

Article type : Research Article

# Comparison of CBCT-based dose calculation methods in head and neck cancer radiotherapy: from Hounsfield unit to density calibration curve to deep learning

Anaïs Barateau<sup>1</sup>, Renaud De Crevoisier<sup>1</sup>, Axel Largent<sup>1</sup>, Eugenia Mylona<sup>1</sup>, Nicolas Perichon<sup>1</sup>, Joël Castelli<sup>1</sup>, Enrique Chajon<sup>1</sup>, Oscar Acosta<sup>1</sup>, Antoine Simon<sup>1</sup>, Jean-Claude Nunes<sup>1</sup>, Caroline Lafond<sup>1</sup>

1. Univ. Rennes, CLCC Eugène Marquis, INSERM, LTSI - UMR 1099, F-35000 Rennes, France

Short title: Deep learning for CBCT dose calculation

Corresponding author: [anais.barateau@gmail.com](mailto:anais.barateau@gmail.com)

Mailing address:

**Centre Eugène Marquis**

Unité de physique médicale

Rue de la Bataille Flandres Dunkerque - CS 44229

35042 RENNES CEDEX

FRANCE

Accepted Article

## Abstract

### Purpose

Anatomical variations occur during head and neck (H&N) radiotherapy treatment. kV cone-beam computed tomography (CBCT) images can be used for daily dose monitoring to assess dose variations owing to anatomic changes. Deep learning methods (DLMs) have recently been proposed to generate pseudo-CT (pCT) from CBCT to perform dose calculation. This study aims to evaluate the accuracy of a DLM and to compare this method with three existing methods of dose calculation from CBCT in H&N cancer radiotherapy.

### Methods

Forty-four patients received VMAT for H&N cancer (70-63-56 Gy). For each patient, reference CT (Bigbore, Philips) and CBCT images (XVI, Elekta) were acquired. The DLM was based on a generative adversarial network. The three compared methods were: i) a method using a density to Hounsfield Unit (HU) relation from phantom CBCT image (HU-D curve method), ii) a water-air-bone density assignment method (DAM), and iii) a method using deformable image registration (DIR). The imaging endpoints were the mean absolute error (MAE) and mean error (ME) of HU from pCT and reference CT ( $CT_{ref}$ ). The dosimetric endpoints were dose discrepancies and 3D gamma analyses (local, 2%/2mm, 30% dose threshold). Dose discrepancies were defined as the mean absolute differences between DVHs calculated from the  $CT_{ref}$  and pCT of each method.

### Results

In the entire body, the MAEs and MEs of the DLM, HU-D curve method, DAM, and DIR method were 82.4 and 17.1 HU, 266.6 and 208.9 HU, 113.2 and 14.2 HU, and 95.5 and -36.6 HU, respectively. The MAE obtained using the DLM differed significantly from those of other methods (Wilcoxon,  $p < 0.05$ ). The DLM dose discrepancies were  $7 \pm 8$  cGy (maximum = 44 cGy) for the ipsilateral parotid gland  $D_{mean}$  and  $5 \pm 6$  cGy (max = 26 cGy) for the contralateral parotid gland mean dose ( $D_{mean}$ ). For the parotid gland  $D_{mean}$ , no significant dose difference was observed between the DLM and other methods.

The mean 3D gamma pass-rate  $\pm$  standard deviation was  $98.1 \pm 1.2\%$ ,  $91.0 \pm 5.3\%$ ,  $97.9 \pm 1.6\%$ , and  $98.8 \pm 0.7\%$  for the DLM, HU-D method, DAM, and DIR method, respectively. The gamma pass-rates and mean gamma results of the HU-D curve method, DAM, and DIR method differed

Accepted Article

significantly from those of the DLM. The mean calculation time to generate one pCT was 30 s for the deep learning and DIR methods.

### **Conclusions**

For H&N radiotherapy, DIR method and DLM appears as the most appealing CBCT-based dose calculation methods among the four methods in terms of dose accuracy as well as calculation time. Using the DIR method or DLM with CBCT images enables dose monitoring in the parotid glands during the treatment course and may be used to trigger replanning.

Keywords: Deep learning; CBCT dose calculation; head and neck cancer



## Introduction

Recent developments in modulated treatment plans and image-guided radiotherapy (IGRT) devices have allowed more precise and targeted head and neck (H&N) treatments, with improved sparing of organs at risk while covering target volumes. However, during H&N radiotherapy, patients are typically subject to anatomical variations such as tumor shrinkage or weight loss<sup>1,2</sup>. These variations can induce discrepancies between planned and delivered doses. Acquisitions of daily or weekly 3D kV cone-beam computed tomography (CBCT) images can be used to quantify dose discrepancies related to anatomical variations. CBCT images can be used to calculate the delivered “dose of the day”<sup>3</sup> to viewable and critical structures such as the spinal cord and parotid glands (PGs). However, dose calculation from CBCT is challenging not only owing to the “poor” image quality, but also to the limited field of view (FOV) and the inconsistency of the Hounsfield units (HUs). Moreover, CBCT dose calculation often lacks ground truth (dose computed from the reference CT acquired at the same time as CBCT) for quantifying the uncertainties.

Several methods to perform CBCT dose calculation have been proposed: i) calibration curve between the HU and densities (HU-D curve)<sup>4-6</sup>, ii) density assignment method (DAM)<sup>5,7-10</sup>, iii) deformable image registration (DIR) between CT and CBCT<sup>3,10-12</sup> and iv) machine learning to generate a pseudo-CT (pCT)<sup>13-15</sup>. i) The HU-D curve established from a CBCT image can be used to convert CBCT HUs to densities for dose calculation. This curve can be defined with either an “adapted” phantom<sup>6</sup> (according to anatomical localization) or patient CBCT images<sup>4,5</sup>. Although these methods are straightforward, they are sensitive to CBCT artefacts<sup>16</sup> and patient scattering. ii) The density assignment method (also known as the bulk density method) involves segmenting an image into two to six tissue classes (e.g., soft tissues, air and bones) before assigning density to each class. Nevertheless, this method is dependent of structure segmentation and provide an image with homogeneous tissues. iii) By deforming CT to CBCT, a “deformed” CT is generated and can be used for dose calculation. CT-CBCT DIR can be difficult owing to intrinsic CBCT limitations, such as noise, low contrast, and reduced FOV. Moreover, the registration step is complex owing to large anatomical variations, such as tumor shrinkage or weight loss<sup>17</sup>. iv) Machine learning methods are based on patches or deep learning (DL), to generate a pCT (i.e., synthetic images) from CBCT. Machine learning methods require a large training cohort and most of these methods

require co-registered data for the training step. The main advantages and drawbacks of the HU-D curve method, DAM, DIR method, and DLM are summarized in the two first columns of Table 1.

Deep learning methods (DLMs) are models composed of multiple processing layers that learn multiscale representations of data through multiple levels of abstraction<sup>18</sup>. These methods have recently been introduced in radiotherapy for various applications, such as image segmentation, image reconstruction, image registration, treatment planning, and radiomics<sup>19–25</sup>. DLMs have been primarily proposed for pCT generation from magnetic resonance imaging (MRI)<sup>26–31</sup>. They are particularly appealing owing to their fast computation time. One of the first DLMs for pCT generation was based on the U-Net architecture. More recently, DLMs using generative adversarial networks (GANs) have been proposed, offering the theoretical advantage of providing more realistic pCTs by obtaining an adversarial feedback from a discriminator network<sup>32</sup>.

