R80) should be used for the WET calculation because R80 represents that the electromagnetic interactions have stopped 50% of the protons (Paganetti, 2018). Therefore, R80 has minimum impacts by the energy spread of proton beams (Hsi *et al.*, 2009; Schuemann *et al.*, 2014). We use Eq. (7) to derive WET, where x denotes R80 from MLSIC or TPS.

$$WET_x = R80_{water} - R80_x \tag{7}$$

2.3 Evaluation

Comparisons of dose-difference distributions can show the consistency between the measurement and simulation (Mah *et al.*, 1989). However, a two-dimensional (2D) comparison is challenging for a steep dose gradient region since small spatial uncertainty can result in significant discrepancies between the measurement and simulation (Low *et al.*, 1998). A distance-to-agreement (DTA) concept is introduced to properly quantify this discrepancy to define the acceptance criterion for dose comparisons (Hogstrom *et al.*, 1984; ICRU42, 1987). The DTA is a distance used to search the calculated dose point, which best agrees with a measured point. Low *et al.* (1998) proposed a quantity, the so-called gamma index, to quantify dose distribution discrepancies using acceptance criteria of percentage dose difference and DTA. The calculation passes the criteria if the gamma index is less than one. Otherwise, the calculation fails. The percentage gamma passing rate (%GP) is derived by taking the passing image pixels divided by total image pixels. High-dose regions need to be verified for radiation therapy to ensure the coverage of lesion volumes, and a 10% dose threshold is typically applied before gamma analysis.

For proton therapy, the Imaging and Radiation Oncology Core (IROC) (Taylor *et al.*, 2016) defines the validation procedure for proton institutions participating in the clinical trials sponsored by National Cancer Institute. IROC provides the standard criteria for various anthropomorphic phantoms. The criteria (Kerns *et al.*, 2016; Taylor *et al.*, 2016; Taylor *et al.*, 2017) are: 5%/3 mm for brain; 7%/ 4mm for head and neck; 7%/5mm for left lung; 7%/ 4 mm for liver; 5%/5 mm for spine; 7%/4 mm for pelvis. The acceptable gamma passing rate is 85%. Our institutional gamma criteria are 3%/2 mm for patient quality assurance using a homogeneous water phantom. However, this study used anthropomorphic phantoms, including heterogeneous effects from the anatomy and geometry that increase the difficulty for gamma comparisons. We used the gamma criteria of 3%/3 mm and a 10% dose threshold to investigate the discrepancy between measured and simulated dose distribution. The optimal simulated planar dose for gamma analysis was searched from a 3D simulated dose matrix with an anteroposterior dose grid size of 1 mm.

The WET variation is another evaluation metric to quantify the discrepancies between the measurement and simulation using 3D measured data. This quantity was explored since the *in vivo* proton range uncertainty can be estimated from the residual ranges of proton penetrating through phantoms (Cormack, 1963; Schneider *et al.*, 2004; Mumot *et al.*, 2010). Eq. (8) gives the WET variation (ΔWET) by using the measured WET (WET_{meas}) from MLSIC and simulated WET (WET_{sim}) from TPS. We used ΔWET for histogram distribution analyses to compare the mean and standard deviation of ΔWET predicted by different models. We took the absolute values of ΔWET to generate a 2D WET map to emphasize the regions that showed significant discrepancies between the measurement and simulation.

$$\Delta WET = WET_{sim} - WET_{meas} \tag{8}$$

3 Results

3.1 Gamma analysis for the anthropomorphic phantom and meatPhan

Figure 3(a)–(d) depict the gamma passing rate versus different proton energies for the CIRS anthropomorphic phantom. The passing rates are above 95% for RN and PRN at brain and HN sites. Figure 3(c) shows that only PRN achieves the gamma pass rates above 90% for all energies at the lung site, while the passing rates drop to 84% and 77% for the SECT and empirical models at 203 MeV. Figure 3(d) illustrates that the empirical model yields the gamma passing rates below 76% for all energies at the pelvic site. Figure 3(e) shows the gamma passing rate for meatPhan at various proton energies. Both RN and PRN models can result in the gamma passing rate above 95% across all energies. However, the passing rates decrease to 75% and 69% for the SECT and empirical models at 193 MeV and 192 MeV. Table 4 indicates that PRN achieves the maximum values of mean gamma passing rates for each phantom.

Figure 4–7 depict 2D dose distributions and gamma index maps by measurement and Monte Carlo simulation for the anthropomorphic phantom at different sites. Dose discrepancies between the measurement and SECT model can be observed at the sternum location for the brain and lung. The empirical model yields high gamma indexes in skull and sternum

regions for the brain, HN, and lung, and RN can improve those local high gamma indexes. Meanwhile, PRN can further reduce the gamma indexes compared to RN results. Figure 8 shows the dose distributions and gamma index maps for the meatPhan. The SECT, empirical, and RN models yield high gamma indexes in the porcine femur region. However, PRN can remedy the gamma indexes in the femur region.

3.2 Water equivalent thickness analysis for the anthropomorphic phantom

Figure 9 depicts the WET and WET variation (ΔWET) maps by measurement and Monte Carlo simulation for the CIRS anthropomorphic phantom at the pelvic site. The SECT and empirical model result in local high WET variation (ΔWET ~4 mm) in the ischium region. The empirical model also yields high local ΔWET in pelvic bone and sacrum. Compared to the SECT and empirical models, PRN can reduce the WET variation in the bone regions. Figure 10 shows WET variance distributions obtained by pixel-by-pixel comparisons of 2D WET maps between the measurement and simulation by each model using Eq. (8). Figure 10 shows that PRN achieves the minimum mean ΔWET of −0.06 mm compared to other models.

RN, and (e1) PRN for the anthropomorphic phantom at a pelvic site with a 216 MeV proton beam. (a2) Digitally reconstructed radiograph of the phantom anatomy with a red box to denote the beam field size. (b2)-(e2) Comparisons of the WET absolute difference (|ΔWET|) between measurement and MCDC with different images described as (b1)-(e1). The anthropomorphic phantom was irradiated with anterior beams and the proton residual ranges were measured with MLSIC for WET analyses.

4 Discussion

The current clinical proton patient quality assurance typically checks 2D dose distribution at two depths using a water phantom. This procedure is equivalent to using two energies for dose-difference comparisons by the proposed framework. Figure 3 demonstrates that randomly selecting two energies may not be entirely trustworthy when we compare the dose distributions using the anthropomorphic and porcine phantoms. The gamma passing rates can be high for some energies, but the passing rates drop below acceptance values for other energies. In contrast, the proposed framework adopts two evaluation metrics, dose and WET distributions, to systematically verify the accuracy of *in vivo* proton scattering and range distributions by Monte Carlo dose calculation. WET evaluation can quantify the proton range uncertainty, which dominates the treatment quality. Although Figure 3 depicts that both RN and PRN show similar performances regarding gamma passing rate, Figure 10 shows that RN results in 0.3 mm more range inaccuracy than PRN. The previous work (Chang *et al.*, 2022b) only focused on developing the material characterization method, and the previous model evaluation cannot directly correlate to clinical impacts. This work demonstrates that the proposed framework can systematically and efficiently validate Monte Carlo dose calculation using various CT-based material characterization models. The framework does not require a special setup for patients, and it can potentially be implemented in treatment rooms for online adaptive treatment planning.

The gamma analyses indicate that PRN improves the mean gamma passing rates by 1.8%, 1.4%, 4.0%, and 6.3% for the brain, HN, lung, and pelvis from the anthropomorphic phantom compared to the SECT model. The phantom study also shows that PRN can increase the overall mean gamma passing rates by 9.9% from the empirical model across all phantom sites. Figure 4, Figure 6, and Figure 7 depict that local high gamma indexes occur in the skull, spine, and femur regions for the SECT and empirical models. At the same time, PRN can enhance the dose agreement to the measurement in these bone regions. For the meatPhan, PRN can improve the mean gamma passing rates by 5.0% and 18.1%, compared to the SECT and empirical models. PRN with physics-informed training can additionally increase the mean gamma passing rate by 1.2% from the conventional DL model (RN). Figure 8 shows that local high gamma indexes happen in the femur bone region for most models except PRN. Generally, the PRN results are consistent with dosimetry measurement, and the gamma passing rates are above 90% for the anthropomorphic phantom and meatPhan.

