



Deep-learning synthetized 4DCT from 4DMRI of the abdominal site in carbon-ion radiotherapy



Anestis Nakas ^{a,*}, Maksym Hladchuk ^a, Giovanni Parrella ^a, Alessandro Vai ^b, Silvia Molinelli ^b, Francesca Camagni ^a, Viviana Vitolo ^c, Amelia Barcellini ^{c,e}, Sara Imparato ^c, Andrea Pella ^d, Mario Ciocca ^b, Ester Orlandi ^c, Chiara Paganelli ^{a,1}, Guido Baroni ^{a,d,1}

^a Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

^b Medical Physics Unit, National Center of Oncological Hadrontherapy (CNAO), Pavia, Italy

^c Clinical Unit, National Center of Oncological Hadrontherapy (CNAO), Pavia, Italy

^d Bioengineering Unit, National Center of Oncological Hadrontherapy (CNAO), Pavia, Italy

^e Department of Internal Medicine and Therapeutics, University of Pavia, Pavia, Italy

ARTICLE INFO

Keywords:

synthetic 4DCT
4DMRI guidance
MRI-only
Image-guided particle therapy
Carbon ion radiotherapy
Deep learning

ABSTRACT

Purpose: To investigate the feasibility of deep-learning-based synthetic 4DCT (4D-sCT) generation from 4DMRI data of abdominal patients undergoing Carbon Ion Radiotherapy (CIRT).

Material and methods: A 3-channel conditional Generative Adversarial Network (cGAN) was trained and tested on twenty-six patients, using paired T1-weighted 4DMRI and 4DCT volumes. 4D-sCT data were generated via the cGAN following a 3-channels segmentation approach (air, bone, soft tissue) in two scenarios: (a) 4DCT-based approach (i.e. segmentation relying on 4DCT) and (b) 4DMRI-based approach (i.e. manual segmentation on 4DMRI, to simulate a 4DMRI-only scenario). The network was first validated on a 4D computational phantom, where a ground truth dataset was available. Subsequently, the network was tested on 6 independent held-out-of-training patients. Generated volumes were evaluated with respect to the original 4DCT based on motion analysis, similarity metrics (e.g. Mean Absolute Error (MAE), Normalized Cross Coefficient (NCC)) and dosimetric criteria, by means of recalculating clinically optimized CIRT plans on the 4D-sCT.

Results: For the phantom, similarity metrics were in line with literature results, while dose volume histogram values were below 0.9 %. 4DCT-based patient results demonstrated an accurate representation with respect to the original 4DCT images (MAE: 50.64–51.29 HU), while 4DMRI-only-based results yielded higher values (MAE: 81.15–90.22 HU). Gamma pass rates (3 %/3mm) were ~ 97 % for the 4DCT-based scenario, showing dosimetric consistency between the compared 4DCT and 4D-sCT dose distributions. D95% values on GTV/CTV were within clinical tolerances for the 4DMRI-only scenario.

Conclusion: Deep learning-based 4D-sCT generation shows potential to support treatment planning in abdominal tumors treated with CIRT.

1. Introduction

Carbon-Ion Radiotherapy (CIRT) is an effective modality for deep seated and radioresistant tumors, offering superior radiation dose conformity and radiobiological effectiveness compared to conventional photon radiotherapy [1]. However, the successful application of CIRT demands precise tumor targeting, which can be challenging for lesions subject to respiratory motion [2,3]. Density variations created along the beam path could result in range uncertainties, thus leading to tumor

underdosage and healthy tissue overdosage.

Those challenges highlight the importance of accounting for the dynamic nature of respiration in CIRT [3,4]. The quest for precision and efficacy has led to the integration of advanced medical imaging techniques within the treatment workflow. Four-dimensional computed tomography (4DCT) has been a cornerstone in addressing respiratory motion during treatment planning. However, 4DCT is associated with additional radiation exposure, low contrast and potential image artifacts during retrospective sorting. Furthermore, the consequent need for re-

* Corresponding author.

¹ Shared co-last authorship.

evaluating 4DCT sessions throughout the treatment course to account for inter-fraction variations further imposes the need for alternative motion management techniques or modalities.

Research efforts have focused on 4D Magnetic Resonance Imaging (4DMRI) as a non-ionizing alternative that provides improved soft tissue contrast and dynamic acquisitions [5]. However, challenges arise as treatment planning systems are predominantly designed to work with 4DCT datasets. Indeed, MRI data do not provide electron density information which is essential to create treatment plans. Addressing this discrepancy, efforts have been focused on the generation of synthetic CT (sCT) from MRI, aiming to support treatment planning and adaptation by seamlessly integrating the benefits of both modalities for more precise treatments, while limiting non-therapeutic radiation exposure.