Some studies have recently proposed DLMs for pCT generation from CBCT, mainly for scatter correction<sup>33,34</sup>. Other studies proposed the DLM for pCT generation from CBCT in prostate<sup>35–37</sup>, pancreas<sup>38</sup>, and H&N<sup>13–15</sup>, for dose calculation. H&N studies have been performed using either U-Net or cycleGAN architectures to generate pCT from CBCT.

To the best of our knowledge, no DLM to generate pCT from H&N CBCT has been compared with other methods for dose calculation from CBCT. In H&N radiotherapy, previous studies showed dose differences of less than 3% for the DAM<sup>5,7–10</sup>, less than 2% for the DIR method<sup>3,10,11</sup>, and less than 1% for the DLM.

This study aims to evaluate the accuracy of the DL method for CBCT H&N dose calculation and to compare this method with the HU-D curve, density assignment, and DIR methods.

## Materials and methods

### **Patient data**

Forty four patients with locally advanced oropharyngeal carcinomas were retrospectively selected.

### **Image acquisition**

All patients had a planning CT scan, and 14 patients had weekly CT scans. All CT images were acquired on a BigBore (Philips) scanner, with a 2 mm slice thickness from the vertex to the carina. Patient positioning was assured using a personalized thermoplastic head and shoulder mask with

five fixation points. The treatment isocenter was set at a reproducible bony place, between the C3-C4 vertebrae.

Weekly CBCTs were acquired with an XVI (Elekta) on a VERSAHD linac (Elekta). The settings for the H&N CBCT acquisition were as follows: 120 kV, 0.4 mAs per frame, 660 frames, M20 FOV, and 2 mm slice thickness. For the 14 patients who had weekly CT scans, the weekly CBCTs were acquired for a time close to the CTs.

### **Delineation and dose calculation**

Structure delineation of the planning CTs was performed by the same radiation oncologist according to the European recommendations<sup>39</sup>. Positron emission tomography and MRI co-registration were used for tumor delineation. The gross tumor volume (GTV) corresponded to the primary tumor and involved lymph nodes. Three target volumes were generated for each patient. A clinical target volume receiving 70 Gy (CTV<sub>70</sub>, 35 fractions) was equal to the GTV plus a 5 mm 3D margin, adjusted to exclude any air cavities and bone mass that indicate no evidence of tumor invasion. CTV<sub>63</sub> (receiving 63 Gy in 35 fractions) corresponded to the high-risk area of the microscopic spread, whereas CTV<sub>56</sub> (receiving 56 Gy in 35 fractions) corresponded to the low-risk subclinical area. The GTV, CTV<sub>63</sub>, CTV<sub>56</sub>, and all the organs at risk, in particular the PGs and spinal cord, were manually delineated on each CT slice. The planning target volumes (PTV<sub>70Gy</sub>, PTV<sub>63Gy</sub>, and PTV<sub>56Gy</sub>) were generated by adding a 5 mm 3D margin around the CTVs.

For all patients, the total prescribed dose was 70 Gy (35 fractions) delivered by a simultaneous integrated boost (70-63-56 Gy) using VMAT technique. The dose calculation was performed with Pinnacle v.9.10 (Philips) treatment planning system (TPS) using the collapsed cone convolution algorithm and a dose grid resolution of 3 mm. GORTEC recommendations were used for organs at risk limitation doses<sup>40</sup>. The dose constraints for the contralateral PG were a mean dose lower than 30 Gy and a median dose lower than 26 Gy. For the ipsilateral PG, the mean dose was as low as possible. For the spinal cord, the maximum dose was lower than 45 Gy. The minimum PTV volume coverage by the 95% isodose was 95%.

On the pCT images generated by the evaluated methods, structures (target volumes, PGs and spinal cord) were propagated from the planning CTs, and the dose distributions were calculated after copying the beam parameters.

### **Intra-patient CBCT to CT registration**

To minimize the effects of anatomic variations between CT and CBCT, each CBCT was registered to its corresponding  $CT_{\text{initial}}$  using a rigid registration, followed by a non-rigid registration (NiftyReg). This  $CT_{\text{initial}}$  was considered as the reference ( $CT_{\text{ref}}$ ).

### **Deep learning method**

The GAN DLM architecture was composed of two networks: a generator (G) and a discriminator (D), which were trained in competition with each other (Fig. 1).

#### ***Generator network***

The generator network aims to provide a pCT from each patient's CBCT. In this study, the generator network used a 2D architecture similar to the U-Net DLM proposed by Han<sup>41</sup>. A single-scale perceptual loss function was used to train this network. This loss function mimics the human visual system to compare the CT and pCT images using similar features as opposed to only the intensities<sup>42,43</sup>. The features inside the CT and pCT images were computed using the Visual Geometry Group (VGG) 16 network, which is used for perceptual loss computation and appears relevant for various tasks (e. g., image deblurring, super-resolution, and computer vision)<sup>43,44</sup>. The perceptual loss function of the generator ( $L_G$ ) is defined as follows:

$$L_G(I, C) = \|VGG(C) - VGG(G(I))\|_2^2$$

where  $VGG$  is the output of the 7<sup>th</sup> VGG16 convolutional layer,  $I$  the CBCT, and  $C$  the corresponding CT.

#### ***Discriminator network***

The discriminator network aims to classify the pCT image as a real or fake CT. Hence, the output of this network is a probability value ranging between 0 and 1 depending on whether the pCT appears to be fake or real. The architecture was composed of six convolutional layers and one fully connected layer. Each convolutional layer was followed by batch normalization and Leaky-ReLu activation functions. The number of filters for these six layers were 8, 16, 32, 64, 64, and 64.

The filter size was  $3 \times 3$  (stride = 2) for the first four layers and  $1 \times 1$  (stride = 1) for the remaining layers. The fully connected layer had one filter followed by a sigmoid activation function.

The loss function of the discriminator ( $L_D$ ) was a binary cross entropy expressed as  $L_D(G(I), C) = -\sum_{i=1}^n C_i \log(G(I)_i) + (1 - C_i) \log(1 - G(I)_i)$ , where  $G(I)$  is the pCT computed by the generator from the target CBCT  $I$ , and  $n$  is the number of voxels inside  $C$  and  $I$  images.

The generator and discriminator losses were combined to form the following adversarial loss:  $L_{adversarial}(I, C) = \lambda_1 L_D(I, C) + \lambda_2 L_G(I, C)$ , where  $L_D(I, C)$  is the discriminator loss,  $L_G(I, C)$  is the generator loss, and  $\lambda_1$  and  $\lambda_2$  are the weights of the discriminator and generator losses, respectively. First, the discriminator was trained using the discriminator loss, followed by generator training using the fully adversarial loss. These training steps were performed iteratively until 300 epochs when the discriminator could not accurately determine if the pCTs provided by the generator were real or false CTs.