Figure 3(e) shows that the gamma passing rates for the SECT and empirical models drop below 85% at 190–195 MeV. Based on the digitally reconstructed radiograph in Figure 8(a2), Figure 11 depicts that the gamma failures happen in the rib and femur bone regions (blue arrow). Figure 11(b1) displays that the SECT model allows fewer protons to penetrate through the rib region (blue arrow) compared to the empirical model in Figure 11(b2). However, both models disagree with the measurement at 190 MeV and 195 MeV in Figure 11(c1)–(c2)/(f1)–(f2). A high dose gradient usually occurs for the regions, including partial proton penetration, and a small spatial uncertainty can result in a significant dose difference(Low *et al.*, 1998). Figure 8(b1) shows that more protons penetrate through the rib region and the gamma passing rate increases due to reducing the dose gradient. Thus, the drop in gamma passing rates observed in Figure 3(e) is due to the range inaccuracy in the bone region. This result is consistent with the WET distributions in Figure 10 that the empirical model predicts shorter WET (longer residual ranges) than the SECT method. The result also suggests that it is more efficient to do radiation therapy quality assurance using the proposed WET measured method compared to the conventional 2D gamma analyses. The 2D gamma method requires measuring multiple energies to validate a model systematically.

Aside from dosimetry analyses, the measurement of WET maps can directly detect range discrepancies induced by TPS simulation with different mass density conversion models. Figure 9(b2) shows that the SECT model yields a WET variation of ~4 mm in pelvic bone regions compared to the measurement. The empirical model further worsens the WET agreement to the measured data in pelvic bone and sacrum regions, as shown in Figure 9(c2). Contrarily, the DL models can improve the WET inconsistency in these bone regions. Figure 10 depicts that PRN can improve the mean of WET distributions by 1.6 mm and 2.4 mm from the SECT and empirical models. Compared to the conventional DL model (RN), PRN with physics-informed training can further reduce the mean WET by 0.3 mm. The WET analyses show consistent results compared to the dosimetry analyses. Both validation experiments indicate that PRN can deliver the most accurate mass density maps for Monte Carlo dose calculation using the CIRS anthropomorphic phantom and meatPhan. The WET analyses require proton beams to penetrate through phantoms such that residual ranges can

be measured by MLSIC. However, the design of MLSIC includes a 2-cm buildup layer, and the residual proton range must be larger than the thickness of this layer to ensure proton penetration for the entire measured region. For a 216-MeV proton beam, the minimum residual range is 2.9 cm across the measured pelvic region from the anthropomorphic phantom. Therefore, the current WET method is limited to using high-energy proton beams (≥ 216 MeV).

Figure 3 and Figure 10 show that the empirical model consistently performs worse than the Hounsfield look-up method in SECT. The Hounsfield look-up table (HLUT) is machine-specific since each CT energy spectra can vary between each scanner due to different machine designs, manufacture of machine compartments, or manufacture of X-ray tubes. In contrast, we did not re-calibrate the empirical model using the institutional CT scanner. We used Siemens Syngo.Via to derive DECT parametric maps as the empirical model inputs, and the original model calibration was based on a different material decomposition method (Hünemohr *et al.*, 2014). Using twin-beam DECT protocols can also compromise the image quality (Almeida *et al.*, 2017), increasing uncertainty for the empirical model. Future work should include machine-specific and software-specific calibration to commission DECT material characterization models using the institutional data before using the model for proton treatment planning.

ML/DL-based methods have exhibited the capability of CT noise suppression and DECT parametric mapping for accurate RSP prediction (Su *et al.*, 2018; Chang *et al.*, 2022a). However, ML and DL belong to inverse modeling, and the methodology is ill-posed: the solution can be impacted by data noise and quantity (O'Sullivan, 1986). The optimal model should base on validation experiments to ensure the robustness of these inverse models. Therefore, establishing validation procedures become a crucial step in exploring the feasibility of these material mapping methods for clinical applications. Such procedures include validation experiments to quantify the clinical impacts and explore the limits of each model. The current work aims to develop a validation workflow using anthropomorphic and heterogeneous tissue phantoms for proton dosimetry and WET map measurement. The current framework adopts the 190-keV virtual monochromatic image, which can suppress metal artifacts, to extend the framework applicability of the framework for future spine implant or hip prosthesis patients (Wellenberg *et al.*, 2018a; Wellenberg *et al.*, 2018b). Future investigation will likely focus on using the proposed workflow to measure WET maps of various animal tissue or tissue substitute phantoms to quantify proton range uncertainty for different tissues. Most importantly, the impact of tissue heterogeneity on the range uncertainty needs to be evaluated, and the results can be potentially used to prevent ML and DL from outputting physically unreasonable values.

5 Conclusions

A validation framework was developed to access the applicability of a DL-based mass density inference model for MCDC using anthropomorphic and porcine tissue phantoms. The dosimetry and WET map analyses indicated that the physics-informed DL model could deliver accurate material mass density maps for MCDC to achieve good agreement with the measured data regarding dose-difference, distance-to-agreement, and WET distributions.

The proposed framework has the potential to quantify the uncertainty of using DL models for proton treatment planning and *in vivo* proton range uncertainty due to the heterogeneity of patient anatomy.

Acknowledgments

This research is supported in part by the National Institutes of Health under Award Number R01CA215718 and R01EB032680. We also would like to thank Dr. Rongxiao Zhang for capturing imaging and proton dose measurement on the porcine tissue phantom.

Appendix A.: Hounsfield look-up table.

Table A1 shows the Hounsfield look-up table to convert SECT numbers to material mass densities.

Table A1.

Hounsfield look-up table.

CT number (Hounsfield unit)	Mass density (g/cm^3)	Relative stopping power
−1024	9.0×10^{-4}	9.0×10^{-4}
−980	1.0×10^{-3}	1.0×10^{-3}
−741	0.26	0.257
−707	0.30	0.288
−560	0.45	0.432
−93	0.94	0.891
−61	0.95	0.916
−48	0.98	0.929
−24	0.99	0.958
0	1.0	1.0
19	1.03	1.001
29	1.05	1.002
48	1.06	1.045
52	1.07	1.049
76	1.09	1.067
101	1.12	1.097
189	1.14	1.094
200	1.15	1.103
242	1.18	1.149
383	1.29	1.263
427	1.34	1.289
549	1.41	1.384
565	1.42	1.394
628	1.46	1.430
702	1.52	1.495
761	1.56	1.513
829	1.61	1.586

CT number (Hounsfield unit)	Mass density (g/cm^3)	Relative stopping power
923	1.68	1.656
1157	1.82	1.780
1260	1.92	1.898
2495	2.71	2.685

Appendix B.: Model structure of ResNet

Table B1 summarizes the model form and model parameters of ResNet.

Table B1.

Model form and model parameters of ResNet. Res., Conv., and ConvA/ConvB denote the residual, convolutional and convolutional layer A/B.

