“Artificial Intelligence supported single detector multi-energy proton radiography system”

Brent van der Heyden¹, Marie Cohilis², Kevin Souris², Luana de Freitas Nascimento³, Edmond Sterpin^{1,2,*}

(1) KU Leuven, Department of Oncology, Laboratory of Experimental Radiotherapy, Leuven, Belgium

(2) UCLouvain, Institut de recherche expérimentale et clinique, Molecular Imaging Radiotherapy and Oncology Lab, Brussels, Belgium

(3) Research in Dosimetric Applications, Belgian Nuclear Research Centre, Mol, Belgium

(*) **Corresponding author:**

Dr. Edmond Sterpin

edmond.sterpin@kuleuven.be

Figures 7

Tables 1

Equations 4

Running title: Artificial Intelligence supported compact proton radiography

Keywords: proton radiography, artificial intelligence, deep learning, ProtonDSE

Abstract (274/300 words)

Proton radiography imaging was proposed as a promising technique to evaluate internal anatomical changes, to enable pre-treatment patient alignment, and most importantly, to optimize the patient specific CT number to stopping-power ratio (SPR) conversion. The clinical implementation rate of proton radiography systems is still limited due to their complex bulky design, together with the persistent problem of (in)elastic nuclear interactions and multiple Coulomb scattering (i.e. range mixing). In this work, a compact multi-energy proton radiography system was proposed in combination with an Artificial Intelligence network architecture (ProtonDSE) to remove the persistent problem of proton scatter in proton radiography. A realistic Monte Carlo model of the Proteus[®]One accelerator was built at 200 MeV and 220 MeV to isolate the scattered proton signal in 236 proton radiographies of 80 digital anthropomorphic phantoms. ProtonDSE was trained to predict the proton scatter distribution at two beam energies in a 60%/25%/15% scheme for training, testing, and validation. A calibration procedure was proposed to derive the Water Equivalent Thickness image based on the detector dose response relationship at both beam energies. ProtonDSE network performance was evaluated with quantitative metrics that showed an overall mean absolute percentage error below $1.4 \pm 0.4\%$ in our test dataset. For one example patient, detector dose to WET conversions were performed based on the total dose (I_{Total}), the primary proton dose (I_{Primary}), and the ProtonDSE corrected detector dose ($I_{\text{Corrected}}$). The determined WET accuracy was compared with respect to the reference WET by idealistic raytracing in a manually delineated region-of-interest inside the brain. The error was determined $4.3 \pm 4.1\%$ for WET(I_{Total}), $2.2 \pm 1.4\%$ for WET(I_{Primary}), and $2.5 \pm 2.0\%$ for WET($I_{\text{Corrected}}$).

1. Introduction

A growing number of radiotherapy clinics worldwide is expanding patient care with proton therapy due to its advantageous particle characteristics compared to conventional megavoltage photons. Proton therapy takes advantage of the finite proton range and the so-called Bragg peak, which is characterized by a sharp distal dose fall-off (Newhauser and Zhang, 2015). The planning of proton treatments in cancer patients relies on the accurate determination of stopping-power ratios (SPR) derived from X-ray computed tomography (CT) scans (Almeida *et al.*, 2018). The ambiguous conversion from CT numbers (Hounsfield Units), a measure of photon attenuation and scatter, to SPR introduces conversion-related uncertainties. These conversion-related uncertainties translate into range uncertainties that are currently managed by robust planning, either by beam specific margins or by robust optimization (Unkelbach and Paganetti, 2018). However, robust planning increases volumes receiving high dose of radiation, which is detrimental for healthy organs-at-risk which should be spared as much as possible to reduce radiation-induced side-effects (McGowan *et al.*, 2013). Innovative methods facilitating the decrease in clinical range uncertainties are desired to enable more conformal treatment plans, and conceivably improve the treatment outcome.

The current magnitude of range uncertainties for pre-treatment SPR prediction has been estimated to be approximately 3% for single-energy CT based planning (Paganetti, 2012; Yang *et al.*, 2012; van der Heyden *et al.*, 2019a), although more sophisticated approaches such as dual-energy CT (Li *et al.*, 2017; Wohlfahrt *et al.*, 2017; Wohlfahrt *et al.*, 2018) or photon-counting CT (Taasti *et al.*, 2018) have shown potential to decrease planning induced uncertainties below 3%. Range uncertainties arising from the CT to patient specific SPR conversion (Wohlfahrt *et al.*, 2019) could probably be diminished by direct or indirect patient imaging with higher energy protons in analogy with X-ray imaging, i.e. proton radiography (Poludniowski *et al.*, 2015). Proton imaging ideally measures the water equivalent thickness (WET) and is equivalent to the SPRs integrated along the proton beam path. Imaging approaches such as proton imaging have the potential to either eliminate (i.e. proton CT) (Krah *et al.*, 2020) or optimize (i.e. proton radiography) the patient-specific CT to SPR conversion (Doolan *et al.*, 2015; Krah *et al.*, 2019b; Deffet *et al.*, 2020).

Earlier, systems have been proposed to perform proton radiography or proton CT (Poludniowski *et al.*, 2015; Johnson, 2018), but the proposed combinational use of proton-tracking equipment with a residual energy-range detector is rather complex and bulky, which makes the technique difficult to deploy in a clinical setting (Meijers *et al.*, 2020). More recent, research groups proposed more compact detector concepts, such as a high-resolution radiation-hardened X-ray flat-panel detector to measure energy resolved dose measurements in a large area (Bentefour el *et al.*, 2016), and to perform time-resolved proton

1 radiography with a rotating range modulator wheel (Jee *et al.*, 2017; Zhang *et al.*, 2017a;
2 Zhang *et al.*, 2017b). (Huo *et al.*, 2019) and (Zhang *et al.*, 2019) have investigated the use of
3 multiple proton energies to determine the WET with a compact single-detector radiography
4 system. Although the WET determination accuracy of single-detector radiography systems is
5 often lower compared to more complex proton range telescope solutions, it is may
6 reasonable to assume that simpler, faster, and more cost-effective proton radiography
7 systems could be sufficient in a clinical workflow, when the device meets the clinical
8 requirements.