### Training data

The DLM was trained using 30 anatomically paired (CT-CBCT) data: axial 2D slices of the training CT and CBCT images (7600 slices). Data augmentation was performed to artificially increase the size of the training cohort. It was conducted by randomly applying affine registrations (by NiftyReg) on the slices (translated from -5% to 5% per axis, rotated from  $-10^\circ$  to  $+10^\circ$ , sheared from  $-10^\circ$  to  $10^\circ$ ). A mini-batch size of four slices and 300 epochs was considered. The network parameters were optimized using the Adam algorithm<sup>45</sup>. The parameters of this algorithm were as follows:  $\alpha = 1 \times 10^{-4}$ ,  $\beta_1 = 0.9$ , and  $\beta_2 = 0.9$ . For the GAN, the weights of the discriminator and generator loss functions were:  $\lambda_1 = 5$  and  $\lambda_2 = 1$ , respectively. The convergence curves of the GAN generator and discriminator are presented in Fig. S-1.

The DLM was implemented in Python using Keras<sup>46</sup>. The training computation time for the networks was approximately 72 h using an Nvidia GTX 1070 TI 8 GB GPU.

### Other CBCT-based dose calculation methods

The three other assessed methods are described following.

i. HU to density curve (HU-D curve) from phantom CBCT image

This method involves establishing an HU-D curve of the CIRS 062 phantom with clinical H&N CBCT protocol acquisition. Only the central part of the phantom was used to mimic the dimensions of the patient head. This phantom contained eight heterogeneous inserts and is considered appropriate for CBCT dose calculation owing to its size and chemical composition<sup>47</sup>. An HU-D H&N curve was then implemented into the TPS.

ii. Density assignment method (DAM)

An automatic thresholding was performed inside the patient body contour for all the CBCT and planning CT images based on HU values to create bone and air cavity classes. Manual corrections of the segmentation were necessary in case of dental artefacts and shoulder areas in the CBCT images. A soft tissue class was then created by the subtraction of air and bone classes within the patient body contour. The HU values of each class from the mean values of the planning CT were assigned to the corresponding classes of the CBCTs for each patient. Hence, three classes were obtained with the following values: -730 HU for the air cavity, 0 HU for the soft tissue, and 550 HU for the bone.

iii. CT-CBCT deformable image registration (DIR)

A deformed image was created with automatic DIR between the CBCT (fixed image) and planning CT (moving image) images using Admire software (research v.3.3.1, Elekta). A study was performed to assess the DIR accuracy of Admire (Appendix 1).

## Endpoints

Imaging and dosimetric endpoints were considered for the 14 patients (total of 64 CT-CBCTs).

### *Imaging endpoints*

A voxel-wise comparison of the HU between  $CT_{ref}$  and pCT was performed. The mean absolute error (MAE) and the mean error (ME) were calculated between the  $CT_{ref}$  and pCT obtained from all the methods. These endpoints were defined as follows:  $MAE = \frac{1}{n}$

$\sum_{i=1}^n |HU_{CTref}(i) - HU_{pCT}(i)|$  and  $ME = \frac{1}{n} \sum_{i=1}^n HU_{CTref}(i) - HU_{pCT}(i)$ , with  $n$  the total number of voxels.

### ***Dosimetric endpoints***

Owing to the limited size of the CBCT FOV, the water equivalent density was assigned inside the CT body contour when no CBCT or pCT information was available to perform an evaluation of the method. The accuracy of the methods was first evaluated by computing the dose and systematic dose discrepancies. The DVH calculated from the CT<sub>ref</sub> was subtracted from the DVH calculated from the pCT. The dose discrepancy was defined by the mean absolute dose and the systematic dose discrepancy by the mean DVH differences. The results were reported for the GORTEC reference DVH points ( $V_{95\%}$  for the PTVs and  $D_{2\%}$  for the spinal cord and mandible), and  $D_{mean}$  to the PGs was considered. A spatial dose evaluation was finally conducted by performing 3D gamma analyses (local, 2%/2 mm, low dose threshold: 30%) using the dose distributions from the CT<sub>ref</sub> and pCTs.

### ***Statistical analysis***

Wilcoxon signed-rank tests were performed to compare the endpoints. For the DVH comparisons based on the pCT generation methods, a nonparametric permutation test was performed<sup>48</sup> to control the presence of false positives in multiple statistical tests (10 cGy DVH bin-wise). In this case, 1000 permutations were performed, where for each permutation  $i$ , randomly selected DVHs were swapped (CT<sub>ref</sub> <->pCT) and the average difference was computed for each dose-bin. For each permuted sample and the original sample, the average difference was normalized to the standard deviation computed over all the 1000 permutations, and the maximum observed difference was selected as the test statistic ( $TS$ ). A distribution of  $TS$  across all the permuted samples ( $TS_{i,max}$ ) was obtained and compared with that one from the observed sample ( $TS_{max}$ ). The adjusted p-value was then computed as the probability of having a  $TS_{max}$  greater than the  $TS_{i,max}$  at the significance level of 5% ( $p \leq 0.05$ ). The corresponding percentile over the distribution of all the  $TS_{i,max}$  provides a threshold value that determines the dose DVH bins where a statistically significant dose difference arises. Unlike bin-wise tests, a permutation test provides a single

number that summarizes the discrepancy in the DVH between the two groups, rather than the discrepancy of a particular bin, which therefore accounts for multiple comparisons. The mathematical formulation of the permutation test has been provided by Chen et al.<sup>49</sup>. The test allowed us to report a robust bin-wise comparison based on the DVH value of each method as well as to compare the lowest MAE among all the methods with the MAE of each method and the ME of each method with a null distribution.

## Results

### Imaging endpoints and calculation time

Examples of CBCT,  $CT_{ref}$ , and pCT generated by each method are illustrated in Fig. 2. Table 2 lists the imaging endpoints for the entire body contour, soft tissue, and bone. In the entire body, the MAE and ME were 266.6 and 208.9 HU, 113.2 and 14.2 HU, 95.5 and -36.6 HU, and 82.4 and 17.1 HU for the HU-D curve method, DAM, DIR method, and DLM, respectively. The MAE results of the DLM differed significantly from those of other methods ( $p < 0.05$ ). The ME results of the four methods differed significantly from the null distribution.

The mean calculation time to generate one pCT was 30 s for the GAN DLM.

### Dosimetric endpoints

Fig. 3 shows the mean DVHs for the  $CT_{ref}$  and each method, by volume of interest. The DVHs of all methods differed significantly from that of  $CT_{ref}$ . Table 3 shows the  $D_{mean}$  differences of DVH points for the PTVs, spinal cord, and mandible and the  $D_{mean}$  to the PGs. The density assignment as well as DIR and DL methods provided dose discrepancies lower than 0.6% for the PTVs, PGs, and spinal cord. The lowest absolute mean  $D_{max}$  difference of the spinal cord was obtained using the DIR method ( $17 \pm 20$  cGy). The lowest absolute mean  $D_{mean}$  differences of the PGs were obtained using the DLM. The mean differences differed significantly from the null distribution for the PTV, spinal cord, and mandible DVH points in the four methods. The mean  $D_{mean}$  differences of the PGs did not significantly differ from the null distribution in the four methods. Fig. 4 shows the dose discrepancies of each method along the DVHs, by volume of interest. The method with the



highest discrepancies was the HU-D curve method. The density assignment and DIR methods were less significantly different from the DLM compared with the HU-D curve method. Fig. S-2 shows the systematic dose discrepancies (ME) of each method along the DVHs, by volume of interest. All methods showed significant differences with a null distribution.