	Network	Layer	Number of channels	Kernel size	Stride	Padding
	ConvA/ConvB	**Conv.**	**64/64**	**7/3**	**2/1**	**3/0**
Res. Block	A1	Conv./Conv./Res.	64/64/64	3/3/1	2/1/2	1/1/0
	A2	Conv./Conv./Res.	128/128/128	3/3/1	2/1/2	1/1/0
	A3	Conv./Conv./Res.	256/256/256	3/3/1	2/1/2	1/1/0
Res. Block	B1	Conv./Conv./Res.	64/64/64	2/2/1	2/1/2	1/1/0
	B2	Conv./Conv./Res.	128/128/128	2/2/1	2/1/2	1/1/0
	B3	Conv./Conv./Res.	256/256/256	2/2/1	2/1/2	1/1/0

References

Almeida IP, Schyns LEJR, Öllers MC, van Elmpt W, Parodi K, Landry G and Verhaegen F 2017 Dual-energy CT quantitative imaging: a comparison study between twin-beam and dual-source CT scanners Medical Physics 44 171–9 [PubMed: 28070917]

Arjomandy B, Sahoo N, Ciangaru G, Zhu R, Song X and Gillin M 2010 Verification of patient-specific dose distributions in proton therapy using a commercial two-dimensional ion chamber array Medical Physics 37 5831–7 [PubMed: 21158295]

Arridge S, Maass P, Öktem O and Schönlieb C-B 2019 Solving inverse problems using data-driven models Acta Numerica 28 1–174

Baumann M, Krause M, Overgaard J, Debus J, Bentzen SM, Daartz J, Richter C, Zips D and Bortfeld T 2016 Radiation oncology in the era of precision medicine Nature Reviews Cancer 16 234–49 [PubMed: 27009394]

Blumer A, Ehrenfeucht A, Haussler D and Warmuth MK 1987 Occam's Razor Information Processing Letters 24 377–80

Bourque AE, Carrier J-F and Bouchard H 2014 A stoichiometric calibration method for dual energy computed tomography Physics in Medicine and Biology 59 2059–88 [PubMed: 24694786]

Champion K, Lusch B, Kutz JN and Brunton SL 2019 Data-driven discovery of coordinates and governing equations Proceedings of the National Academy of Sciences 116 22445

Chang C-W and Dinh NT 2019 Classification of machine learning frameworks for data-driven thermal fluid models International Journal of Thermal Sciences 135 559–79

Chang C-W, Gao Y, Wang Q, Lei Y, Wang T, Zhou J, Lin L, Bradley JD, Liu T and Yang X Proc.SPIE,2022a), vol. Series 12031)

Chang C-W, Gao Y, Wang T, Lei Y, Wang Q, Pan S, Sudhyadhom A, Bradley JD, Liu T, Lin L, Zhou J and Yang X 2022b Dual-energy CT based mass density and relative stopping power estimation for proton therapy using physics-informed deep learning Physics in Medicine & Biology 67 115010

Chang C-W, Huang S, Harms J, Zhou J, Zhang R, Dhabaan A, Slopsema R, Kang M, Liu T, McDonald M, Langen K and Lin L 2020 A standardized commissioning framework of Monte Carlo dose calculation algorithms for proton pencil beam scanning treatment planning systems Medical Physics 47 1545–57 [PubMed: 31945191]

Chang C-W, Lei Y, Charyyev S, Leng S, Yoon T, Zhou J, Yang X and Lin L Proc.SPIE,2022c), vol. Series 12034) p 120340X

Chang C-W, Marants R, Gao Y, Goette M, Scholey JE, Bradley JD, Liu T, Zhou J, Sudhyadhom A and Yang X 2022d Multimodal Imaging-based Material Mass Density Estimation for Proton Therapy Using Physics-Constrained Deep Learning arXiv preprint arXiv:2207.13150

Cormack AM 1963 Representation of a Function by Its Line Integrals, with Some Radiological Applications Journal of Applied Physics 34 2722–7

Deffet S, Farace P and Macq B 2020 Sparse deconvolution of proton radiography data to estimate water equivalent thickness maps Medical Physics 47 509–17 [PubMed: 31705805]

Domingos P 1999 The Role of Occam's Razor in Knowledge Discovery Data Mining and Knowledge Discovery 3 409–25

Farace P, Righetto R and Meijers A 2016 Pencil beam proton radiography using a multilayer ionization chamber Physics in Medicine and Biology 61 4078–87 [PubMed: 27164479]

He K, Zhang X, Ren S and Sun J 2016 Deep Residual Learning for Image Recognition 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 770–8

Hogstrom KR, Mills MD, Eyer JA, Palta JR, Mellenberg DE, Meoz RT and Fields RS 1984 Dosimetric evaluation of a pencil-beam algorithm for electrons employing a two-dimensional heterogeneity correction Int J Radiat Oncol Biol Phys 10 561–9 [PubMed: 6725043]

Hornik K, Stinchcombe M and White H 1989 Multilayer feedforward networks are universal approximators Neural Networks 2 359–66

Hsi WC, Moyers MF, Nichiporov D, Anferov V, Wolanski M, Allgower CE, Farr JB, Mascia AE and Schreuder AN 2009 Energy spectrum control for modulated proton beams Medical Physics 36 2297–308 [PubMed: 19610318]

Hünemohr N, Krauss B, Tremmel C, Ackermann B, Jäkel O and Greilich S 2014 Experimental verification of ion stopping power prediction from dual energy CT data in tissue surrogates Physics in Medicine and Biology 59 83–96 [PubMed: 24334601]

ICRP23 1975 Report on the Task Group on Reference Man ICRP Publication 23

ICRU42 1987 Use of Computers in External Beam Radiotherapy Procedures with High-Energy Photons and Electrons ICRU Publication 42

ICRU44 1989 Tissue Substitutes in Radiation Dosimetry and Measurement ICRU Publication 44

ICRU49 1993 Stopping Power and Ranges for Protons and Alpha Particles ICRU Publication 49

Jackson DF and Hawkes DJ 1981 X-ray attenuation coefficients of elements and mixtures Physics Reports 70 169–233

Karger CP, Jäkel O, Palmans H and Kanai T 2010 Dosimetry for ion beam radiotherapy Physics in Medicine and Biology 55 R193–R234 [PubMed: 20952816]

Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S and Yang L 2021 Physics-informed machine learning Nature Reviews Physics 3 422–40

Kerns JR, Followill DS, Lowenstein J, Molineu A, Alvarez P, Taylor PA and Kry SF 2016 Agreement Between Institutional Measurements and Treatment Planning System Calculations for Basic Dosimetric Parameters as Measured by the Imaging and Radiation Oncology Core-Houston Int J Radiat Oncol Biol Phys 95 1527–34 [PubMed: 27315667]

Knopf A-C and Lomax A 2013 In vivo proton range verification: a review Physics in Medicine and Biology 58 R131–R60 [PubMed: 23863203]

LeCun Y, Bengio Y and Hinton G 2015 Deep learning Nature 521 436–44 [PubMed: 26017442]

Low DA, Harms WB, Mutic S and Purdy JA 1998 A technique for the quantitative evaluation of dose distributions Medical Physics 25 656–61 [PubMed: 9608475]

Mah E, Antolak J, Scrimger JW and Battista JJ 1989 Experimental evaluation of a 2D and 3D electron pencil beam algorithm Physics in Medicine and Biology 34 1179–94

Mayneord W 1937 The significance of the roentgen Acta Int Union Against Cancer 2 271–82

Meyer E, Raupach R, Lell M, Schmidt B and Kachelrieß M 2010 Normalized metal artifact reduction (NMAR) in computed tomography Medical Physics 37 5482–93 [PubMed: 21089784]

Mumot M, Algranati C, Hartmann M, Schippers JM, Hug E and Lomax AJ 2010 Proton range verification using a range probe: definition of concept and initial analysis Physics in Medicine and Biology 55 4771–82 [PubMed: 20679697]

Nair V and Hinton GE 2010 Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning, (Haifa, Israel: Omnipress) pp 807–14

O'Sullivan F 1986 A Statistical Perspective on Ill-Posed Inverse Problems Statistical Science 1 502–18

Paganetti H 2012 Range uncertainties in proton therapy and the role of Monte Carlo simulations Physics in Medicine and Biology 57 R99–R117 [PubMed: 22571913]

Paganetti H 2018 Proton Therapy Physics: CRC Press)