9
10 The accuracy of proton radiography systems is often limited by range mixing in the
11 therapeutic energy range; multiple Coulomb scatter interactions (Schneider and Pedroni,
12 1994) inside the patient make protons deviate in small angles instead of moving along straight
13 tracks. Range mixing is still a persistent challenge in proton radiography and proton CT
14 (Khellaf *et al.*, 2020). Previous studies have already investigated computationally intensive
15 most-likely proton path corrections to improve the WET determination accuracy (Li *et al.*,
16 2006; Krah *et al.*, 2019a; Brooke and Penfold, 2020), but still need improvement in certain
17 geometries that contain significant heterogeneities in the beam path, such as a patient
18 (Brooke and Penfold, 2020).

19
20 Artificial Intelligence (AI) has already been demonstrated as a very powerful tool to
21 recognize and predict image patterns in patient imaging (Yang *et al.*, 2018; van der Heyden *et al.*,
22 2019b; Vandewinckele *et al.*, 2020). To our best knowledge, the use of AI algorithms has
23 not yet been investigated to tackle the problem of range mixing and proton scatter, which
24 limits the accuracy of proton radiography. However, U-Net and deep-convolutional
25 autoencoder network architectures have recently been adopted as possible solution to
26 remove X-ray scatter distributions from measured X-ray projections (Hansen *et al.*, 2018;
27 Maier *et al.*, 2019; van der Heyden *et al.*, 2020; Lalonde *et al.*, 2020). All studies showed
28 superior Hounsfield Unit determination and significantly improved image quality (i.e. reduced
29 X-ray beam hardening) compared to heuristic X-ray scatter correction methods. In analogy
30 with X-ray scatter correction algorithms, we propose a U-Net architecture, trained on realistic
31 Monte Carlo simulation data, to correct for (in)elastic scattered protons and multiple
32 Coulomb scattering in single-detector proton radiography.

33
34 In this work, we propose a compact multi-energy single-detector proton radiography
35 design with clinical implementation opportunities, to calculate WETs by measuring energy-
36 specific dose function with a flat-panel scintillation detector. In addition, we developed a
37 simulation-based deep proton scatter estimation (ProtonDSE) AI-model with the goal to
38 improve the WET calculation accuracy in single-detector proton radiography.

2. Materials and Methods

2.1 Monte Carlo simulation model

A single detector proton radiography Monte Carlo model was developed in TOPAS (Version 3.5, built on Geant4.10.06.p1) for proton imaging by pencil beam scanning. Machine output measurements were performed at the Proteus[®]One (IBA Particle Therapy, Louvain-la-Neuve, Belgium) accelerator housed in the PARTICLE Proton Therapy Center of Leuven. The Monte Carlo simulation output was validated with single pencil beam profiles measured at two nominal proton energies: 200 MeV and 220 MeV. The integrated depth doses were measured using a large air-vented plane-parallel StingRay[®] ionization chamber (IBA Dosimetry, Schwarzenbruck, Germany) with a diameter of 120 mm. In-air spot profile measurements were performed with a scintillator-based Lynx[®] detector (IBA Dosimetry, Schwarzenbruck, Germany) at the machine isocenter with an effective 0.5 mm resolution. The “g4em-standard_opt3”, “g4stopping”, “g4decay”, “g4h-elastic_HP”, “g4ion-binarycascade”, and “g4h_phy_QGSP_BIC_HP” physics lists were adopted in the TOPAS model for proton radiography simulations. The lower-edge particle production cut was set to 5 keV, and the source to isocenter distance was set to 455 mm. A TOPAS model was developed for research purposes only.

A flat-panel detector with an active area of 41 x 41 cm² was simulated in TOPAS as imaging detector. The realistic detector model included a protective entrance carbon plate of ~1 mm, a small air gap of ~3 mm, a 600 µm CsI scintillator, a 1 mm glass substrate (for the photodiodes), and an aluminum back plate of 1.5 mm. The protective entrance plate of the flat-panel detector was positioned 200 mm away from the machine isocenter. Although submillimeter pixel sizes (< 0.4 mm) exist in flat-panel technology, 1 x 1 mm² pixel sizes are simulated to increase calculation efficacy of our Monte Carlo model. A dedicated dose scorer was developed as TOPAS extension and attached to the CsI scintillation layer. The dedicated dose scorer distinguished two separate dose outputs: (i) the total proton dose, and (ii) the dose by protons that entered the CsI scintillator with a polar angle that deviated less than 2.5 degrees from its original momentum at the start of the Monte Carlo run. The constraint of 2.5 degrees was determined empirically to keep reasonable calculation times with sufficient simulation statistics. In this work, a primary proton is denoted as a proton that deviated less than 2.5 degrees from its initial direction. The second category (< 2.5 deg) intends to eliminate contaminating detector dose by protons that underwent nuclear (in)elastic scattering or that underwent multiple Coulomb scattering inside the medium and exited the medium at a large angle compared to the initial direction of the proton pencil beam. Monte Carlo simulation speed was improved by disabling the generation of secondary particles different from protons inside the patient, but the generation and transfer of secondary electrons was enabled inside

the scintillation layer of the flat-panel detector. All calculations were performed on the Tier2 high-performance supercomputer of the Flemish supercomputer center.