Table 4 shows the mean gamma and gamma pass-rate values calculated from the  $CT_{ref}$  and pCT dose distributions for each method. The density assignment, DIR, and DL methods provided mean gamma pass-rates that exceeded 97.9%. The gamma pass-rates and mean gamma results of the HU-D curve method, DAM, and DIR method differed significantly from those of the DLM. Fig. S-3 shows examples of gamma maps in the coronal plane for one patient.

The last two columns of Table 1 are a qualitative analysis of the image and dose endpoints among the four compared methods.

## Discussion

The aim of this study was to compare a DLM (GAN) with three other methods (HU-D curve method, DAM, and DIR method) to perform dose calculation from H&N CBCT. Compared with the  $CT_{ref}$ , the DAM, DIR method, and DLM provided low dose discrepancies, thereby rendering them clinically acceptable for CBCT-based dose calculation (Fig. 3). Regarding dose accuracy as well as calculation time ( $< 30$  s), the DIR method and DLM appeared to be the most attractive methods (Table 1).

The main disadvantages of CBCT for dose calculation is HU inconsistency. Three studies have proposed comparing several CBCT dose calculation methods (scatter correction, DAM, DIR) but not with the DLM<sup>7,10,11</sup>. We conducted a comparison study using the DLM and three other methods (well-known methods in the literature) for CBCT dose calculation. This comparison study was performed using imaging and dose endpoints. Table 1 summarizes the benefits, drawbacks and qualitative comparison of the four investigated methods for H&N CBCT dose

calculation. The DLM provided the lowest MAE (Table 2) in the entire body (82.4 HU), soft tissue (69.2 HU), and bones (207.6 HU). However, the DLM did not show the lowest dose discrepancies. The relationship between image and dose uncertainties is particularly complex. In fact, the dose uncertainty in a voxel is not only related to the image uncertainty (HU) in the considered voxel, but also to several parameters, such as the image uncertainty of the surrounding voxels or dose gradients. For the bone tissues, the systematic error of the methods (HU–D curve method, DAM, and DLM) showed an underestimation of HU values (Table 2), whereas dose underestimation was recorded for the spinal cord (surrounded by bone) and mandible (Table 3). The density assignment and DIR methods provided better or similar dose results to the DLM (Table 3, Fig. 4, Fig. S-2, and Table 4). The method with the worst dose discrepancy (Table 3, Table 4, Fig. 4), except for the mandible, was the method using the HU–D conversion curve. This simple method appeared to be insufficient owing to artefacts and patient scattering<sup>50,51</sup>. Using this method, the main dose discrepancies were obtained at the shoulder level (Fig. S-3), where scatter artefacts are more present. The PTV 56 Gy was the most affected structure because of this issue (Fig. 2).

We proposed a DLM to generate pCT from CBCT to perform dose calculation. Such methods have already been used for pCT generation from MRI in the context of an MRI-only workflow<sup>52</sup>. The two recent MRI H&N studies using conditional GAN<sup>53</sup> or U-Net<sup>54</sup> provided an MAE of approximately 70–75 HU in the entire body contour. However, the aim of pCT generation from CBCT images is different. In fact, pCT generated from CBCT are used to monitor delivered doses or to estimate the cumulative delivered dose during the treatment course in the context of dose-guided adaptive radiotherapy. Studies using DL for pCT generation from CBCT are scarce for brain<sup>55</sup>, H&N<sup>13–15</sup>, pancreas<sup>38</sup> or prostate cancer<sup>35–37,55,56</sup>. The studies involved an imaging analysis (pCT versus reference CT), but only half of them evaluated the dose accuracy. Among the three H&N studies using DL for pCT generation from CBCT<sup>13–15</sup>, one involved training a U-Net neural network on 50 co-registered CBCT/CT images and performing a test based on data from 10 patients<sup>13</sup>. MAEs from 6 to 27 HU were obtained for the pCTs generated from Varian CBCTs. The dose differences were lower than 1%, and the mean gamma pass-rate (global, dose difference: 1%, DTA: 1 mm) was 95.5%. The second study involved training a cycle-consistent generative adversarial network (cycleGAN) on 81 CBCTs<sup>15</sup>. The evaluation performed on data from 20 patients provided a mean MAE of 29.9 HU. The mean gamma pass-rate (3D, dose difference: 1%,

DTA: 1 mm) was 96.3%. The cycleGAN DLM involved two GANs: one to generate pCT from CBCT and a second GAN to generate pCBCT from pCT (the output of the first GAN). These two dual GANs learn simultaneously to improve theoretically the efficiency. The cycleGAN DLM does not require paired data (no challenging intra-patient registration). Therefore, the amount of training data can largely be increased. However, the cycleGAN requires more complex training compared with the GAN because a larger number of parameters require optimization. The third study involved U-Net training on 37 patients to generate pCT from CBCT<sup>14</sup>, where only image analysis was performed. Furthermore, the MAE of 19 HUs for seven patients was obtained. We obtained a higher MAE with a larger number of patients and Elekta CBCT acquisitions. It is noteworthy that the comparison can only be indirect because the patient data differed in each study (CBCT imaging device, FOV size, tumor location, etc.).

Our study has some limitations. First, we trained the DL network with data from only 30 patients. Second, before the learning process, non-rigid registration was used to obtain the same H&N anatomy between CBCT and  $CT_{ref}$ , with the uncertainties depending of the deformable image registration algorithm. Third, the limited size of the CBCT FOV is still an issue. In this study, we added water equivalent density to override it and evaluate only the methods. Moreover, no scatter correction<sup>57</sup> was applied on CBCT images. Such image correction in a pre-processing step could decrease image and dose discrepancies. For image analysis, we did not consider metrics other than HU differences (MAE and ME) because the focus was on dose calculation. Image quality metrics such as Peak Signal-to-Noise Ratio (PSNR), Normalized Mutual Information (NMI), Structural SIMilarity (SSIM) or Visual Information Fidelity (VIF) should be implemented for further image analysis. For the DLM, our GAN parameters (e.g., loss function, VGG layer, discriminator weight, and mini-batch size) should be optimized to improve the image and dose accuracies. Moreover, the DL network was trained with 2D axial slices and not with fully 3D images because of memory GPU limitations. Finally, DL architectures such as the cycleGAN, which may have enabled some intra-individual co-registration issues to be overcome, could be further investigated.

## Conclusions

Four methods of H&N CBCT-based dose calculation were compared in this study. The use of an HU-D curve from CBCT was the simplest but the least accurate method. The DAM, use of DIR, and GAN DLM provided similar dose discrepancies. The DIR method and DLM appeared to be the most attractive methods in terms of dose accuracy as well as calculation time. Using one of such a method with CBCT images would enable dose monitoring in the PGs during the H&N treatment course to trigger replanning.

**Acknowledgments:**

With financial support from ITMO Cancer AVIESAN (Alliance Nationale pour les Sciences de la Vie et de la Santé/ National Alliance for Life Sciences & Health) within the framework of the Cancer Plan.

ADMIRE software was supplied by Elekta AB, Stockholm, Sweden as part of a research agreement.

**Conflict of Interest Statement:**

The authors have no relevant conflicts of interest to disclose.