Paganetti H, Jiang H, Parodi K, Slopsema R and Engelsman M 2008 Clinical implementation of full Monte Carlo dose calculation in proton beam therapy Physics in Medicine and Biology 53 4825–53 [PubMed: 18701772]

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J and Chintala S 2019 PyTorch: An Imperative Style, High-Performance Deep Learning Library Advances in Neural Information Processing Systems 32

RayPhysics 2020 RayStation 10B Reference Manual. (Stockholm, Sweden: RaySearch Laboratories AB)

Rinaldi I, Brons S, Jäkel O, Voss B and Parodi K 2014 Experimental investigations on carbon ion scanning radiography using a range telescope Physics in Medicine and Biology 59 3041–57 [PubMed: 24842455]

Schneider U, Besserer J, Pemler P, Dellert M, Moosburger M, Pedroni E and Kaser-Hotz B 2004 First proton radiography of an animal patient Medical Physics 31 1046–51 [PubMed: 15191291]

Schneider U, Pedroni E and Lomax A 1996 The calibration of CT Hounsfield units for radiotherapy treatment planning Physics in Medicine and Biology 41 111–24 [PubMed: 8685250]

Schneider W, Bortfeld T and Schlegel W 2000 Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions Physics in Medicine and Biology 45 459–78 [PubMed: 10701515]

Schuemann J, Dowdell S, Grassberger C, Min CH and Paganetti H 2014 Site-specific range uncertainties caused by dose calculation algorithms for proton therapy Physics in Medicine and Biology 59 4007–31 [PubMed: 24990623]

Spiers FW 1946 Effective Atomic Number and Energy Absorption in Tissues The British Journal of Radiology 19 52–63 [PubMed: 21015391]

Su K-H, Kuo J-W, Jordan DW, Van Hedent S, Klahr P, Wei Z, Al Helo R, Liang F, Qian P, Pereira GC, Rassouli N, Gilkeson RC, Traughber BJ, Cheng C-W and Muzic RF 2018 Machine learning-based dual-energy CT parametric mapping Physics in Medicine & Biology 63 125001 [PubMed: 29787382]

Taylor PA, Kry SF, Alvarez P, Keith T, Lujano C, Hernandez N and Followill DS 2016 Results From the Imaging and Radiation Oncology Core Houston's Anthropomorphic Phantoms Used for Proton Therapy Clinical Trial Credentialing International Journal of Radiation Oncology*Biology*Physics 95 242–8 [PubMed: 27084644]

Taylor PA, Kry SF and Followill DS 2017 Pencil Beam Algorithms Are Unsuitable for Proton Dose Calculations in Lung Int J Radiat Oncol Biol Phys 99 750–6 [PubMed: 28843371]

van de Water TA, Lomax AJ, Bijl HP, de Jong ME, Schilstra C, Hug EB and Langendijk JA 2011 Potential Benefits of Scanned Intensity-Modulated Proton Therapy Versus Advanced Photon Therapy With Regard to Sparing of the Salivary Glands in Oropharyngeal Cancer Int J Radiat Oncol Biol Phys 79 1216–24 [PubMed: 20732761]

van Elmpt W, Landry G, Das M and Verhaegen F 2016 Dual energy CT in radiotherapy: Current applications and future outlook Radiotherapy and Oncology 119 137–44 [PubMed: 26975241]

Wang F, Han J, Zhang S, He X and Huang D 2018 CSI-Net: Unified human body characterization and pose recognition arXiv preprint arXiv:1810.03064

Wellenberg RHH, Donders JCE, Kloen P, Beenen LFM, Kleipool RP, Maas M and Streekstra GJ 2018a Exploring metal artifact reduction using dual-energy CT with pre-metal and post-metal implant cadaver comparison: are implant specific protocols needed? Skeletal Radiology 47 839–45 [PubMed: 28842739]

Wellenberg RHH, Hakvoort ET, Slump CH, Boomsma MF, Maas M and Streekstra GJ 2018b Metal artifact reduction techniques in musculoskeletal CT-imaging European Journal of Radiology 107 60–9 [PubMed: 30292274]

Wohlfahrt P, Möhler C, Enghardt W, Krause M, Kunath D, Menkel S, Troost EGC, Greilich S and Richter C 2020 Refinement of the Hounsfield look-up table by retrospective application of patient-specific direct proton stopping-power prediction from dual-energy CT Medical Physics 47 1796–806 [PubMed: 32037543]

Wohlfahrt P, Möhler C, Hietschold V, Menkel S, Greilich S, Krause M, Baumann M, Enghardt W and Richter C 2017 Clinical Implementation of Dual-energy CT for Proton Treatment Planning on Pseudo-monoenergetic CT scans International Journal of Radiation Oncology*Biology*Physics 97 427–34 [PubMed: 28068248]

Zhang R, Taddei PJ, Fitzek MM and Newhauser WD 2010 Water equivalent thickness values of materials used in beams of protons, helium, carbon and iron ions Physics in Medicine and Biology 55 2481–93 [PubMed: 20371908]

Zhou S, Rao W, Chen Q, Tan Y, Smith W, Sun B, Zhou J, Chang C-W, Lin L, Darafsheh A, Zhao T and Zhang T 2022 A multi-layer strip ionization chamber (MLSIC) device for proton pencil beam scan quality assurance Physics in Medicine & Biology

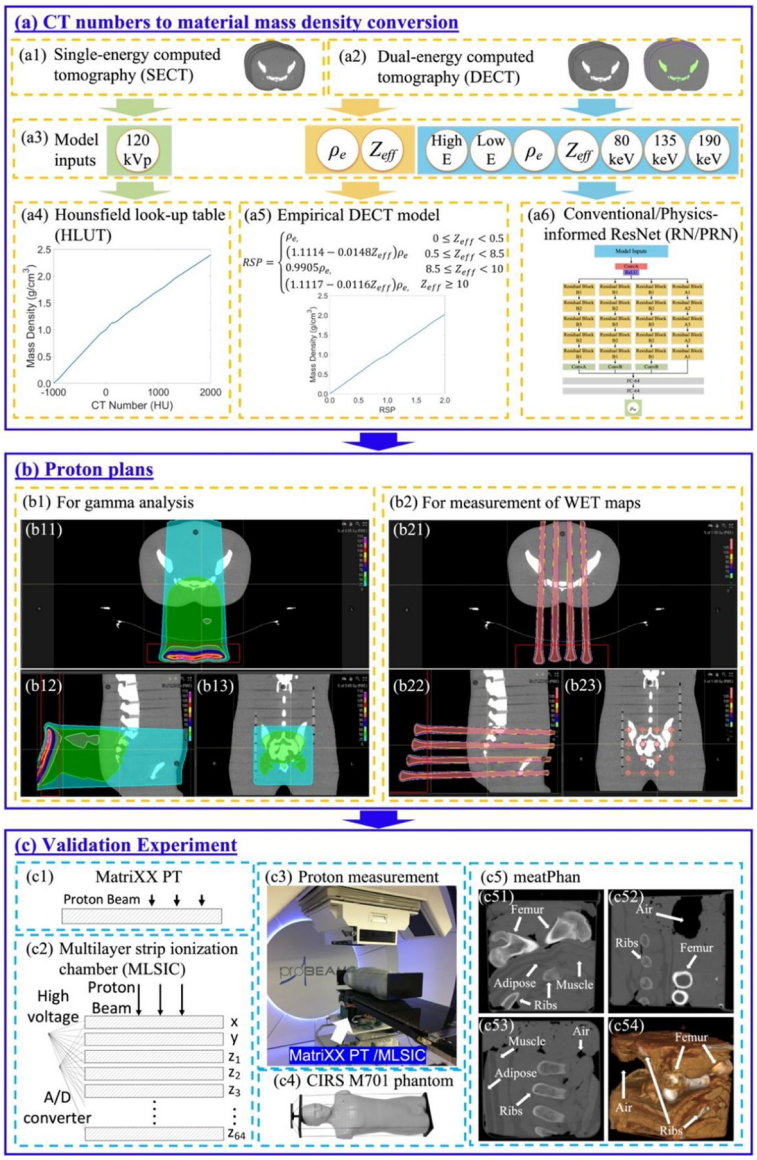

Figure 1.