2.2 System calibration measurements

A 10 x 10 cm² proton imaging field at isocenter was simulated for two nominal beam energies (200 MeV and 220 MeV) to construct the dose calibration database in function of WET for the total dose (first category), and the primary dose (second category, < 2.5 deg). The spot grid of the proton imaging field was composed of 2601 realistic proton pencil beams (3E5 histories/spot to keep statistical uncertainty below 1%) with 2 mm spacing at the isocenter. The flat-panel detector dose was scored in all dose categories with varying PMMA material thickness from 0 mm to 260 mm in 10 mm steps. The material thickness resolution was increased to 5 mm near the Bragg peak. PMMA material thickness was converted to WET from its known chemical composition (C₅O₂H₈) and density (1.19 g/cm³) with the Bethe-Bloch equation and the Bragg additivity rule. PMMA was adopted as solid affordable material with known chemical composition because it is composed out of three major chemical elements found in soft tissues (carbon, oxygen, and hydrogen) (White *et al.*, 1987; Woodard and White, 1986). The mean ionization potential of water was set to 78 eV (Faddegon *et al.*, 2015), and 200 MeV was considered as reference kinetic proton energy in the equations.

An energy dependent WET-to-dose look-up table for calibration purposes was constructed at 0.1 mm WET resolution by applying a piecewise cubic Hermite polynomial interpolating technique (PCHIP) between the measurement points. Two look-up tables were composed per imaging energy; one for the total dose, and one for the primary dose.

2.3 ProtonDSE - network architecture

The Monte Carlo simulation model scored both the total dose and the primary dose. Here, a conventional two-dimensional U-Net architecture is adopted to predict the scattered proton contribution (> 2.5 deg). Our Proton Deep Scatter Estimation (ProtonDSE) network architecture (Figure 1) consists of 7 layers, existing of a contracting and an expansive path. The output of the contraction layer is concatenated in the channel dimension with the input of expansion layer.

Figure 1. Proton Deep Scatter Estimation (ProtonDSE) network architecture

Every block of the contracting path applies successive 3 x 3 convolution layers followed by a rectified linear unit (ReLU). The spatial resolution in every dimension is halved every block by applying 2 x 2 max pooling layer. The number of feature maps after each block doubles along the contraction path from 16 to 1024. The expansive path is symmetrical to the contracting path, with the exception that the convolution of each block is replaced by a bilinear upsampling layer. Adam (Kingma and Lei, 2015) was employed as stochastic training optimizer with an initial learning rate of 1E-5 and parametrized with $\beta_1 = 0.900$, $\beta_2 = 0.999$. The fixed batch size was set to 16, the number of epochs to 250, and the mean absolute error (MAE) was adopted as loss function to train the convolutional neural network. No batch-normalization was applied to prevent unwanted scaling of the scattered proton distribution. Individual network training was performed for 200 MeV and 220 MeV on the Google Cloud Platform, in a virtual machine with an Intel® Xeon® CPU @ 2.20GHz and a 16 GB T4 NVIDIA® GPU. ProtonDSE_{200MeV} and ProtonDSE_{220MeV} were implemented in Keras (v. 2.4.3) with Tensorflow as backend (v. 2.2.0).

2.4 Water Equivalent Thickness (WET) determination

The WET in every flat-panel detector pixel is determined according to Equation 1 with a similar method described by (Huo *et al.*, 2019).

$$\text{WET} = \underset{\text{WET}}{\operatorname{argmin}} \sum_{i=E_1, E_2} (D_{\text{DB},i}(\text{WET}) - D_{\text{M},i})^2 \quad (\text{Eq. 1})$$

Where D_{DB} is the detector dose in the calibration look-up table, and D_{M} is the scatter-corrected or measured detector dose in one pixel after the proton beam passed through the medium at energy E_1 or E_2 , respectively 200 MeV and 220 MeV. This method is used to determine the WET in both the total dose measurements and the primary dose measurements by particles that underwent limited scattering. The WET calculation accuracy (Equation 1) includes a comparison with an idealistic in-house three-dimensional proton raytracing solution. Raytracing was calculated according to a geometry existing of an ideal point source and a perfect virtual detector, which had the same image resolution and image size as the simulated flat-panel imager (Jacobs *et al.*, 1998; Siddon, 1985).

Following proton radiography workflow is proposed in a clinical setting to determine the WET. First, the multi-energy proton radiographs are acquired from the patient with a single flat-panel detector. Then, the large angle scattered proton distributions are predicted by ProtonDSE for all beam energies. Next, the ProtonDSE-predicted scatter distributions are removed from the multi-energy measurements to generate scatter-corrected radiographs. Finally, the WET of the patient is determined with Equation 1 by applying the calibration curve for primary detector dose to the ProtonDSE-corrected radiograph.

2.5 Proton radiographs as dataset

The open-access ‘Head-Neck-Radiomics-HN1’ dataset (Wee and Dekker, 2019) consisting of 137 cancer patient CTs was downloaded as Monte Carlo simulation dataset. Manual filtering was applied on the entire dataset to omit patient CTs with severe artifacts caused by dental fillings. Only head and neck patients with no, or very minor, CT artifacts (n=80) were considered as Monte Carlo simulation geometry. The CT to mass density conversion curve, which is essential information required for Monte Carlo simulations, was made available on request by the dataset owner.

All digital anthropomorphic CT geometries were converted into a Monte Carlo simulation geometry interpretable by TOPAS MC. The mass density inside the CT voxels was calculated by applying the CT to mass density conversion curve (piecewise linear

interpolation), and seven ICRU (Report 44) materials or tissues were assigned with respect to the CT numbers in Hounsfield Units. Air was assigned below -850 HU, inflated lung between -850 HU and -200 HU, adipose tissue between -200HU and 0 HU, brain (gray/white matter: 50/50) between 0 HU and 25 HU, skeletal muscle between 25 HU and 100 HU, cartilage bone between 100 HU and 300 HU and cortical bone above 300 HU.