## References

1. Brouwer CL, Steenbakkers RJHM, Langendijk JA, Sijtsema NM. Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? *Radiother Oncol.* 2015;115(3):285-294. doi:10.1016/j.radonc.2015.05.018
2. Castadot P, Lee JA, Geets X, Grégoire V. Adaptive Radiotherapy of Head and Neck Cancer. *Semin Radiat Oncol.* 2010;20(2):84-93. doi:10.1016/j.semradonc.2009.11.002
3. Veiga C, McClelland J, Moinuddin S, et al. Toward adaptive radiotherapy for head and neck patients: Feasibility study on using CT-to-CBCT deformable registration for “dose of the day” calculations. *Med Phys.* 2014;41(3):031703.
4. Richter A, Hu Q, Steglich D, et al. Investigation of the usability of conebeam CT data sets for dose calculation. *Radiat Oncol.* 2008;3(1):42. doi:10.1186/1748-717X-3-42
5. Fotina I, Hopfgartner J, Stock M, Steininger T, Lütgendorf-Caucig C, Georg D. Feasibility of CBCT-based dose calculation: Comparative analysis of HU adjustment techniques. *Radiother Oncol.* 2012;104(2):249-256. doi:10.1016/j.radonc.2012.06.007
6. Yoo S, Yin F-F. Dosimetric feasibility of cone-beam CT-based treatment planning compared to CT-based treatment planning. *Int J Radiat Oncol.* 2006;66(5):1553-1561. doi:10.1016/j.ijrobp.2006.08.031
7. Dunlop A, McQuaid D, Nill S, et al. Comparison of CT number calibration techniques for CBCT-based dose calculation. *Strahlenther Onkol.* 2015;191(12):970-978. doi:10.1007/s00066-015-0890-7
8. van Zijtveld M, Dirkx M, Heijmen B. Correction of conebeam CT values using a planning CT for derivation of the “dose of the day.” *Radiother Oncol.* 2007;85(2):195-200. doi:10.1016/j.radonc.2007.08.010
9. Barateau A, Perichon N, Castelli J, et al. A density assignment method for dose monitoring in head-and-neck radiotherapy. *Strahlenther Onkol.* Published online October 9, 2018. doi:10.1007/s00066-018-1379-y

- Accepted Article
10. Giacometti V, King RB, Agnew CE, et al. An evaluation of techniques for dose calculation on cone beam computed tomography. *Br J Radiol.* 2019;92(1096):20180383. doi:10.1259/bjr.20180383
  11. Marchant TE, Joshi KD, Moore CJ. Accuracy of radiotherapy dose calculations based on cone-beam CT: comparison of deformable registration and image correction based methods. *Phys Med Biol.* Published online February 20, 2018. doi:10.1088/1361-6560/aab0f0
  12. Kurz C, Kamp F, Park Y-K, et al. Investigating deformable image registration and scatter correction for CBCT-based dose calculation in adaptive IMPT: CBCT correction to enable IMPT dose calculation. *Med Phys.* 2016;43(10):5635-5646. doi:10.1118/1.4962933
  13. Li Y, Zhu J, Liu Z, et al. A preliminary study of using a deep convolution neural network to generate synthesized CT images based on CBCT for adaptive radiotherapy of nasopharyngeal carcinoma. *Phys Med Biol.* Published online June 6, 2019. doi:10.1088/1361-6560/ab2770
  14. Chen L, Liang X, Shen C, Jiang S, Wang J. Synthetic CT generation from CBCT images via deep learning. *Med Phys.* Published online January 13, 2020. doi:10.1002/mp.13978
  15. Liang X, Chen L, Nguyen D, et al. Generating Synthesized Computed Tomography (CT) from Cone-Beam Computed Tomography (CBCT) using CycleGAN for Adaptive Radiation Therapy. :21.
  16. Schulze R, Heil U, Groß D, et al. Artefacts in CBCT: a review. *Dentomaxillofacial Radiol.* 2011;40(5):265-273. doi:10.1259/dmfr/30642039
  17. Rigaud B, Simon A, Castelli J, et al. Deformable image registration for radiation therapy: principle, methods, applications and evaluation. *Acta Oncol.* Published online June 3, 2019:1-13. doi:10.1080/0284186X.2019.1620331
  18. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436-444. doi:10.1038/nature14539
  19. Meyer P, Noblet V, Mazzara C, Lallement A. Survey on deep learning for radiotherapy. *Comput Biol Med.* 2018;98:126-146. doi:10.1016/j.compbiomed.2018.05.018
  20. Higaki T, Nakamura Y, Tatsugami F, Nakaura T, Awai K. Improvement of image quality at CT and MRI using deep learning. *Jpn J Radiol.* Published online November 29, 2018. doi:10.1007/s11604-018-0796-2

21. Alkadi R, Taher F, El-baz A, Werghi N. A Deep Learning-Based Approach for the Detection and Localization of Prostate Cancer in T2 Magnetic Resonance Images. *J Digit Imaging*. Published online November 30, 2018. doi:10.1007/s10278-018-0160-1
22. Laukamp KR, Thiele F, Shakirin G, et al. Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI. *Eur Radiol*. 2019;29(1):124-132. doi:10.1007/s00330-018-5595-8
23. Liang S, Tang F, Huang X, et al. Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning. *Eur Radiol*. Published online October 9, 2018. doi:10.1007/s00330-018-5748-9
24. Nyflot MJ, Thammasorn P, Wootton LS, Ford EC, Chaovalitwongse WA. Deep learning for patient-specific quality assurance: Identifying errors in radiotherapy delivery by radiomic analysis of gamma images with convolutional neural networks. *Med Phys*. Published online December 13, 2018. doi:10.1002/mp.13338
25. Sahiner B, Pezeshk A, Hadjiiski LM, et al. Deep learning in medical imaging and radiation therapy. *Med Phys*. 2019;46(1):e1-e36. doi:10.1002/mp.13264
26. Gong K, Yang J, Kim K, El Fakhri G, Seo Y, Li Q. Attenuation correction for brain PET imaging using deep neural network based on Dixon and ZTE MR images. *Phys Med Biol*. 2018;63(12):125011. doi:10.1088/1361-6560/aac763
27. Ladefoged CN, Marner L, Hindsholm A, Law I, Højgaard L, Andersen FL. Deep Learning Based Attenuation Correction of PET/MRI in Pediatric Brain Tumor Patients: Evaluation in a Clinical Setting. *Front Neurosci*. 2019;12. doi:10.3389/fnins.2018.01005
28. Leynes AP, Yang J, Wiesinger F, et al. Zero-Echo-Time and Dixon Deep Pseudo-CT (ZeDD CT): Direct Generation of Pseudo-CT Images for Pelvic PET/MRI Attenuation Correction Using Deep Convolutional Neural Networks with Multiparametric MRI. *J Nucl Med*. 2018;59(5):852-858. doi:10.2967/jnumed.117.198051
29. Torrado-Carvajal A, Vera-Olmos J, Izquierdo-Garcia D, et al. Dixon-VIBE Deep Learning (DIVIDE) Pseudo-CT Synthesis for Pelvis PET/MR Attenuation Correction. *J Nucl Med*. 2019;60(3):429-435. doi:10.2967/jnumed.118.209288