In this regard, Deep Learning (DL) solutions have been thoroughly investigated in literature for sCT generation from MRI data. All studies, however, have mainly focused on Photon or Proton Radiotherapy applications and static anatomies. In specific, Convolutional Neural Networks (CNN) – especially U-net architectures, conditional Generative Adversarial Networks (cGAN) and cycle GAN, have shown promising results for head and neck or pelvic/prostate cases [6,7,8,9,10,11,12,13,14]. Regarding dynamic anatomies (abdomen or thorax) efforts have been more challenging, due to the anatomical complexity and heterogeneity in electron density [15,16,17,18,19,20,21]. So far, a limited amount of research groups have investigated applications in CIRT [22,23]. Parrella et al. [22] utilized a 3-channel cGAN to generate abdominal 3D-sCT from end-exhale T1-weighted MRI for application in CIRT. The results demonstrated the networks' potential to generate accurate abdominal sCT, allowing CIRT dose recalculations comparable to clinical plans. To the authors' best knowledge, no study has investigated the feasibility of abdominal 4D-sCT, DL-based generation for application in CIRT so far.

Starting from the work proposed in [23,24], we utilized a 3-channel cGAN [22] to generate 4D-sCT from 4DMRI data. The network was trained and tested on abdominal 4DCT and virtual T1-weighted (T1-w) 4DMRI (v4DT1) data of patients undergoing CIRT at the National Center for Oncologic Handronotherapy (CNAO) in Italy. The generated 4D-sCT data were evaluated with respect to the original 4DCT based on motion analysis, similarity metrics and dosimetric criteria, by means of recalculating clinically optimized CIRT plans on the 4D-sCT. Due to the lack of a proper patient ground truth, a 4D CT/MRI computational phantom was used to validate the network's performance.

2. Materials and Methods

2.1. Patient data

4DCT and 4DMRI datasets were acquired for 26 abdominal (pancreatic/liver) cancer patients, treated with CIRT at CNAO using the respiratory gating technique. For standard patient setup purposes, customized pillows (MOLDCARE Cushion, QFix, Avondale, PA, USA) and non-perforated thermoplastic abdominal masks (Klarity Medical Products, Health, OH, USA) were used in the treatment delivery, as well as the 4DCT-4DMRI imaging sessions.

Planning 4DCT data were acquired on free breathing on a Siemens SOMATOM Sensation Open CT scanner with a resolution of $0.98 \times 0.98 \times 2 \text{ mm}^3$. A pressure sensor (AZ-733 V system, Anzai Medical Co. Ltd., Japan) was placed between the patient's abdomen and the thermoplastic mask, to provide a surrogate respiratory signal for retrospective sorting. Four respiratory phases were reconstructed, namely the end-exhale (0EX), the 30 %-exhale (30EX), 30 %-inhale (30IN) and 100 %-inhale (100IN).

4DMRI were acquired on a Siemens Magnetom Verio 3 T scanner on the same day as the 4DCT. T2/T1-weighted (T2/T1-w) balanced steady-state free precession sequences (TrueFISP sequence, pixel spacing: $1.33 \times 1.33 \text{ mm}$; slice thickness 5 mm; repetition time: 228.07 ms; echo time: 1.5 ms; accelerating factor: 2; acquisition time: 230 ms) were acquired to

obtain multi-slice sagittal images of the abdominal area during free breathing. A limited field of view in the right-left direction (12.5 cm) was covered due to temporal-spatial trade-off [25]. The same four respiratory phases as in the 4DCT were retrospectively reconstructed following the approach proposed by Meschini et al. [25]. Additionally, a T1-w volumetric interpolated breath-hold examination (VIBE) sequence was acquired at end-exhale as reference 3D OEX T1-w image, with $1.06 \times 1.06 \times 3 \text{ mm}^3$ resolution (repetition time: 3.87 ms, echo time: 1.92 ms).

Treatment plans were optimized with standard robustness optimization parameters of the abdominal site (3 mm set-up; 3 % range uncertainties) on the OEX-CT and recalculated within the gating window (30EX/30IN-CT) for clinical approval in CNAO. The Raystation (Raysearch Laboratories, Sweden) Treatment Planning System (TPS) was used. CIRT plan information are summarized in Table 1 for testing patients along with the range of motion of the liver, calculated in the superior – inferior (SI) orientation.

2.2. Data preparation

The reference OEX T1-w images presented higher contrast than the reconstructed T2-w 4DMRI. For this reason, by applying the Deformable Image Registration (DIR) approach described in [26,27], the T2-w 4DMRI motion field was imposed on the reference T1-w volume to generate v4DT1 data of the following phases: 0EX, 30EX, 30IN and 100IN (Fig. 1A). DIR was performed using a multiresolution B-splines algorithm (plastimatch.org).