A total number of 236 proton radiographs per nominal proton energy were simulated through 80 different digital anthropomorphic patient geometries. Every proton radiograph was simulated with a 20 x 24 cm² proton imaging field with 2 mm spot spacing at isocenter. The imaging protocol includes the simulation of 9888 proton pencil beams (3E5 histories/beam). On average, three proton radiographs were simulated per digital anthropomorphic phantom. The first radiograph was simulated in the left-right direction without random rigid image transformation, the second simulation included random patient-specific isocenter shifts and isocenter rotations limited within ± 7.5 degrees and 3 cm, the third simulation allowed larger random rotations within ± 15 degrees. All pseudorandom numbers were sampled from a uniform distribution.

Limited preprocessing was applied to the dataset prior to training. The simulated radiographs were cropped to 256 x 256 pixels with respect to the radiograph isocenter and without spatial resolution loss (1 x 1 mm²). All radiographs that scored the total imaging dose were normalized with respect to the highest calibration value prior to distal fall-off of the Bragg peak (Figure 3, left). The primary dose radiographs were normalized to the highest dose in the calibration database (i.e. WET = 0 mm, no- or negligible proton scatter leads to a maximum primary dose) (Figure 3, right). Except for 2D image flipping and rotations, no other data augmentation operations such as deformable image registration, shearing, or zooming are applied to respect the laws of particle physics. Network evaluation was performed in a 60%/25%/15% scheme for training, validation, and testing.

2.6 Quantitative evaluation metrics

The ProtonDSE network performance was evaluated at 200 MeV and 220 MeV by three quantitative metrics between the vectorized primary ground truth dose by Monte Carlo simulations (I_{MC}) and the vectorized scatter-corrected dose ($I_{corr} = I_{total} - I_{ProtonDSE}$). The mean absolute error (MAE), the mean absolute percentage error (MAPE), and the mean squared error (MSE) were calculated according to Equation 2-4, where N is the total pixel number of the proton radiograph.

$$MAE = \frac{1}{N} \sum_n |I_{MC} - I_{corr}| \quad (Eq. 2)$$

$$\text{MAPE} = \frac{100\%}{N} \sum_n \frac{|I_{\text{MC}} - I_{\text{corr}}|}{I_{\text{MC}}} \quad (\text{Eq. 3})$$

$$\text{MSE} = \frac{1}{N} \sum_n (I_{\text{MC}} - I_{\text{corr}})^2 \quad (\text{Eq. 4})$$

3. Results

3.1 Validation of the Monte Carlo model

Figure 2 shows the dose measurements and simulations that were normalized to their peak energy. Proton ranges in water were compared at 200 MeV and 220 MeV at the depth of the 80% dose distal fall-off (R80). $R80_{200\text{MeV}}$ and $R80_{220\text{MeV}}$ were both in <0.1% agreement between Monte Carlo simulation (261.19 mm, 306.79 mm) and measurement (261.23 mm, 306.84 mm). The MAE between the normalized distributions of the simulated and the measured spot profiles was $5.5\text{E-}4$ for 200 MeV, and $5.4\text{E-}4$ for 220 MeV. Subsequently, the horizontal cross profiles of the spot distributions were in good agreement (Figure 2).

Figure 2. Monte Carlo model validation of the simulated dose profiles and spot distributions (dashed lines) with proton beam measurements (solid lines) at 200 MeV (blue) and 220 MeV (orange).

3.2 WET calibration database

Two WET to detector dose calibration curves were simulated with PMMA slabs of varying thickness in the proton beam path, one curve was obtained at 200 MeV, and one at 220 MeV. The PMMA thickness was theoretically converted to WET by applying a 1.1672 ratio. One calibration curve was obtained for the total imaging dose (Figure 3, left), and a second calibration curve was obtained for the primary detector dose calibration (Figure 3, right). Only calibration data on the left from the vertical dashed black line (WET = 256.8 mm) was used to prevent incorrect data supply to our WET optimization function (Equation 1). This means that only geometries with a maximum WET lower than 256.8 mm are suited for this method (with 200 MeV as the lowest energy).

Figure 3. The dose to water equivalent tickness (WET, in mm) calibration curves for the total detector dose (left) and the primary detector dose (right, < 2.5 degrees). The vertical dashed black line indicates the maximum WET value to be measured with a 200-220 MeV energy combination.

3.3 ProtonDSE: model accuracy and performance

The average computation time required to train ProtonDSE_{200MeV} and ProtonDSE_{220MeV} was 4660 ± 70 seconds for 250 epochs, and the average ProtonDSE prediction time was 50 ± 2 milliseconds. Figure 4 depicts the MAE loss curves associated with the training of ProtonDSE networks.

Figure 4. The mean absolute error (MAE) loss functions of the ProtonDSE_{200MeV} network (left, blue) and the ProtonDSE_{220MeV} network (orange, right) plotted against the number of epochs for the training and validation data.

All quantitative network performance metrics with ± 1 standard deviation were listed in Table 1. The MAE, MAPE, and MSE were calculated on the entire normalized (Figure 3) test dataset of 35 proton radiographs per proton energy. All metrics ignored air-related detector pixels by thresholding, together with pixels that were out of the imaging field-of-view.

Table 1. The mean absolute error (MAE), the mean squared error (MSE), and the mean absolute percentage error (MAPE, in %) ± 1 standard deviation for the ProtonDSE_{200MeV} and ProtonDSE_{220MeV} networks applied on our test dataset of 35 subjects.