Validation framework for Monte Carlo dose calculation using physics-informed deep learning models, including three steps: (a) CT to material mass density conversion using SECT with (a4) HLUT and DECT with (a5) the empirical and (a6) conventional/physics-informed ResNet models; (b) proton plans for (b1) dosimetry and (b2) WET measurement using (b11)-(b13) single field proton beams and (b21)-(b23) proton beams with sparse spots; (c) validation experiment using (c1) MatriXX PT and (c2) MLSIC to measure proton beams, penetrating through (c4) anthropomorphic and (c5) porcine tissue phantoms.

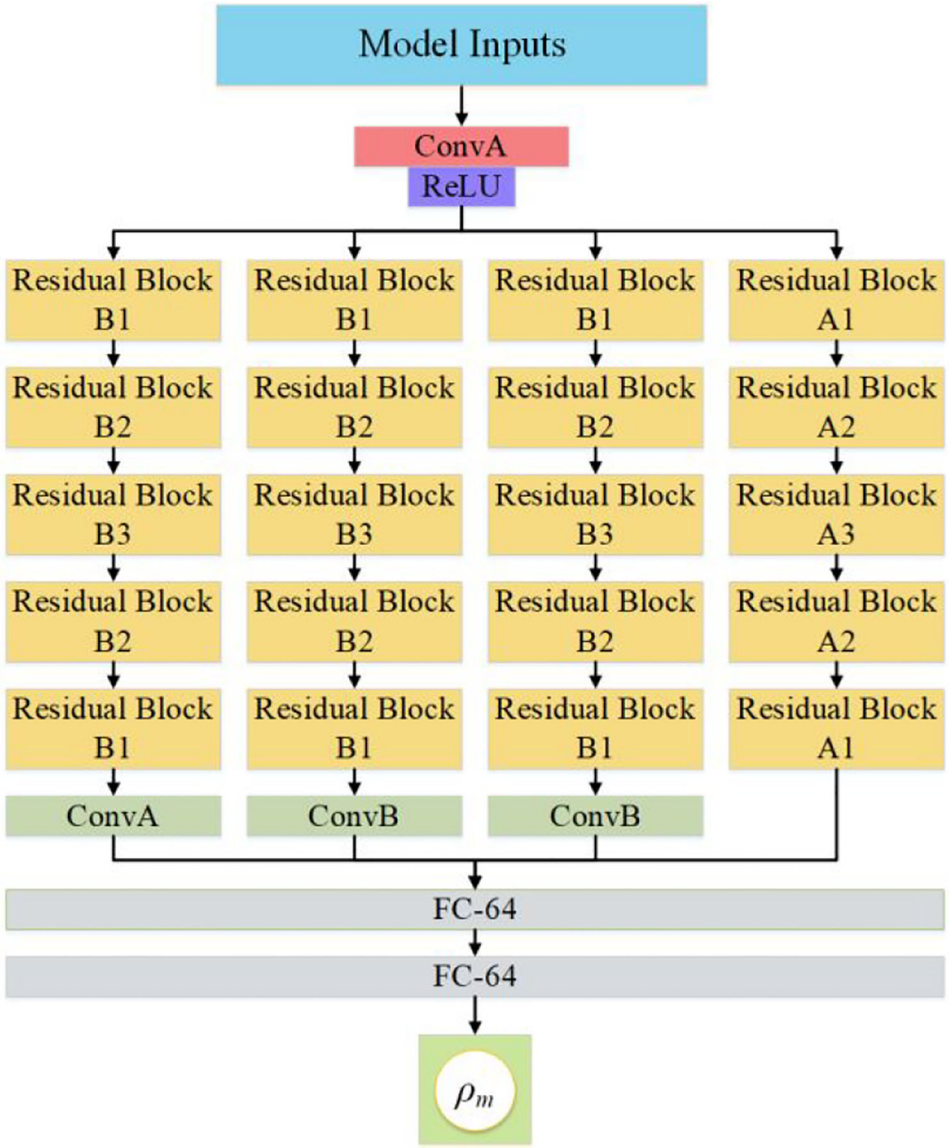

Figure 2.

Model structure of the residual network (ResNet). ConvA and ConvB denote two distinct convolutional layer A and B defined in Table B1. FC is the fully connected layer with 64 hidden units.

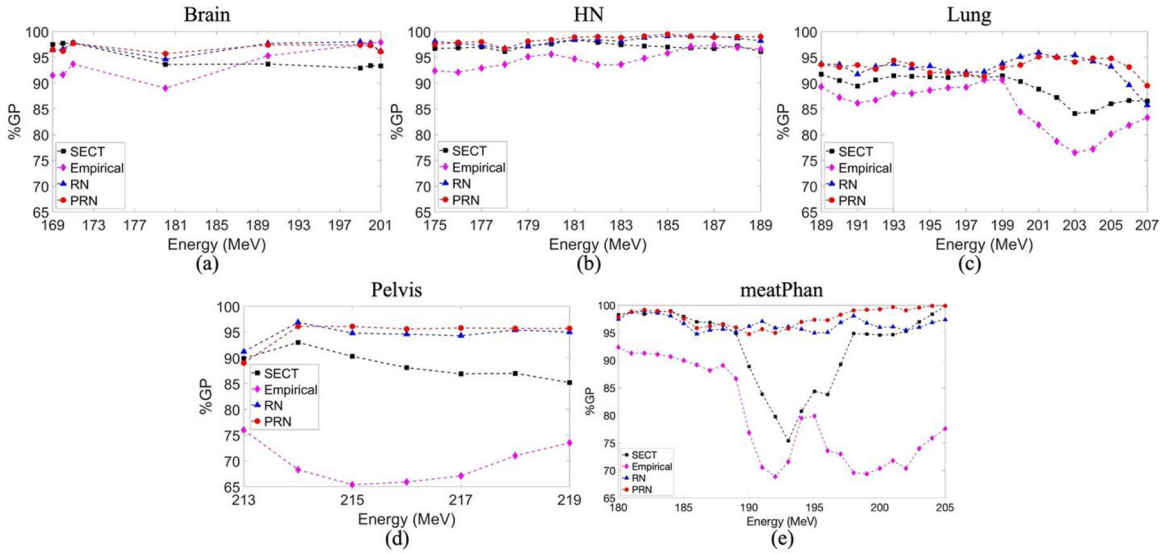

Figure 3.

Variation of gamma passing rates (%GP) at different energies of the CIRS adult male phantom at (a) brain, (b) head-and-neck (HN), (c) lung, and (d) pelvic sites and the porcine tissue phantom (meatPhan).

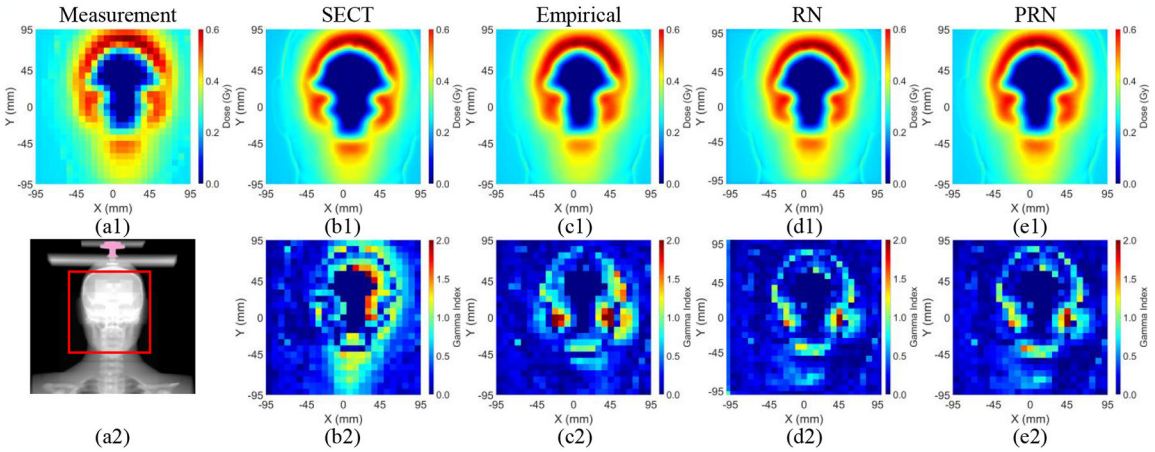

Figure 4.