Then, all 4DCT and v4DT1 volume pairs were pre-processed to improve their quality before training. Although 4DCT and 4DMRI were acquired minimizing the time interval between the scans on the same day and with the same setup (patient-specific cushion and thermoplastic mask), residual differences in patient positioning and anatomical variations (inter-acquisition motion) might occur between CT-MRI scans. First, a manual rigid alignment between CT – T1-w scans with respect to the spinal cord was performed. Application of DIR was excluded as in [22,28]. Registered CT slices were resampled to T1-w MRI and clipped to $[-1024, +1047] \text{ HU}$, while background values were set at -1024 HU . Similarly to [22], for the preprocessing of the v4DT1 the following steps were utilized: bias field correction, Gaussian noise reduction, contrast enhancement, histogram matching and setting background values to zero.

To match the network's input criteria, all pre-processed volumes were resized to $256 \times 256 \times 64$ [29]. To facilitate better results and account for the anatomical complexity of the abdominal region, a 3-channel input approach was adopted as in [22]. In more detail, 4DCT-based masks were applied on the input 4DCT and v4DT1 transversal slices, obtaining a 3-channel segmentation of air, bone and soft tissue ($[-1024, -800] \text{ HU}$, $[150, 1047] \text{ HU}$ and $[-800, 150] \text{ HU}$, respectively). Each channel was further normalized to $[-1, 1]$.

Additionally, to increase the diversity of the training set and facilitate the model's generalization on unseen data, data augmentation was performed on each 4DCT/v4DT1 training slices. Seven transformations (Horizontal/Vertical Flip, addition of Gaussian Noise, Shearing, Rotation and Cropping) were applied on each input, resulting in a training set

Table 1

CIRT dose plan information for the test patients. The liver's motion in the superior-inferior direction as quantified from the 4DCT is also reported.

Patient	Prescribed RBE Dose (Gy)	Fractions	Beams	Tumor	4DCT Liver SI Motion (mm)
P1	43	10	1	Pancreas	3.19
P2	55.2	12	1	Pancreas	3.51
P3	38.4	8	1	Pancreas	1.53
P4	38.4	8	1	Pancreas	2.36
P5	40	10	1	Liver	4.91
P6	38.4	8	1	Pancreas	3.07