	ProtonDSE _{200MeV}	ProtonDSE _{220MeV}
MAE	$4.9\text{E-}3 \pm 1.2\text{E-}3$	$4.8\text{E-}3 \pm 1.4\text{E-}3$
MSE	$6.8\text{E-}5 \pm 4.5\text{E-}5$	$6.5\text{E-}5 \pm 4.5\text{E-}5$
MAPE (in %)	1.4 ± 0.4	1.2 ± 0.4

Figure 5 shows the normalized distributions of the total measured dose (I_{total}), the dose by protons that deviated more than 2.5 degrees from their initial path (I_{MC} and $I_{\text{ProtonDSE}}$), the corrected proton radiographs (I_{Primary} and $I_{\text{Corr(ected)}}$), and the reference WET image by idealistic raytracing.

Figure 5. Proton radiographs and particle scatter distributions at 200 MeV of one example head-and-neck patient of our test dataset. $I_{\text{ProtonDSE}}$ was predicted by $\text{ProtonDSE}_{200\text{MeV}}$ from I_{Total} , with I_{MC} considered as reference. $\text{WET}_{\text{Raytracing}}$ (in mm) is the reference image calculated by idealistic raytracing.

The predicted and reference doses of four random example patients at 200 MeV are depicted in Figure 6. The top row indicates the normalized total dose used as input by our network architecture. The corrected total dose by the $\text{ProtonDSE}_{200\text{MeV}}$ network architecture (I_{Corr}) was evaluated relatively to the primary Monte Carlo reference dose (I_{Primary}). The largest relative dose errors were found in anatomical regions containing air such as the nasal cavities, in relatively thin regions such as the neck or the nose, or in thick regions above 256.8 mm (Figure 3) such as the shoulders. The MAPE (excluding air voxels and pixels outside the imaging field-of-view) for the particular patients depicted in Figure 6 (bottom row) was equal to 2.3%, 1.2%, 1.6%, and 1.9%, respectively.

Figure 6. The normalized total dose, primary dose by Monte Carlo simulation, primary dose predicted by ProtonDSE_{200MeV}, and the relative primary dose difference (%) for 4 randomly selected head-and-neck patients of the test dataset

Figure 7 illustrates the WET calculation for one randomly selected patient out of our test dataset. The middle row of Figure 7 shows the WET calculation by Equation 1 at 200 MeV and 220 MeV on the total dose images without correction $WET(I_{Total})$, the Monte Carlo predicted primary dose image as reference $WET(I_{Primary})$, and the ProtonDSE corrected primary dose images $WET(I_{Corrected})$. The normalized dose distributions were converted back to absolute dose prior to WET calculation. In the bottom row of Figure 7, the relative WET error was calculated with respect to the direct WET image obtained by idealistic three-dimensional proton raytracing. The optimization function (Equation 1) is shown for one random pixel of the radiograph, indicated by the black arrow in $WET_{Raytracing}$.

A mean absolute WET error (in %) with respect to $WET_{Raytracing}$ was calculated in a manually delineated region-of-interest inside the brain projection of this example patient (ROI_{Brain}) (Figure 7), which excludes the proton imaging field border. The error was determined $4.3 \pm 4.1\%$ for $WET(I_{Total})$, $2.2 \pm 1.4\%$ for $WET(I_{Primary})$, and $2.5 \pm 2.0\%$ for $WET(I_{Corrected})$.

Figure 7. Top row: the primary (=reference) and ProtonDSE corrected dose images at 200 MeV and 220 MeV for one example patient of our test dataset. Middle row: Water Equivalent Thickness (WET, in mm) images calculated with Equation 1 (except for the raytracing image) on the original measured signal I_{Total} , on the primary signal $I_{Primary}$, and on the corrected signal $I_{Corrected}$. Bottom row: Relative WET differences (in %) with respect to the raytracing radiograph as reference. The vertical dashed black line in the optimization plot from Equation 1 (bottom right) indicates the reference value by raytracing for one random pixel, indicated by a black arrow in $WET(I_{Raytracing})$.

4. Discussion

It remains a persistent challenge to design a relatively light, compact, and clinically implementable proton radiography system with sufficient accuracy to evaluate relevant anatomy changes that would affect treatment planning, or to optimize CT numbers from proton radiographic images (Krah *et al.*, 2019b). The use of flat-panel detector technology in proton radiography was proposed before in simulation studies (Huo *et al.*, 2019), and in studies that performed actual measurements (Zhang *et al.*, 2019). However, multiple Coulomb scattering (i.e. range mixing) and (in)elastic nuclear scatter interactions in the medium limited the spatial resolution of earlier proposed single detector solutions. Range mixing is especially difficult to resolve in highly complex heterogeneous geometries such as a patient (Brooke and Penfold, 2020; Krah *et al.*, 2019a; Khellaf *et al.*, 2019), resulting in a proton pencil beam with mixed residual energies. (Zhang *et al.*, 2019) aimed to solve the problem of range mixing by applying a pixel rejection technique on their measured WET image, based on empirical local standard deviation thresholding. Nevertheless, their solution is maybe less optimal in terms of sufficiency because some pixel information is currently ignored after thresholding.

In this work, we proposed a single detector multi-energy proton radiography system supported by Artificial Intelligence and Monte Carlo simulations, e.g. to avoid the need of pixel rejection techniques. A Monte Carlo simulation dataset of multiple proton radiographies was composed from 80 different head-and-neck patients at nominal proton energies of 200 MeV and 220 MeV. A dedicated TOPAS extension was developed to separate protons that traversed through the digital anthropomorphic patient geometry and exited the patient with an angle deviating more than 2.5 degrees from their initial momentum (Figure 5). This criterion partially eliminates the contaminating contribution of scattered protons, and protons that underwent too many consecutive Coulomb scattering events. By definition, scatter corrected proton radiographs have an improved spatial resolution due to the removal of unbalanced proton scatter at boundaries. A multi-layer two-dimensional U-Net network architecture (ProtonDSE) (Figure 1) was trained on the Monte Carlo dataset to predict and correct for the contaminating scattered proton distribution (Figure 5 and Figure 6). The prediction performance of ProtonDSE was evaluated with three quantitative metrics that indicated an accuracy below 1.5% in our test dataset for both 200 MeV and 220 MeV (Table 1). Further, simulations with PMMA slabs of varying thicknesses were performed to compose a calibration database that relates the total and primary doses scored in the flat-panel scintillation layer to the corresponding WET in look-up tables (Figure 3 and Equation 1). Results of an illustrative patient case have indicated an improved WET accuracy in the brain for the ProtonDSE corrected proton radiograph ($2.5 \pm 2.0\%$) compared to the proton radiograph calculated on the total dose image ($4.3 \pm 4.1\%$). Outside the manually delineated

1 region-of-interest in the projection of the brain, errors were larger than 2.5%, mostly in
2 regions that had a lower WET than the brain, such as the neck.