Dose distributions by (a1) measurement and Monte Carlo dose calculation (MCDC) using images from (b1) SECT with a Hounsfield look-up table (HLUT) and DECT with (c1) the empirical model, (d1) RN, and (e1) PRN for the anthropomorphic phantom at a brain site with a 190 MeV proton beam. (a2) Digitally reconstructed radiograph of the phantom anatomy with a red box to denote the beam field size. (b2)-(e2) Comparisons of dose distributions using gamma index between measurement and MCDC with different images as described in (b1)-(e1). The anthropomorphic phantom was irradiated with an anterior beam and the exit dose was measured with MatriXX PT (IBA Dosimetry, Germany).

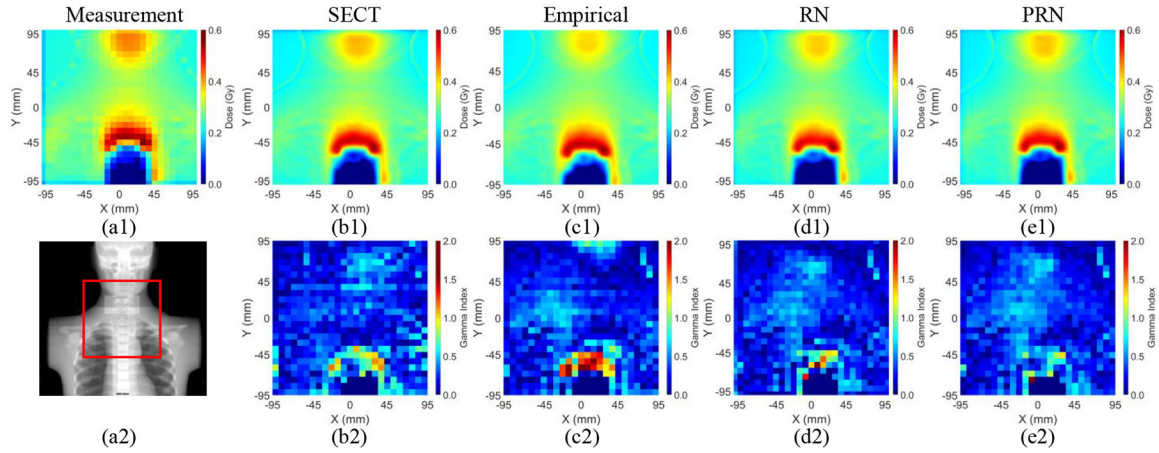

Figure 5.

Dose distributions by (a1) measurement and Monte Carlo dose calculation (MCDC) using images from (b1) SECT with a Hounsfield look-up table (HLUT) and DECT with (c1) the empirical model, (d1) RN, and (e1) PRN for the anthropomorphic phantom at a head-and-neck (HN) site with a 188 MeV proton beam. (a2) Digitally reconstructed radiograph of the phantom anatomy with a red box to denote the beam field size. (b2)-(e2) Comparisons of dose distributions using gamma index between measurement and MCDC with different images as described in (b1)-(e1). The anthropomorphic phantom was irradiated with an anterior beam and the exit dose was measured with MatriXX PT (IBA Dosimetry, Germany).

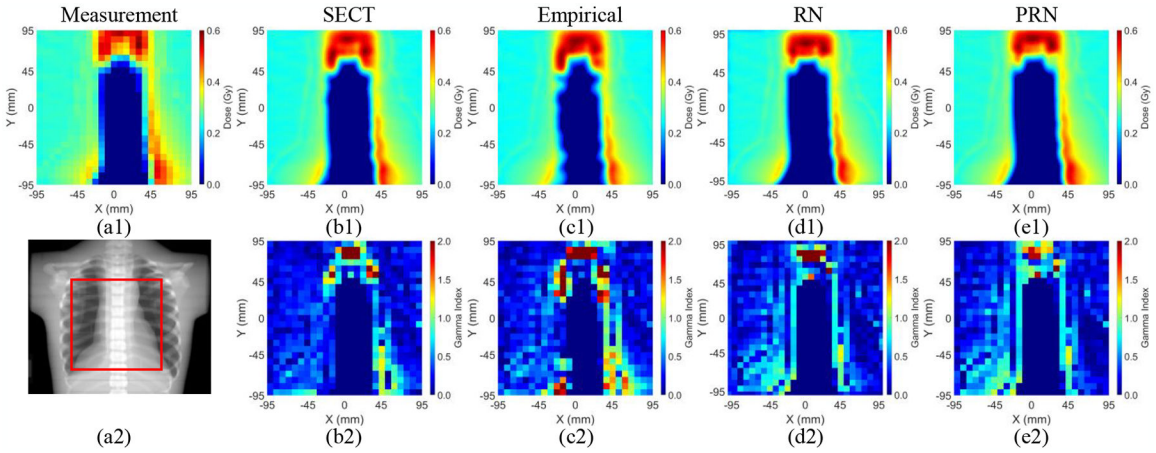

Figure 6.

Dose distributions by (a1) measurement and Monte Carlo dose calculation (MCDC) using images from (b1) SECT with a Hounsfield look-up table (HLUT) and DECT with (c1) the empirical model, (d1) RN, and (e1) PRN for the anthropomorphic phantom at a thoracic site with a 199 MeV proton beam. (a2) Digitally reconstructed radiograph of the phantom anatomy with a red box to denote the beam field size. (b2)-(e2) Comparisons of dose distributions using gamma index between measurement and MCDC with different images as described in (b1)-(e1). The anthropomorphic phantom was irradiated with an anterior beam and the exit dose was measured with MatriXX PT (IBA Dosimetry, Germany).

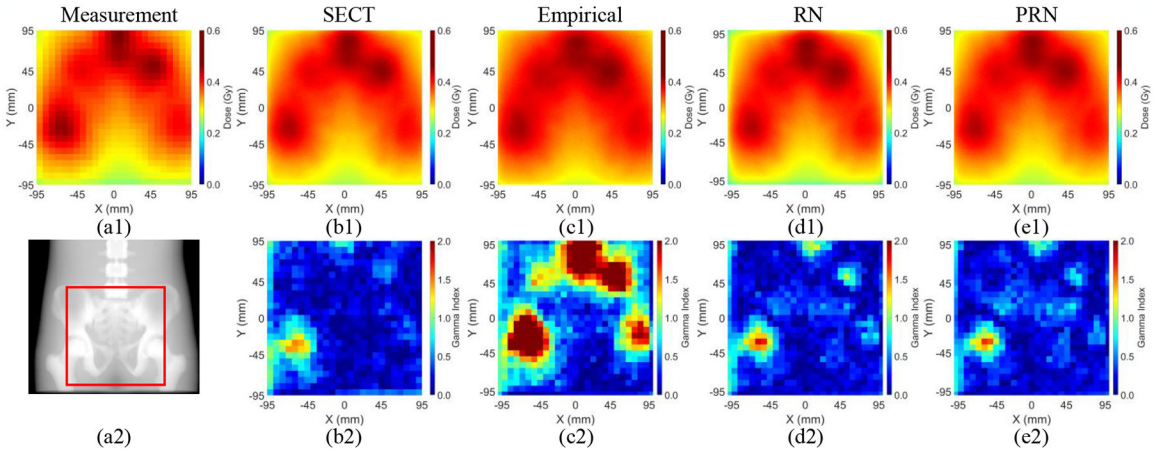

Figure 7.