30. Kläser K, Markiewicz P, Ranzini M, et al. Deep Boosted Regression for MR to CT Synthesis. In: Gooya A, Goksel O, Oguz I, Burgos N, eds. *Simulation and Synthesis in Medical Imaging*. Vol 11037. Springer International Publishing; 2018:61-70. doi:10.1007/978-3-030-00536-8\_7
31. Nyholm T, Svensson S, Andersson S, et al. MR and CT data with multiobserver delineations of organs in the pelvic area-Part of the Gold Atlas project. *Med Phys*. 2018;45(3):1295-1300. doi:10.1002/mp.12748
32. Yi X, Walia E, Babyn P. Generative Adversarial Network in Medical Imaging: A Review. *Med Image Anal*. 2019;58:101552. doi:10.1016/j.media.2019.101552
33. Maier J, Eulig E, Vöth T, et al. Real-time scatter estimation for medical CT using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation. *Med Phys*. 2019;46(1):238-249. doi:10.1002/mp.13274
34. Nomura Y, Xu Q, Shirato H, Shimizu S, Xing L. Projection-domain scatter correction for cone beam computed tomography using a residual convolutional neural network. *Med Phys*. Published online June 5, 2019. doi:10.1002/mp.13583
35. Landry G, Hansen D, Kamp F, et al. Comparing Unet training with three different datasets to correct CBCT images for prostate radiotherapy dose calculations. *Phys Med Biol*. Published online November 28, 2018. doi:10.1088/1361-6560/aaf496
36. Hansen DC, Landry G, Kamp F, et al. ScatterNet: A convolutional neural network for cone-beam CT intensity correction. *Med Phys*. 2018;45(11):4916-4926. doi:10.1002/mp.13175
37. Kurz C, Maspero M, Savenije MHF, et al. CBCT correction using a cycle-consistent generative adversarial network and unpaired training to enable photon and proton dose calculation. *Phys Med Biol*. Published online October 14, 2019. doi:10.1088/1361-6560/ab4d8c
38. Liu Y, Lei Y, Wang T, et al. CBCT-based Synthetic CT Generation using Deep-attention CycleGAN for Pancreatic Adaptive Radiotherapy. *Med Phys*. Published online March 6, 2020. doi:10.1002/mp.14121



39. Brouwer CL, Steenbakkers RJHM, Bourhis J, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother Oncol*. 2015;117(1):83-90. doi:10.1016/j.radonc.2015.07.041
40. Toledano I, Graff P, Serre A, et al. Intensity-modulated radiotherapy in head and neck cancer: Results of the prospective study GORTEC 2004–03. *Radiother Oncol*. 2012;103(1):57-62. doi:10.1016/j.radonc.2011.12.010
41. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44(4):1408-1419. doi:10.1002/mp.12155
42. Nie D, Trullo R, Lian J, et al. Medical Image Synthesis with Context-Aware Generative Adversarial Networks. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017*. Lecture Notes in Computer Science. Springer, Cham; 2017:417-425. doi:10.1007/978-3-319-66179-7\_48
43. Yang Q, Yan P, Zhang Y, et al. Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss. *IEEE Trans Med Imaging*. 2018;37(6):1348-1357. doi:10.1109/TMI.2018.2827462
44. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv14091556* Cs. Published online September 4, 2014. Accessed October 2, 2018. <http://arxiv.org/abs/1409.1556>
45. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *ArXiv14126980* Cs. Published online December 22, 2014. Accessed September 24, 2018. <http://arxiv.org/abs/1412.6980>
46. *Deep Learning for Humans. Contribute to Keras-Team/Keras Development by Creating an Account on GitHub*. Keras; 2018. Accessed September 24, 2018. <https://github.com/keras-team/keras>
47. Guan H, Dong H. Dose calculation accuracy using cone-beam CT (CBCT) for pelvic adaptive radiotherapy. *Phys Med Biol*. 2009;54(20):6239-6250. doi:10.1088/0031-9155/54/20/013
48. Ross SM. Chapter 14-Nonparametric Hypotheses Tests. In: *Introductory Statistics (fourth edition)*; 2017, Sheldon M. Ross (Eds), Academic Press, 621-666.
49. Chen C, Witte M, Heemsbergen W, Herk M van. Multiple comparisons permutation test for image based data mining in radiotherapy. *Radiat Oncol*. 2013;8(1). doi:10.1186/1748-717X-8-293

50. Schröder L, Stankovic U, Remeijer P, Sonke J-J. Evaluating the impact of cone-beam computed tomography scatter mitigation strategies on radiotherapy dose calculation accuracy. *Phys Imaging Radiat Oncol*. 2019;10:35-40. doi:10.1016/j.phro.2019.04.001
51. Schröder L, Stankovic U, Sonke J. Technical Note: Long-term stability of Hounsfield unit calibration for cone-beam computed tomography. *Med Phys*. Published online January 10, 2020. doi:10.1002/mp.14015
52. Wang T, Lei Y, Fu Y, Curran WJ, Liu T, Yang X. Medical Imaging Synthesis using Deep Learning and its Clinical Applications: A Review. *ArXiv200410322 Phys*. Published online April 21, 2020. Accessed April 29, 2020. <http://arxiv.org/abs/2004.10322>
53. Qi M, Li Y, Wu A, et al. Multi-sequence MR image-based synthetic CT generation using a generative adversarial network for head and neck MRI-only radiotherapy. *Med Phys*. Published online February 6, 2020. doi:10.1002/mp.14075
54. Dinkla AM, Florkow MC, Maspero M, et al. Dosimetric evaluation of synthetic CT for head and neck radiotherapy generated by a patch-based 3D convolutional neural network. *Med Phys*. Published online June 17, 2019. doi:10.1002/mp.13663
55. Harms J, Lei Y, Wang T, et al. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. *Med Phys*. Published online July 17, 2019. doi:10.1002/mp.13656
56. Kida S, Kaji S, Nawa K, et al. Visual enhancement of Cone-beam CT by use of CycleGAN. *Med Phys*. Published online January 3, 2020. doi:10.1002/mp.13963
57. Rührschopf E-P, Klingensbeck K. A general framework and review of scatter correction methods in x-ray cone-beam computerized tomography. Part 1: Scatter compensation approaches: Scatter compensation approaches. *Med Phys*. 2011;38(7):4296-4311. doi:10.1118/1.3599033

## Figure captions

### **Fig. 1. Deep learning method for pseudo-CT generation from CBCT**

*The training of the generative adversarial network (GAN) comprises two competing multilayer networks: the generator and the discriminator. The discriminator aims to distinguish the real image (ground truth) from the realistic fake image (pCT) produced by the generator. The input data of the generator are CBCT and CT images that provide pCTs. The discriminator classifies these pCTs as real or fake CTs until the discriminator cannot determine whether the pCT is a real CT. In the testing step, for a new test patient, the CBCT goes through the trained network to obtain the corresponding pCT.*

### **Fig. 2. CBCT, reference CT, and pseudo-CT images from the deep learning method**

### **Fig. 3. Mean DVHs PTV, parotid glands, and spinal cord**

*Permutation tests were performed to compare the DVHs from the reference CT to those of the four methods. Significant differences ( $p < 0.05$ ) between the DVHs are displayed at the top of each figure using the symbol \*.*

### **Fig. 4. Dose discrepancies for CBCT-based dose calculation methods along the entire DVH for the PTV, PGs, and spinal cord**