3
4 The proton energies used in this work (200 MeV and 220 MeV) only start to give
5 distinguishable dose values above 100 mm WET (Figure 3). Below 100 mm WET, Equation 1
6 becomes highly unstable, which means that the smallest dose deviations, or possibly
7 measurement noise, could yield WET values between 0 mm and 100 mm WET. The usage of
8 lower proton energies ($70 \text{ MeV} < E < 160 \text{ MeV}$) for imaging should therefore be considered in
9 anatomical sites that lack a desired thickness for equation stability, such as the neck, the chin,
10 or the nasal cavity. The dose response functions of lower energy proton beams for imaging
11 would give more distinguishable values at lower WETs. One interesting solution would be a
12 patient-specific energy selection method that calculates the optimized beam energy based
13 on WET (+range) estimations from the planning CT. The development of such a system is likely
14 to give more distinguishable dose values at body sites with large varying WET values. It should,
15 however, be prevented in any scenario that low energy beams are stopped inside the patient,
16 in analogy with treatment beams. Pencil beam scanning protocols and energy selection
17 procedures could potentially be optimized (e.g. based on the initial planning CT) to perform
18 personalized multi-energy proton imaging for every treatment site by beam painting. This
19 personalized proton beam painting technique for imaging also comes with the additional
20 advantage of reduced imaging dose prior to treatment.

21
22 Although the results presented in this work mostly rely on a realistic Monte Carlo
23 simulation model that was validated with experimental data (Figure 2), confirmation of our
24 method is still required in real proton imaging experiments with a flat-panel detector. This
25 also includes the complexity of composing a primary dose to WET calibration curve (Figure 3,
26 right) which is not directly measurable in a realistic scenario. The use of slit camera
27 measurements should however be investigated to compose the primary dose to WET
28 calibration curve. Alternatively, a hybrid method could potentially be adopted in which the
29 total dose measurements are scaled with the Monte Carlo simulations, and the scattered
30 Monte Carlo distributions of the protons are scaled accordingly. Moreover, flat-panel artifacts
31 known from X-ray imaging have to be considered in real measurements and should therefore
32 be investigated thoroughly for proton imaging. These include flat-panel detector calibration,
33 detector saturation, image lag (Siewerdsen and Jaffray, 1999), radiation hardness (Zhang *et al.*,
34 2017b), defect pixel artifacts, and ghosting artifacts (Mail *et al.*, 2007).

35
36 Following the general trend of using artificial intelligence techniques in radiotherapy,
37 it should be noted that it remains challenging to commission Artificial Intelligence techniques
38 in medicine of which the performance cannot directly be verified by a human observer in
39 clinical routine (Fiorino *et al.*, 2020; Huynh *et al.*, 2020; Vandewinckele *et al.*, 2020). The

1 authors therefore foresee a proton radiography commissioning protocol with
2 anthropomorphic ground-truth phantoms (Wohlfahrt *et al.*, 2018) of which the elemental
3 composition and density, and thus the stopping powers, are provided by the manufacturer,
4 or are measured with conventional protocols that include proton irradiation and Bragg peak
5 measurements (Zhang and Newhauser, 2009; Zhang *et al.*, 2010).

6 7 **5. Conclusion**

8
9 A deep proton scatter estimation network architecture (ProtonDSE) was developed
10 and trained to improve the accuracy of compact single-detector and multi-energy proton
11 radiography. The trained ProtonDSE networks for 200 MeV and 220 MeV showed good
12 performance in terms of proton scatter estimation below 1.5% in the entire test database.
13 The mean WET error with respect to idealistic ray tracing was determined 2.5% for a digital
14 anthropomorphic phantom case. The simulation results of our current configuration at two
15 energies should now be confirmed in actual anthropomorphic ground-truth phantom
16 measurements, and in the future, it is likely that more simulations are essential to extend the
17 training database with proton radiographs of different anatomical sites, pencil beam scanning
18 protocols, proton beam energies, and imaging angles. Furthermore, the development of
19 personalized multi-energy proton imaging protocols by beam scanning is likely to perform
20 more accurate proton radiography in thin anatomical regions, and therefore gained our
21 interest for future studies, together with the personalized imaging dose optimization prior to
22 proton therapy treatment.

23 24 25 **Acknowledgements**

26
27 This work was partially performed with research funding from Google Cloud. Martin Uray
28 from the Fachhochschule Salzbrug (Austria) is acknowledged for his insights related to the
29 network architecture used in this work. Dr. Leonard Wee and Dr. Michel Öllers are
30 acknowledged for providing additional information regarding the Head-Neck-Radiomics-HN1
31 dataset used in this work. Marie Cohilis is supported by the Télévie Grant from the Belgian
32 Fonds National pour la Recherche Scientifique F.R.S-FNRS (Grant No. 7450517F). Dr. Kevin
33 Souris is funded by the Walloon region (MECATECH/BIOWIN, grant number 8090)