Dose distributions by (a1) measurement and Monte Carlo dose calculation (MCDC) using images from (b1) SECT with a Hounsfield look-up table (HLUT) and DECT with (c1) the empirical model, (d1) RN, and (e1) PRN for the anthropomorphic phantom at a pelvic site with a 219 MeV proton beam. (a2) Digitally reconstructed radiograph of the phantom anatomy with a red box to denote the beam field size. (b2)-(e2) Comparisons of dose distributions using gamma index between measurement and MCDC with different images as described in (b1)-(e1). The anthropomorphic phantom was irradiated with an anterior beam and the exit dose was measured with MatriXX PT (IBA Dosimetry, Germany).

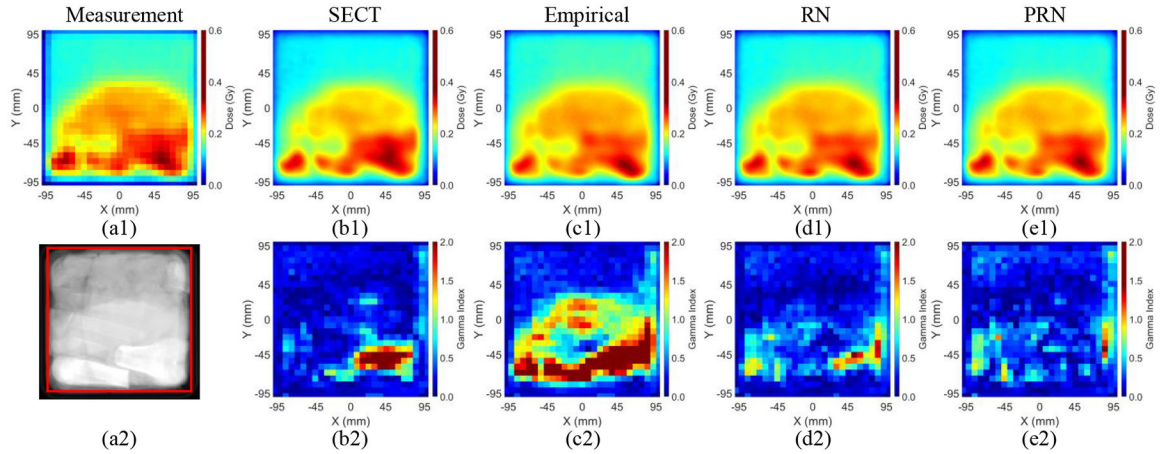

Figure 8.

Dose distributions by (a1) measurement and Monte Carlo dose calculation (MCDC) using images from (b1) SECT with a Hounsfield look-up table (HLUT) and DECT with (c1) the empirical model, (d1) RN, and (e1) PRN for the porcine tissue phantom (meatPhan) with a 200 MeV proton beam. (a2) Digitally reconstructed radiograph of the phantom anatomy with a red box to denote the beam field size. (b2)-(e2) Comparisons of dose distributions using gamma index between measurement and MCDC with different images as described in (b1)-(e1). The meatPhan was irradiated with an anterior beam and the exit dose was measured with MatriXX PT (IBA Dosimetry, Germany).

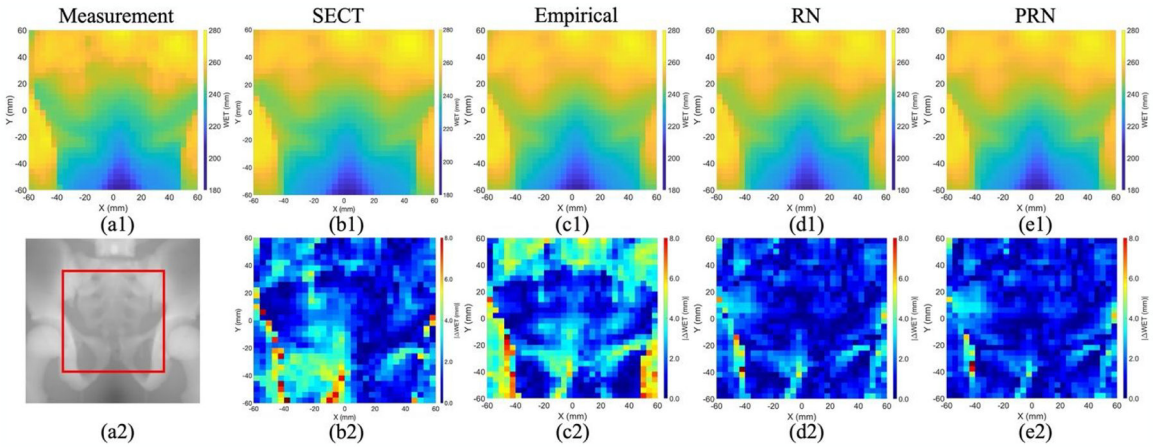

Figure 9.

Water equivalent thickness (WET) maps by (a1) measurement and Monte Carlo dose calculation (MCDC) using images from (b1) SECT with a Hounsfield look-up table (HLUT) and DECT with (c1) the empirical model, (d1) RN, and (e1) PRN for the anthropomorphic phantom at a pelvic site with a 216 MeV proton beam. (a2) Digitally reconstructed radiograph of the phantom anatomy with a red box to denote the beam field size. (b2)-(e2) Comparisons of the WET absolute difference (|ΔWET|) between measurement and MCDC with different images described as (b1)-(e1). The anthropomorphic phantom was irradiated with anterior beams and the proton residual ranges were measured with MLSIC for WET analyses.

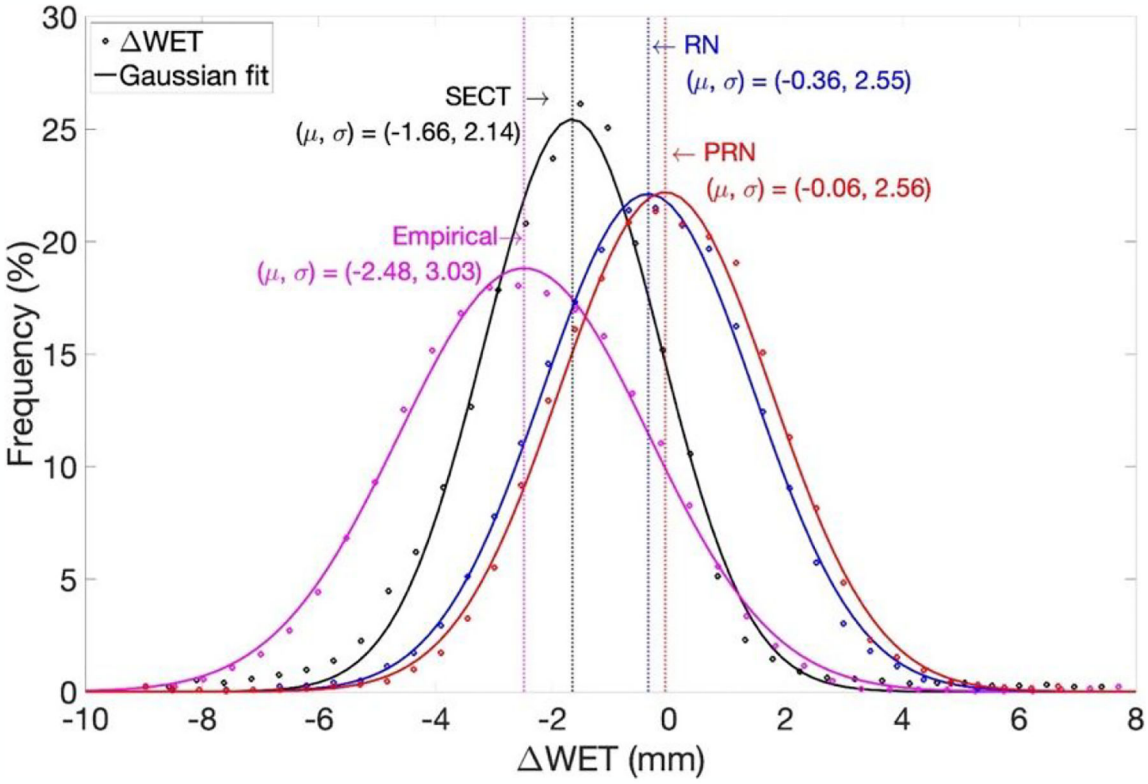

Figure 10.