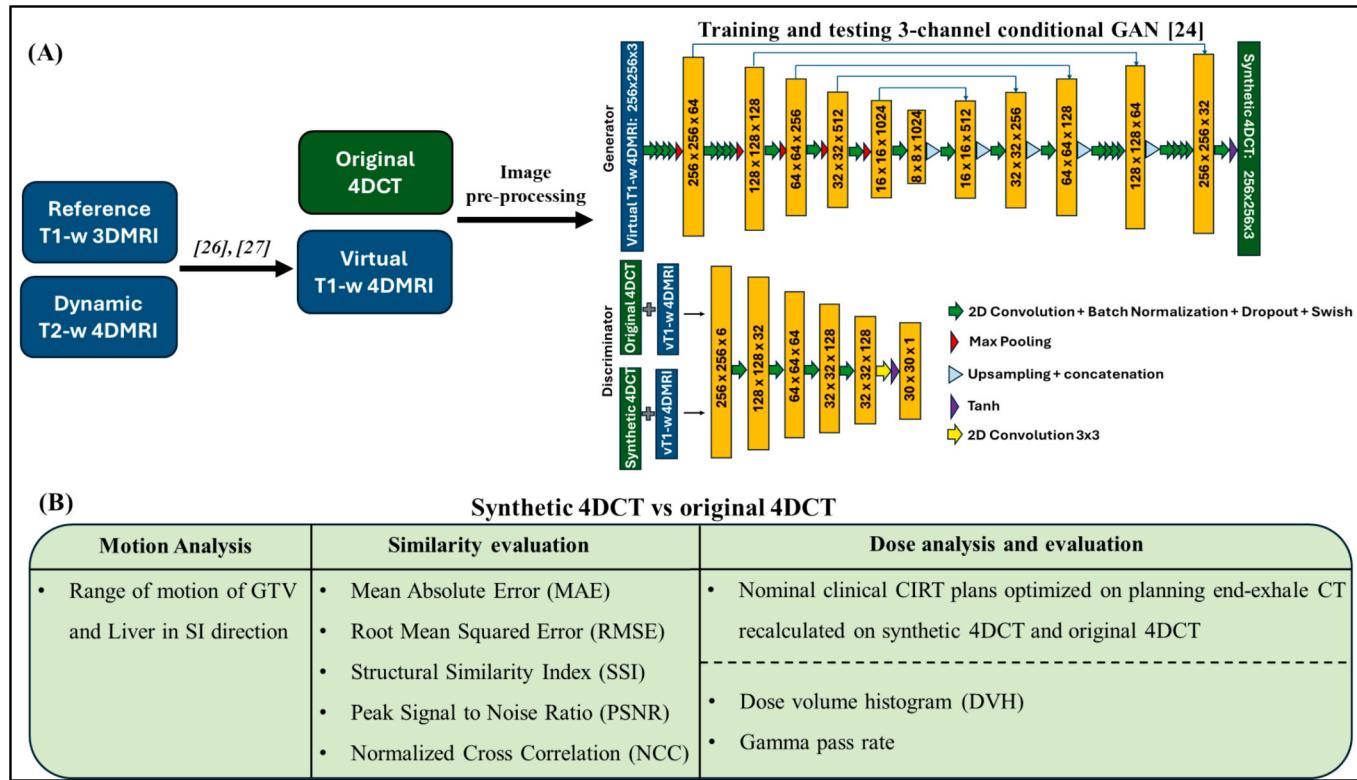


Fig. 1. (A) Workflow for data preparation, network training and 4D-sCT data generation; The training dataset consisted of paired 4DCT and virtual T1-weighted 4DMRI (v4DT1) data, with the latter generated through a Deformable Image Registration (DIR) pipeline exploiting dynamic T2-weighted 4DMRI data as in [26,27]; The network resembles the architecture of a Conditional Generative Adversarial Network (cGAN) consisting of: a U-net Generator aiming to create 4D-sCT data and a Discriminator seeking to differentiate between original 4DCT and generated volumes; (B) evaluation of generated 4D-sCT with respect to the original 4DCT.

of 40,960 slices, entailing all different respiratory phases.

2.3. Network and training

A 3-channel 2D cGAN [24] was trained and tested to generate 4D-sCT. The training dataset consisted of paired 4DCT and v4DT1 data, including four different respiratory phases of 20 abdominal patients, while an independent held-out-of-training test set of 6 patients was used.

The network consisted of two components trained adversarially, engaging in a min–max optimization: a 256x256x3 U-net Generator aiming to create 4D-sCT indistinguishable from the real 4DCT data, and a 70x70 PatchGAN Discriminator seeking to accurately differentiate between real and generated volumes (Fig. 1A). The blocks of the U-net Generator were used to process input data through multiple parallel convolutional operations with filters of varying size, i.e. 1×1 , 5×5 , 7×7 and 11×11 . The outputs of these parallel operations were concatenated along the channel dimension, forming the input for subsequent layers. The remaining layers of the U-net generator were composed of standard convolutional blocks with Swish activation functions and Batch Normalization. To mitigate overfitting, dropout layers with a 50 % rate were incorporated into the middle layers of the U-Net architecture. Additionally, max-norm regularization was employed to constrain the magnitude of layer weights, preventing the model from learning excessively large weights and improving generalization on unseen data. The max-norm parameter was fine-tuned for optimal performance, with a final selected value of 3. The Discriminator evaluated smaller, fixed size 70x70 patches of the image, using convolutional blocks, with batch normalization and ReLU activation functions.

The batch size was fixed at 1 throughout the optimization process to accommodate computational constraints, while the loss function

combined the cGAN adversarial cross-entropy loss with L1-distance, as described in [22]. The lambda parameter controlling the weight of the L1 loss, was tested over a range of 10 to 100. However, reducing the L1 loss weight did not yield improvements in model performance. The network was trained in an alternative manner; one gradient descent step on the discriminator and one on the generator. The ADAM optimizer was used while various learning rates for both the generator and discriminator were explored. The search range extended from 2×10^{-4} to 2×10^{-7} , and multiple combinations were evaluated during the tuning process. Final learning rates of 2×10^{-5} for the generator and 2×10^{-7} for the discriminator were selected, as they offered the best trade-off between convergence speed and model performance. Hyperparameters yielding the lowest Mean Absolute Error (MAE) and the highest Peak Signal-to-Noise Ratio (PSNR) within the patient body, were identified as optimal for network optimization.

All computational steps and network training were performed on an Ubuntu LTS Linux server workstation with a 25.7 GB RAM, NVIDIA Corporation GPU (CUDA version 11.8), with the model trained for 50 epochs.

2.4. Evaluation of generated sCT

Due to the lack of a patient ground truth motion information, the performance of the network was validated on a 4D XCAT phantom [30] and subsequently tested on the held-out-of-training patient set. For the phantom, a 4DCT of the same four respiratory phases as those in the patients' 4DCT, was available. XCAT 4DCT volumes were blurred with a Gaussian noise filter to simulate a realistic ground truth 4DCT, with the latter used to derive corresponding 4DMRI volumes with the 4D-ComBAT phantom [31]. The 4DCT-MRI volume pairs were pre-processed as in Section 2.2 and used to test the network.

The generated 4D-sCT, for both hold-out test set and phantom, were evaluated with respect to the original 4DCT by means of similarity metrics. MAE, Root Mean Square Error (RMSE), Structural Similarity Index (SSI), PSNR and Normalized Cross Correlation (NCC) were calculated between the 4D-sCT and 4DCT, with excluded backgrounds. Testing was performed on reassembled volumes utilizing 3-channel (i) 4DCT-based and (ii) 4DMRI-based masks. The latter approach adopted masks directly derived from the v4DT1 via manual segmentation, under consultation of a clinical expert, to simulate an MRI-only workflow (4DMRI-only workflow).