34 35 36 **Conflicts of Interest Statement**

37 The authors report no conflicts of interest.

References

- Almeida I P, Schyns L, Vaniqui A, van der Heyden B, Dedes G, Resch A F, Kamp F, Zindler J D, *et al.* 2018 Monte Carlo proton dose calculations using a radiotherapy specific dual-energy CT scanner for tissue segmentation and range assessment *Phys Med Biol* **63** 115008
- Bentefour el H, Schnuerer R and Lu H M 2016 Concept of proton radiography using energy resolved dose measurement *Phys Med Biol* **61** N386-93
- Brooke M D and Penfold S N 2020 An inhomogeneous most likely path formalism for proton computed tomography *Phys Med* **70** 184-95
- Deffet S, Cohilis M, Souris K, Salvo K, Depuydt T, Sterpin E and Macq B 2020 openPR - A computational tool for CT conversion assessment with proton radiography *Med Phys*
- Doolan P J, Testa M, Sharp G, Bentefour E H, Royle G and Lu H M 2015 Patient-specific stopping power calibration for proton therapy planning based on single-detector proton radiography *Phys Med Biol* **60** 1901-17
- Faddegon B A, Shin J, Castenada C M, Ramos-Mendez J and Daftari I K 2015 Experimental depth dose curves of a 67.5 MeV proton beam for benchmarking and validation of Monte Carlo simulation *Med Phys* **42** 4199-210
- Fiorino C, Jeraj R, Clark C H, Garibaldi C, Georg D, Muren L, van Elmpt W, Bortfeld T and Jornet N 2020 Grand challenges for medical physics in radiation oncology *Radiother Oncol*
- Hansen D C, Landry G, Kamp F, Li M, Belka C, Parodi K and Kurz C 2018 ScatterNet: A convolutional neural network for cone-beam CT intensity correction *Med Phys* **45** 4916-26
- Huo W, Zwart T, Cooley J, Huang K, Finley C, Jee K W, Sharp G C, Rosenthal S, *et al.* 2019 A single detector energy-resolved proton radiography system: a proof of principle study by Monte Carlo simulations *Phys Med Biol* **64** 025016
- Huynh E, Hosny A, Guthier C, Bitterman D S, Petit S F, Haas-Kogan D A, Kann B, Aerts H and Mak R H 2020 Artificial intelligence in radiation oncology *Nat Rev Clin Oncol*
- Jacobs F, Sundermann E, De Sutter B, Christiaens M and Lemahieu I 1998 A fast algorithm to calculate the exact radiological path through a pixel or voxel space. *J. Comput Inf Technol* **6** 89-94
- Jee K W, Zhang R, Bentefour E H, Doolan P J, Cascio E, Sharp G, Flanz J and Lu H M 2017 Investigation of time-resolved proton radiography using x-ray flat-panel imaging system *Phys Med Biol* **62** 1905-19
- Johnson R P 2018 Review of medical radiography and tomography with proton beams *Rep Prog Phys* **81** 016701
- Khellaf F, Krah N, Letang J M, Collins-Fekete C A and Rit S 2020 A comparison of direct reconstruction algorithms in proton computed tomography *Phys Med Biol* **65** 105010
- Khellaf F, Krah N, Rinaldi I, Letang J M and Rit S 2019 Effects of transverse heterogeneities on the most likely path of protons *Phys Med Biol* **64** 065003
- Kingma D and Lei J 2015 Adam: A method for stochastic optimization *arXiv*
- Krah N, Letang J M and Rit S 2019a Polynomial modelling of proton trajectories in homogeneous media for fast most likely path estimation and trajectory simulation *Phys Med Biol* **64** 195014
- Krah N, Patera V, Rit S, Schiavi A and Rinaldi I 2019b Regularised patient-specific stopping power calibration for proton therapy planning based on proton radiographic images *Phys Med Biol* **64** 065008

- 1 Krah N, Quinones C T, Letang J M and Rit S 2020 Scattering proton CT *Phys Med Biol*
- 2 Lalonde A, Winey B A, Verburg J M, Paganetti H and Sharp G C 2020 Evaluation of CBCT scatter
- 3 correction using deep convolutional neural networks for head and neck adaptive
- 4 proton therapy *Phys Med Biol*
- 5 Li B, Lee H C, Duan X, Shen C, Zhou L, Jia X and Yang M 2017 Comprehensive analysis of proton
- 6 range uncertainties related to stopping-power-ratio estimation using dual-energy CT
- 7 imaging *Phys Med Biol* **62** 7056-74
- 8 Li T, Liang Z, Singanallur J V, Satogata T J, Williams D C and Schulte R W 2006 Reconstruction
- 9 for proton computed tomography by tracing proton trajectories: a Monte Carlo study
- 10 *Med Phys* **33** 699-706
- 11 Maier J, Eulig E, Voth T, Knaup M, Kuntz J, Sawall S and Kachelriess M 2019 Real-time scatter
- 12 estimation for medical CT using the deep scatter estimation: Method and robustness
- 13 analysis with respect to different anatomies, dose levels, tube voltages, and data
- 14 truncation *Med Phys* **46** 238-49
- 15 Mail N, O'Brien P and Pang G 2007 Lag correction model and ghosting analysis for an indirect-
- 16 conversion flat-panel imager *J Appl Clin Med Phys* **8** 2483
- 17 McGowan S E, Burnet N G and Lomax A J 2013 Treatment planning optimisation in proton
- 18 therapy *Br J Radiol* **86** 20120288
- 19 Meijers A, Free J, Wagenaar D, Deffet S, Knopf A C, Langendijk J A and Both S 2020 Validation
- 20 of the proton range accuracy and optimization of CT calibration curves utilizing range
- 21 probing *Phys Med Biol* **65** 03NT2
- 22 Newhauser W and Zhang R 2015 The physics of proton therapy *Physics in Medicine and*
- 23 *Biology* **60** R155-R209
- 24 Paganetti H 2012 Range uncertainties in proton therapy and the role of Monte Carlo
- 25 simulations *Phys Med Biol* **57** R99-117
- 26 Poludniowski G, Allinson N M and Evans P M 2015 Proton radiography and tomography with
- 27 application to proton therapy *Br J Radiol* **88** 20150134
- 28 Schneider U and Pedroni E 1994 Multiple Coulomb scattering and spatial resolution in proton
- 29 radiography *Med Phys* **21** 1657-63
- 30 Siddon R L 1985 Fast calculation of the exact radiological path for a three-dimensional CT
- 31 array *Med Phys* **12** 252-5
- 32 Siewerdsen J H and Jaffray D A 1999 Cone-beam computed tomography with a flat-panel
- 33 imager: effects of image lag *Med Phys* **26** 2635-47
- 34 Taasti V T, Hansen D C, Michalak G J, Deisher A J, Kruse J J, Muren L P, Petersen J B B and
- 35 McCollough C H 2018 Theoretical and experimental analysis of photon counting
- 36 detector CT for proton stopping power prediction *Med Phys* **45** 5186-96
- 37 Unkelbach J and Paganetti H 2018 Robust Proton Treatment Planning: Physical and Biological
- 38 Optimization *Semin Radiat Oncol* **28** 88-96
- 39 van der Heyden B, Almeida I P, Vilches-Freixas G, Van Beveren C, Vaniqui A, Ares C, Terhaag
- 40 K, Fonseca G P, *et al.* 2019a A comparison study between single- and dual-energy CT
- 41 density extraction methods for neurological proton monte carlo treatment planning
- 42 *Acta Oncol* 1-9
- 43 van der Heyden B, Uray M, Fonseca G P, Huber P, Us D, Messner I, Law A, Parii A, *et al.* 2020
- 44 A Monte Carlo based scatter removal method for non-isocentric cone-beam CT
- 45 acquisitions using a deep convolutional autoencoder *Phys Med Biol* **65** 145002