*The dose discrepancy is defined as the mean absolute DVH differences between the reference CT and the pCT corresponding to each method. Permutation tests were performed to compare the absolute DVH differences of the DLM to those of the other methods. Significant differences ( $p \leq 0.05$ ) are displayed at the top of each figure with \*.*

### **Fig. S-1. Convergence of generator and discriminator loss functions on the training cohort**

### **Fig. S-2. Systematic dose discrepancy of each CBCT-based dose calculation method**

*The systematic dose discrepancy was defined as the mean DVH differences between the reference CT and the pCTs generated by each method. Permutation tests were performed to compare the DVH differences of each method to a null distribution. Significant differences ( $p \leq 0.05$ ) are displayed at the top of each figure using the symbol \*.*

**Fig. S-3. Illustration of pseudo-CTs, dose distributions, and gamma maps for one patient in coronal plane**

*The dose distributions are displayed for each of the four investigated methods. The dose differences are illustrated in terms of gamma analysis. 3D gamma pass-rate values corresponding to each method are indicated for this patient.*

**Table 1. Advantages, drawbacks, and main results of each investigated method for head-and-neck CBCT dose calculation**

<b>Methods</b>	<b>Advantages*</b>	<b>Drawbacks*</b>	<b>Image comparison<sup>#</sup></b>	<b>Dose comparison<sup>#</sup></b>
<b>HU-density curve method</b>	- Simplicity	- Subject to artefacts	+	++
<b>Density assignment method (DAM)</b>	- Simplicity	- Dependent of structure segmentation - Homogeneity of the tissues / limited number of tissue classes (1 to 6)	++	+++
<b>Deformable image registration (DIR) method</b>	- Fully automatic - Heterogeneity of the tissues - Fast computation time	- Not robust to large anatomical variations (tumor shrinkage, weight loss, etc.)	+++	+++
<b>Deep learning method (DLM)</b>	- Fully automatic - Heterogeneity of the tissues - Fast computation time	- Require cohort training (same imaging device, and size-dependent) - Paired data (except cycleGAN)	++++	+++

\*: From literature; #: From our study results

In the two last columns, “+” represents the qualitative comparison of the four methods in terms of image and dose endpoints. The more the number of “+”, the more accurate is the method. The dose endpoints are the DVH and gamma results.

**Table 2. Imaging endpoints comparing the reference CT to the pseudo-CTs obtained by each method**

	<b>Endpoints</b>	<b>HU-D curve method</b>	<b>Density assignment method</b>	<b>DIR method</b>	<b>Deep learning method</b>
Entire body	MAE (HU)	266.6* ± 25.8	113.2* ± 6.7	95.5* ± 21.2	82.4 ± 10.6
	ME (HU)	208.9* ± 36.1	14.6* ± 11.2	-36.6* ± 14.7	17.1* ± 19.9
Soft tissue only	MAE (HU)	260.9* ± 25.9	91.2* ± 6.4	85.0* ± 20.2	69.2 ± 15.3
	ME (HU)	206.4* ± 35.7	9.6* ± 11.0	-33.3* ± 14.7	12.9* ± 19.4
Bone only	MAE (HU)	344.6* ± 64.7	388.1* ± 38.3	226.2* ± 44.4	207.6 ± 41.8
	ME (HU)	253.1* ± 127.2	84.7* ± 77.6	-77.4* ± 28.6	64.9* ± 56.1

*MAE: mean absolute error of HU values defined as the mean difference (in absolute value) of HU values per voxel between the reference CT and pseudo-CT; ME: mean error, defined as the mean difference of HU values per voxel between the reference CT and the pseudo-CT.*

*The imaging endpoint values are expressed as mean ± standard deviation.*

*The Wilcoxon test was used to first compare the MAE of the DLM to those of the other methods and then to compare the ME of the methods to a null distribution. Significant differences ( $p \leq 0.05$ ) are displayed using the symbol \*.*

**Table 3. Reference dose values, dose discrepancies, and systematic dose discrepancies for each volume of interest**

Volumes of interest		PTV			Parotid gland		Spinal cord	Mandible
		70 Gy	63 Gy	56 Gy	Ipsilateral	Contralateral		
Dosimetric endpoints		V <sub>95%</sub> (%)			D <sub>mean</sub> (cGy)		D <sub>max</sub> (cGy)	
CT <sub>ref</sub> values		99.2 ± 0.7	95.2 ± 1.2	98.7 ± 1.3	2633 ± 995	1958 ± 507	4087 ± 112	6584 ± 485
Absolute mean differences (dose discrepancy)	HU-density curve method	0.2 ± 0.2	0.7 ± 0.5*	0.3 ± 0.4	14 ± 13	9 ± 8	38 ± 24	102 ± 56
	Density assignment method	0.2 ± 0.2	0.4 ± 0.3	0.4 ± 0.7	7 ± 9	6 ± 6	23 ± 19*	117 ± 79*
	Deformable image registration	0.3 ± 0.4*	0.6 ± 0.5	0.5 ± 0.7	10 ± 10	8 ± 7	17 ± 20*	28 ± 45*
	Deep learning method	0.2 ± 0.2	0.4 ± 0.4	0.7 ± 1.3	7 ± 8	5 ± 6	42 ± 17	99 ± 54
Mean differences (systematic dose discrepancy)	HU-density curve method	-0.1 ± 0.3*	-0.6 ± 0.6*	-0.3 ± 0.5*	-12 ± 15	-7 ± 10	-29 ± 75*	-90 ± 35*
	Density assignment method	-0.1 ± 0.2*	-0.1 ± 0.5*	0.4 ± 0.8*	-3 ± 11	1 ± 8	-20 ± 23*	-117 ± 79*
	Deformable image registration	0.1 ± 0.5*	0.3 ± 0.7*	0.5 ± 0.8*	4 ± 13	6 ± 9	8 ± 25*	3 ± 54

	<b>Deep learning method</b>	$-0.1 \pm 0.3^*$	$-0.2 \pm 0.5^*$	$0.6 \pm 1.4^*$	$-4 \pm 10$	$0 \pm 8$	$-42 \pm 17^*$	$-96 \pm 59^*$
--	-----------------------------	------------------	------------------	-----------------	-------------	-----------	----------------	----------------

*The mean values of the DVH points are reported for the reference CT. The dose discrepancy is defined as the mean absolute DVH differences between the DVH calculated from the reference CT and those obtained from the different methods. The systematic dose discrepancy is defined as the mean DVH differences between the DVH calculated from the reference CT and those obtained from the different methods. The Wilcoxon test was used to: compare the dose discrepancy of the DLM with those of the other methods and to compare the systematic dose discrepancy of the methods to a null distribution. Significant differences ( $p \leq 0.05$ ) are displayed using the symbol \*.*