Distribution of WET variation (ΔWET) by Eq. (7) between the measurement (Meas.) and simulation (Sim.) by each model. The solid lines are Gaussian fitted ΔWET data from each model with different means (μ) and standard deviations (σ).

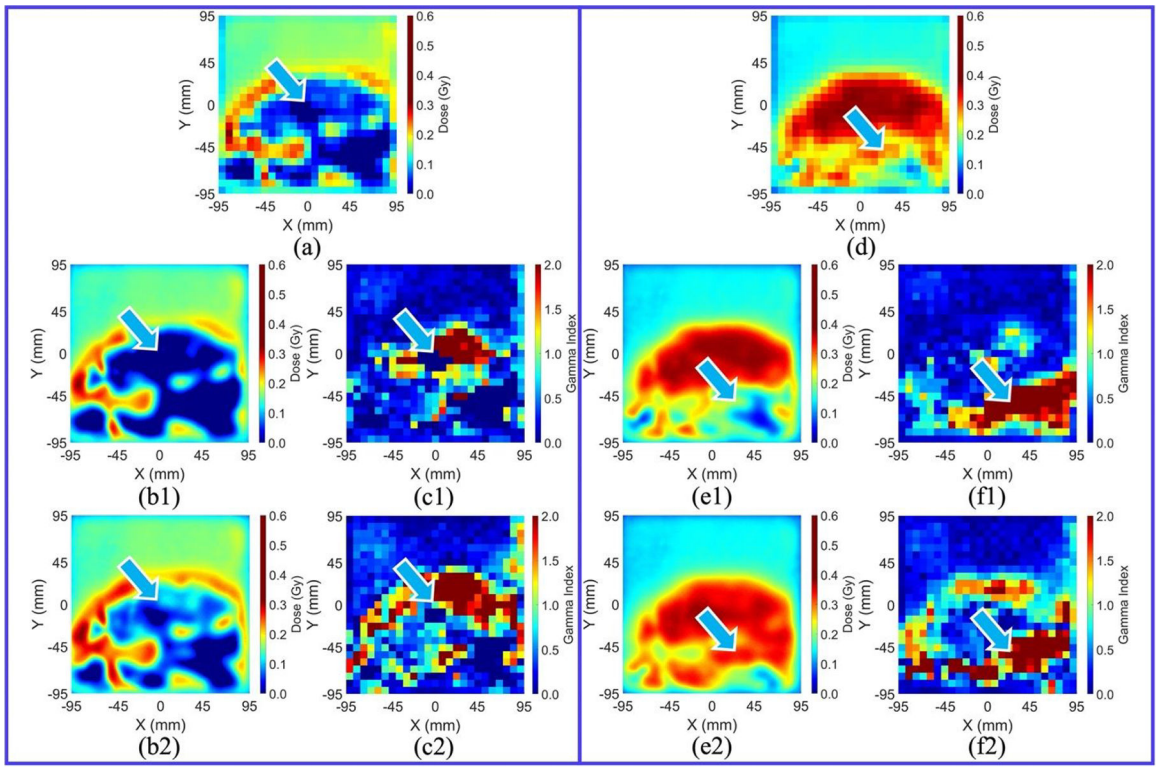

Figure 11.

Measured dose distribution for the porcine phantom (meatPhan) using (a) 190 MeV and (b) 195 MeV proton beams. Monte Carlo dose calculation using images from (b1) SECT with a Hounsfield look-up table (HLUT) and (b2) the empirical model. (c1)-(c2) Corresponding gamma index maps to (b1)-(b2). Monte Carlo dose calculation using images from (e1) SECT with HLUT and (e2) the empirical model. (f1)-(f2) Corresponding gamma index maps to (e1)-(e2).

Table 1.

CT acquisition parameters.

		CIRS anthropomorphic phantom	Porcine tissue phantom (meatPhan)
Slice thickness		1.5	1.0
Field of view		500 mm	
Voxel size		0.977×0.977 mm^2	
Collimation		64×0.6 mm	
SECT	$CTDI_{vol,\ 32\ cm}$	26.5 mGy	23.2 mGy
	mAs_{eff}	393 mAs	344 mAs
	X-ray tube voltage	120 kVp	
DECT	$CTDI_{vol,\ 32\ cm}$	8.5 mGy	9.1 mGy
	mAs_{eff}	397 mAs	424 mAs
	X-ray tube voltage	120 kVp with Au/Sn filters	

Table 2.

Basic materials for interpolating 50 fixed materials used by RayStation proton Monte Carlo dose calculation. Information including mass density, mean ionization energy (*I*), elemental composition in weight percent (%).

	Density (g/cm^3)	*I* (eV)	H	C	N	O	Na	Mg	Al	P	S	Cl	Ar	K	Ca
Air	0.001	85.7			75.5	23.2							1.3		
Lung	0.26	75.3	10.3	10.5	3.1	74.9	0.2			0.2	0.3	0.3		0.2	
Adipose	0.95	63.2	11.4	59.8	0.7	27.8	0.1				0.1	0.1			
Muscle	1.05	74.7	10.2	14.3	3.4	71	0.1			0.2	0.3	0.1		0.4	
Cartilage	1.10	75.0	9.6	9.9	2.2	74.4	0.5			2.2	0.9	0.3			
Bone1	1.85	106.4	4.7	14.5	4.2	44.6		0.2		10.5	0.3				21
Bone2	2.10	106.4	4.7	14.5	4.2	44.6		0.2		10.5	0.3				21
Aluminum	2.70	166							100						

Table 3.

Proton planning parameters for measurements, and information of the Monte Carlo dose calculation (MCDC).

Proton plan type	**Dosimetry measurement**					**WET measurement**
Phantom	CIRS M701	CIRS M701	CIRS M701	CIRS M701	meatPhan	CIRS M701
Site	Brain	HN	Lung	Pelvis	-	Pelvis
Energy (MeV)	169–201	175–189	189–207	200–213	180–205	216
Field size (mm^2)	200×200	200×200	200×200	200×200	180×180	120×120
Spot spacing (mm)	4	4	4	4	4	4
Number of proton spots	2601	2601	2601	2601	2116	961
Number of simulated particles	3.2×10^9	5.7×10^9	7.9×10^9	7.0×10^9	8.9×10^9	3.6×10^9
Dose voxel size[+] (mm^3)	2×1×2	2×1×2	2×1×2	2×1×2	2×1×2	2×1×2
MCDC time (min)	5.2	14.6	19.4	16.4	20.3	8.5

[+](Right-Left) × (Posterior-Anterior) × (Inferior-Superior)

Table 4.

Means and standard deviations of gamma passing rates from different models.

	SECT	Empirical	RN	PRN
Brain	95.0% ± 2.2%	94.3% ± 3.4%	96.8% ± 1.1%	**96.8% ± 0.8%**
HN	97.1% ± 0.6%	94.8% ± 1.8%	98.2% ± 0.7%	**98.5% ± 0.7%**
Lung	89.2% ± 2.6%	85.1% ± 4.6%	93.0% ± 2.3%	**93.2% ± 1.4%**
Pelvis	88.6% ± 2.6%	69.6% ± 4.0%	94.6% ± 1.7%	**94.9% ± 2.6%**
meatPhan	92.8% ± 7.1%	79.7% ± 8.8%	96.6% ± 1.2%	**97.8% ± 1.6%**