- 1 van der Heyden B, Wohlfahrt P, Eekers D B P, Richter C, Terhaag K, Troost E G C and Verhaegen
- 2 F 2019b Dual-energy CT for automatic organs-at-risk segmentation in brain-tumor
- 3 patients using a multi-atlas and deep-learning approach *Sci Rep* **9** 4126
- 4 Vandewinckele L, Claessens M, Dinkla A, Brouwer C, Crijns W, Verellen D and van Elmpst W
- 5 2020 Overview of artificial intelligence-based applications in radiotherapy:
- 6 Recommendations for implementation and quality assurance *Radiother Oncol*
- 7 Wee L and Dekker A 2019 Data from Head-Neck-Radiomics-HN1 [Data set]. The Cancer
- 8 Imaging Archive.)
- 9 White D R, Woodard H Q and Hammond S M 1987 Average soft-tissue and bone models for
- 10 use in radiation dosimetry *Br J Radiol* **60** 907-13
- 11 Wohlfahrt P, Mohler C, Richter C and Greilich S 2018 Evaluation of Stopping-Power Prediction
- 12 by Dual- and Single-Energy Computed Tomography in an Anthropomorphic Ground-
- 13 Truth Phantom *Int J Radiat Oncol Biol Phys* **100** 244-53
- 14 Wohlfahrt P, Mohler C, Stutzer K, Greilich S and Richter C 2017 Dual-energy CT based proton
- 15 range prediction in head and pelvic tumor patients *Radiother Oncol* **125** 526-33
- 16 Wohlfahrt P, Mohler C, Troost E G C, Greilich S and Richter C 2019 Dual-Energy Computed
- 17 Tomography to Assess Intra- and Inter-Patient Tissue Variability for Proton Treatment
- 18 Planning of Patients With Brain Tumor *Int J Radiat Oncol Biol Phys* **105** 504-13
- 19 Woodard H Q and White D R 1986 The composition of body tissues *Br J Radiol* **59** 1209-18
- 20 Yang J, Veeraraghavan H, Armato S G, 3rd, Farahani K, Kirby J S, Kalpathy-Kramer J, van Elmpst
- 21 W, Dekker A, *et al.* 2018 Autosegmentation for thoracic radiation treatment planning:
- 22 A grand challenge at AAPM 2017 *Med Phys* **45** 4568-81
- 23 Yang M, Zhu X R, Park P C, Titt U, Mohan R, Virshup G, Clayton J E and Dong L 2012
- 24 Comprehensive analysis of proton range uncertainties related to patient stopping-
- 25 power-ratio estimation using the stoichiometric calibration *Phys Med Biol* **57** 4095-
- 26 115
- 27 Zhang R, Baer E, Jee K W, Sharp G C, Flanz J and Lu H M 2017a Investigation of real tissue
- 28 water equivalent path lengths using an efficient dose extinction method *Phys Med Biol*
- 29 **62** 5640-51
- 30 Zhang R, Jee K W, Cascio E, Sharp G C, Flanz J B and Lu H M 2017b Improvement of single
- 31 detector proton radiography by incorporating intensity of time-resolved dose rate
- 32 functions *Phys Med Biol* **63** 015030
- 33 Zhang R and Newhauser W D 2009 Calculation of water equivalent thickness of materials of
- 34 arbitrary density, elemental composition and thickness in proton beam irradiation
- 35 *Phys Med Biol* **54** 1383-95
- 36 Zhang R, Sharp G C, Jee K W, Cascio E, Harms J, Flanz J B and Lu H M 2019 Iterative optimization
- 37 of relative stopping power by single detector based multi-projection proton
- 38 radiography *Phys Med Biol* **64** 065022
- 39 Zhang R, Taddei P J, Fitzek M M and Newhauser W D 2010 Water equivalent thickness values
- 40 of materials used in beams of protons, helium, carbon and iron ions *Phys Med Biol* **55**
- 41 2481-93
- 42