The range of motion of the liver and GTV in the SI direction was calculated for the phantom and all patients, through DIR between the 0EX-100IN phases of the 4DCT, 4D-sCT and v4DT1. A Friedman test ($\alpha = 5\%$) was performed to check the significance between the corresponding motions.

The 3-channel cGAN utilizes 2D image slices and as such this could compromise the smooth continuity between slices. To address this, the regularity of the liver profile was evaluated between 4DCT and 4D-sCT (4DMRI-only-based) by computing the RMSE in fitting the liver profile in all respiratory phases, as performed in [32,33]. The median, interquartile range and maximum of the RMSE were calculated for each

patient in the test set [32,33]. A Wilcoxon signed rank test ($\alpha = 5\%$) was used to check the statistical significance between the corresponding RMSE values.

Following the clinical approval procedure in CNAO, nominal clinical plans optimized on the 0EX-CT were recalculated within the gating window of the generated 4D-sCT and 4DCT, using the RayStation TPS. For each corresponding phase, dose differences were estimated by means of gamma analysis (3 %–3mm and 2 %–2mm) and dose-volume histogram (DVH) values relative to the Gross Tumor Volume (GTV) and Clinical Target Volume (CTV) D95% as well as relevant Organs at Risk (OARs) D2%, such as duodenum, stomach, colon, small bowel, kidneys and spinal cord. To determine the statistical significance of DVH differences between the 4D-sCT and 4DCT, a Two One Sided Test (TOST) for equivalence was performed, with a 95 % confidence and [-0.5, +0.5] % equivalence intervals.

3. Results

The training procedure of the network lasted 7–10 h, whereas a single sCT volume generation took 5 sec. Fig. 2 illustrates the original 4DCT and 4DMRI and the generated 4D-sCT (via 4DCT-based and