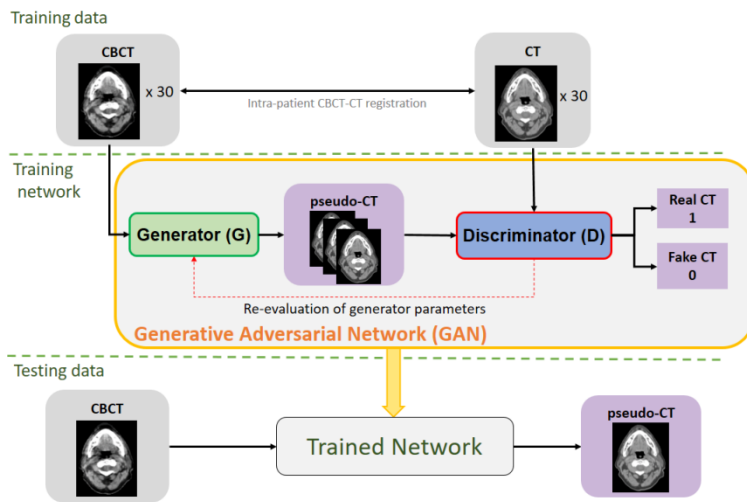


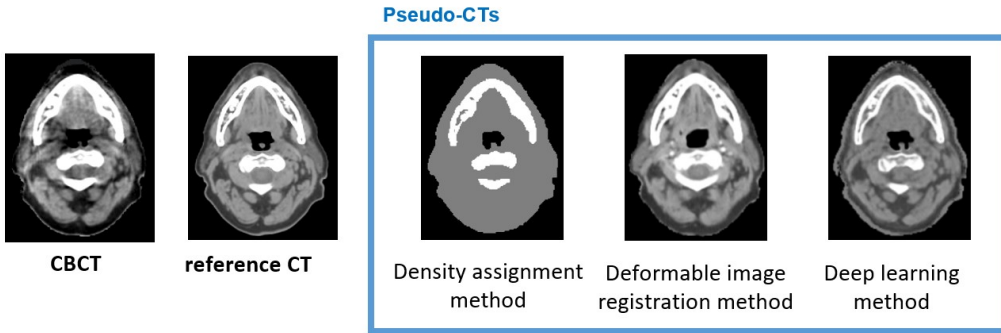
**Table 4. Mean gamma and gamma pass-rate calculated from reference CT and pseudo-CT dose distributions according to each method**

	<b>HU-density curve method</b>	<b>Density assignment method</b>	<b>Deformable image registration</b>	<b>Deep learning method</b>
Gamma pass-rate (%)	91.0 ± 5.3*	97.9 ± 1.6*	98.8 ± 0.7*	98.1 ± 1.2
Mean gamma	0.66 ± 0.18*	0.47 ± 0.17*	0.36 ± 0.10*	0.54 ± 0.13

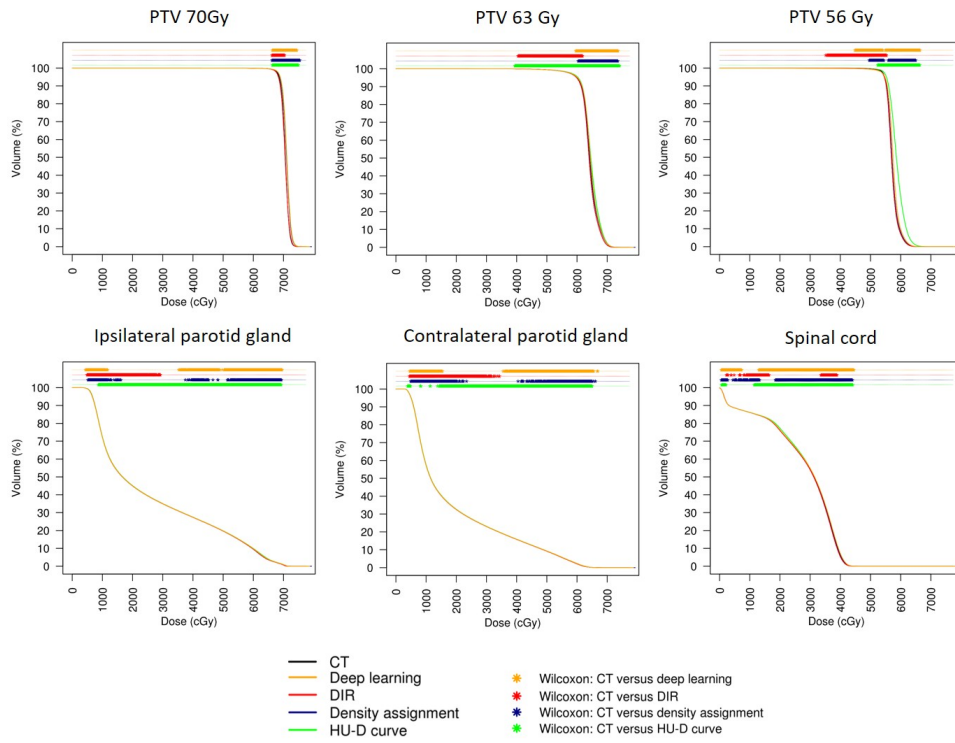
*Values are mean ± standard deviation.*

*The Wilcoxon test was performed to compare the gamma values of the DLM to those of the other methods. Significant differences ( $p \leq 0.05$ ) are displayed using the symbol \*.*

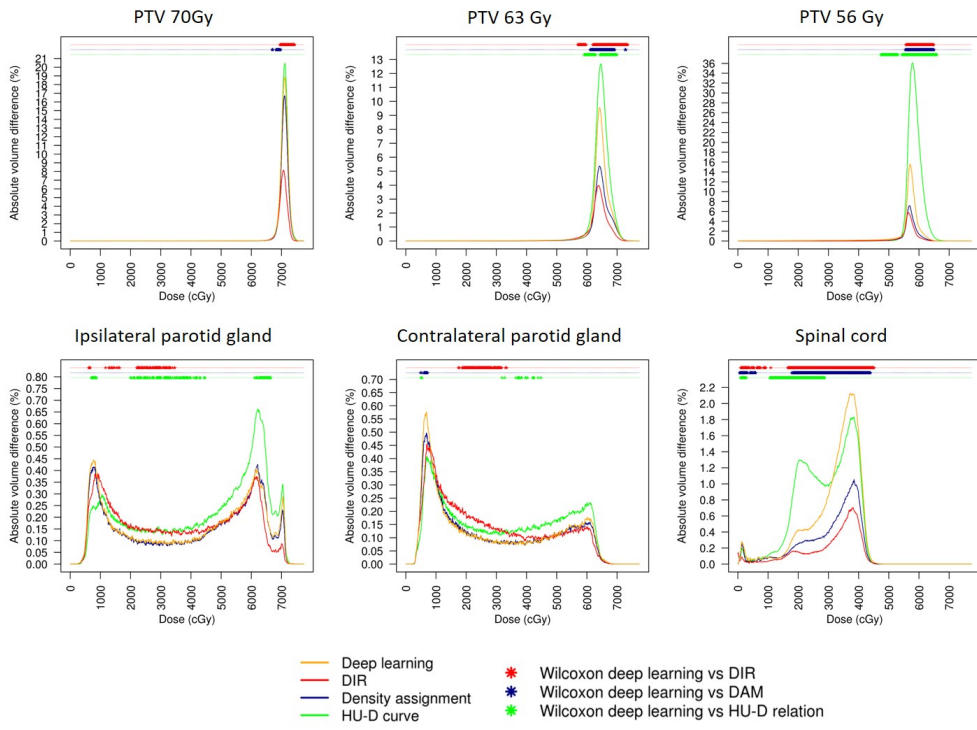




mp\_14387\_f2.jpg



mp\_14387\_f3.jpg



mp\_14387\_f4.jpg