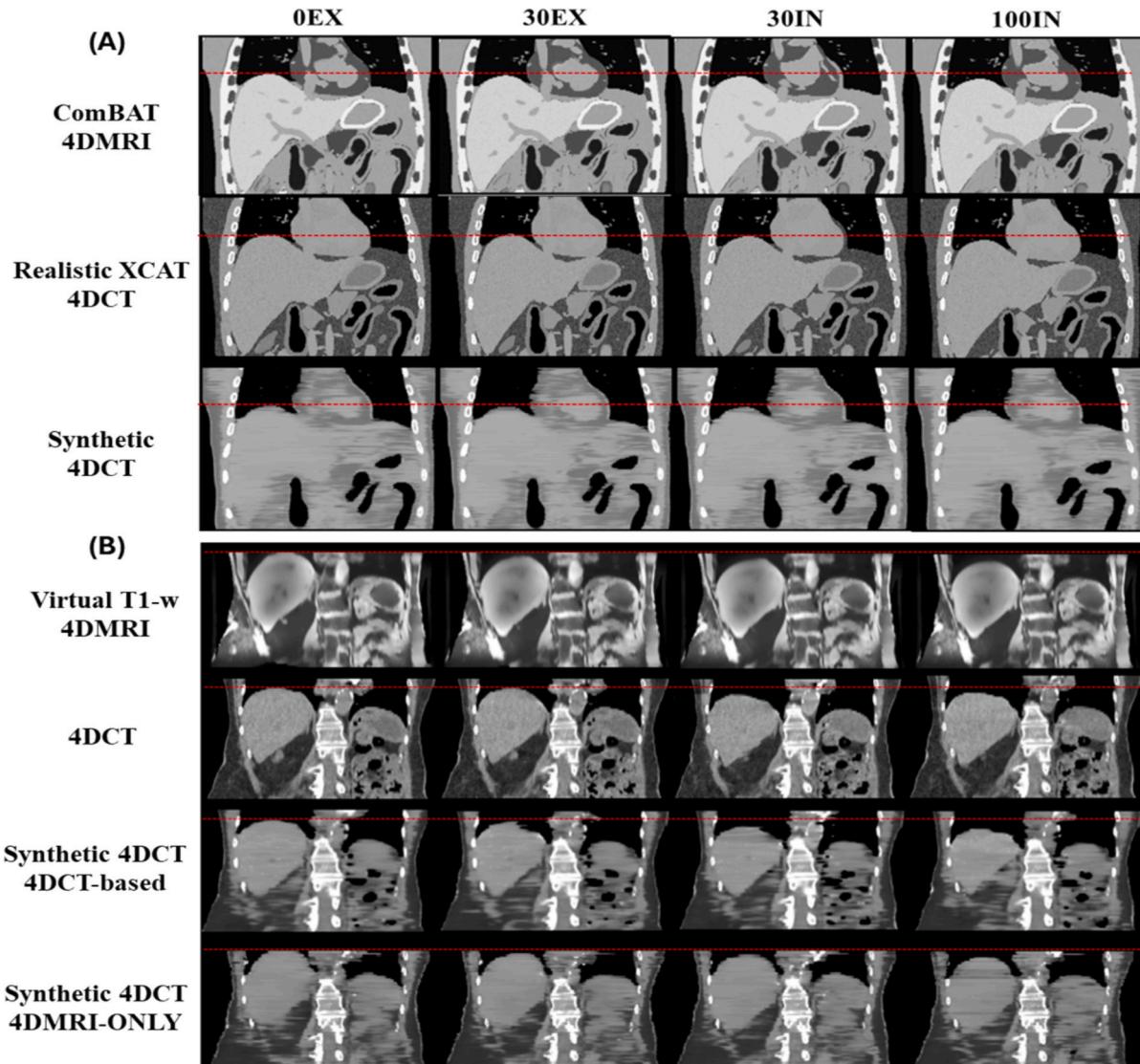


Fig. 2. Illustration of 4DMRI, 4DCT and generated 4D-sCT volumes in the coronal plane for the XCAT phantom (A) and an example abdominal patient for the 4DCT-based and 4DMRI-only scenario (B). The horizontal red dashed line reveals the diaphragm's excursion along the different respiratory phases. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4DMRI-only workflow), for both the phantom and an example patient.

3.1. Phantom

The quantitative results of the 4D-phantom similarity analysis are reported in [Supplementary Materials](#) (Section S1, [Table S1.1](#)) for each corresponding respiratory phase, considering only the values within the body contour. MAE was in the range of [72.93—73.10] HU while NCC reached values of 0.93 in all phases. The motion depicted in the 4DCT and 4DMRI was reflected in the 4D-sCT ([Table S2.1-S2.2](#)). Regarding the dosimetric analysis, [Fig. S1.1](#) shows the ComBAT 4DMRI, 4D-XCAT and 4D-sCT volumes overlayed with corresponding dose distributions, with the clinically optimized plan on the OEX-CT, recalculated on the 30EX/30IN-CT phases and on the 4D-sCT volume. [Fig. S1.2](#) depicts dose differences in DVH values between the realistic 4D-XCAT and the 4D-sCT, relative to the GTV and CTV D95% and D2% of gastrointestinal organs (duodenum, stomach and liver). For all organs, values were lower than 0.9 %.

3.2. Patients

The results of the similarity evaluation between the original 4DCT and 4D-sCT, for the 4DCT-based and 4DMRI-only scenario, are summarized in [Table 2](#). The average metrics' values are presented along with corresponding standard deviations, for each respiratory phase. In the 4DCT-based case, results demonstrated a good reproduction of the sCT volume with respect to the original 4DCT. For the 4DMRI-only evaluation values were substantially larger, especially in air where MAE grew up to 501.39 HU, due to the inter-acquisition variation of air fillings between the 4DCT and v4DT1/4D-sCT.

The SI motion of the liver and GTV among all patients and the phantom was overall in the range of [0.67–8.41] mm and [0.93–5.74] mm for the v4DT1 and 4DCT, respectively ([Table S2.1-S2.2](#)). For the 4DMRI-only scenario, 4D-sCT motion range was [1.55–9.86] mm, whereas for the 4DCT-based [1.07–5.92] mm. In both scenarios, no significant differences ($p > 0.05$) were found between the corresponding datasets. Additionally, the RMSE of the liver profile in both the 4DCT and 4DMRI-only 4D-sCT is reported in [Table S3.4](#). Statistical analysis resulted in $p > 0.05$, yielding no significant differences among the data.

[Fig. 3](#) illustrates v4DT1 volumes along with 4DCT and 4D-sCT volumes, overlayed with their corresponding dose distribution maps for an example patient. Regarding the 4DMRI-only scenario, [Fig. S3.1\(B\)](#) depicts the overlayed DVHs of the original 4DCT and recalculated 4D-sCT plans. Gamma analysis between corresponding respiratory phases, resulted in pass rates up to 98.26 % (3 %, 3 mm) for P1, while values of ~ 76 % were recorded for P6 ([Table S3.1](#)), due to inter-acquisition motion of the air-channel between the 4DCT and v4DT1 data, causing severe dose discrepancies ([Fig. S3.2](#)). When restricting the gamma criterion (2 %, 2 mm), pass rates degraded down to ~ 30 % for P6 ([Table S3.1](#)).

Table 2

Average (standard deviation) results of the similarity evaluation between the 4DCT and 4D-sCT volumes generated by the network for all six testing abdominal patients, in the 4DCT-based and 4DMRI-only scenarios.

Phase/Metric	MAE (HU)	MAE BONE (HU)	MAE SOFT (HU)	MAE AIR (HU)	RMSE (HU)	SSIM	PSNR (dB)	NCC
OEX	50.99 (3.73)	74.78 (15.27)	44.23 (3.35)	39.91 (8.84)	85.20 (9.62)	0.67 (0.05)	25.24 (3.58)	0.94 (0.03)
	51.29 (3.20)	75.59 (14.78)	44.88 (2.30)	35.97 (7.74)	85.37 (9.38)	0.67 (0.04)	25.24 (3.53)	0.95 (0.02)
100IN								
	50.75 (3.63)	74.72 (14.41)	44.01 (2.75)	38.52 (7.59)	84.31 (10.31)	0.67 (0.04)	25.35 (3.67)	0.94 (0.02)
	50.64 (3.49)	73.43 (13.45)	44.20 (2.54)	37.38 (7.96)	84.09 (9.82)	0.67 (0.04)	25.32 (3.68)	0.95 (0.02)
30IN	81.15 (11.08)	101.64 (12.69)	65.33 (11.30)	358.79 (120.04)	170.43 (24.26)	0.56 (0.08)	20.79 (0.59)	0.63 (0.16)
	90.22 (10.39)	94.38 (10.33)	57.72 (8.00)	501.39 (125.31)	192.20 (26.23)	0.55 (0.07)	20.18 (0.54)	0.57 (0.11)
OEX	81.35 (9.61)	94.03 (10.81)	57.73 (10.50)	442.17 (104.80)	172.94 (21.26)	0.57 (0.07)	20.72 (0.39)	0.61 (0.13)
	83.48 (8.96)	94.16 (10.03)	58.16 (9.82)	447.08 (102.47)	177.39 (20.87)	0.56 (0.07)	20.61 (0.36)	0.61 (0.12)
100IN	81.35 (9.61)	94.03 (10.81)	57.73 (10.50)	442.17 (104.80)	172.94 (21.26)	0.57 (0.07)	20.72 (0.39)	0.61 (0.13)
	83.48 (8.96)	94.16 (10.03)	58.16 (9.82)	447.08 (102.47)	177.39 (20.87)	0.56 (0.07)	20.61 (0.36)	0.61 (0.12)
30EX	81.35 (9.61)	94.03 (10.81)	57.73 (10.50)	442.17 (104.80)	172.94 (21.26)	0.57 (0.07)	20.72 (0.39)	0.61 (0.13)
	83.48 (8.96)	94.16 (10.03)	58.16 (9.82)	447.08 (102.47)	177.39 (20.87)	0.56 (0.07)	20.61 (0.36)	0.61 (0.12)
30IN	81.35 (9.61)	94.03 (10.81)	57.73 (10.50)	442.17 (104.80)	172.94 (21.26)	0.57 (0.07)	20.72 (0.39)	0.61 (0.13)
	83.48 (8.96)	94.16 (10.03)	58.16 (9.82)	447.08 (102.47)	177.39 (20.87)	0.56 (0.07)	20.61 (0.36)	0.61 (0.12)

Relative D95% differences on the GTV and CTV are depicted in [Fig. 4](#). Results showed small deviations with GTV values being < 1 %, except for P3, who reached a 3.26 % variation in the 30IN phase. Similarly, CTV values demonstrated comparable results between corresponding 4DCT/4D-sCT dose distributions, with a maximum deviation of 4.46 % for P3.

Regarding OARs, the interquartile range (IQR) of the D2 difference was [-0.34, 0.13] Gy(RBE) ([Fig. 5](#)). In terms of D2% relative error, results lied on an IQR of [-40, 10.31]%, with discrepancies observed in the gastrointestinal organs such as the duodenum and the small bowel, with the stomach reaching a peak of 60.19 %. However, such high values stemmed from the very small D2% doses deposited on the OARs ([Table S3.2](#)), thus emphasizing relative errors. Section S5 in [Supplementary materials](#) highlights regions of dose discrepancies for certain patients' cases, mainly due to inter-acquisition motion and especially air cavity mismatches between the 4DCT and 4D-sCT.

Alternatively, in the 4DCT-based scenario, results were more favorable with respect to the 4DMRI-only scenario with an average pass rate of ~ 97 % among all patients and phases ([Table S3.1](#)), while relative D95% values on the GTV/CTV remained lower than 0.7 % and 1.1 %, respectively ([Fig. S3.3](#)). For OARs, the IQR of the relative D2 difference dropped to [-2.02, 1.49] % ([Fig. S3.4](#)). However, errors above clinical tolerances were still observed in the gastrointestinal organs (reported in section S6), again mainly due to air pockets dispositions, with a maximum value of 33.53 % in the stomach reported due to inner organ inter-acquisition motion ([Fig. S6.2\(C\)](#)).

4. Discussion

In this work, we utilized our 3-channel network to generate for the first time abdominal 4D-sCTs for applications in CIRT [[23](#)] using paired v4DT1 and 4DCT volumes. Literature studies have mainly focused on the synthesis of 3D-sCT thus, to facilitate a meaningful comparison, our analysis was organized and presented for each separate respiratory phase. We first validated the 4D network using a computational 4D-CT/MRI phantom and then evaluated its performance on six abdominal test patients, by means of motion, similarity and dosimetric analysis.

The 4D-CT/MRI phantom guaranteed anatomical consistency between the 4DCT and 4DMRI/4D-sCT with respect to patients for validating the network. The similarity metrics results were in agreement with the literature, demonstrating our network's potential in reproducing anatomical structures. The absence of inter-acquisition variation between 4DCT-4DMRI in the phantom resulted in low differences between CT-sCT dose distributions, with DVH metrics on GTV, CTV and OARs being less than 1 %. Future analysis will also consider the development of more realistic computational phantoms [[34](#)] which can further support the validation process.

Regarding the application on patients, the similarity analysis revealed better results for the 4DCT-based scenario. This could be attributed to the application of 4DCT-derived masks on the reassembled

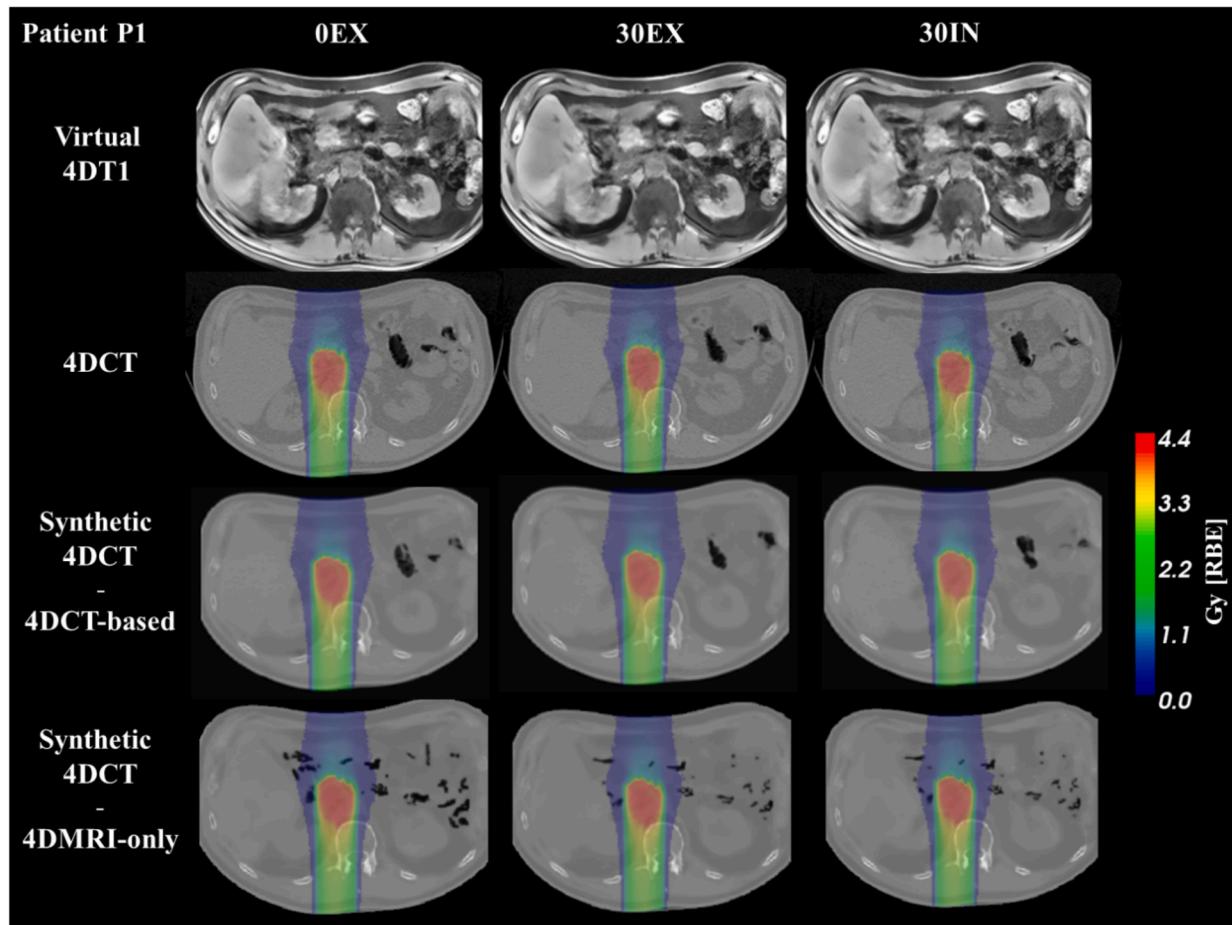


Fig. 3. Axial plane of the virtual 4DT1, 4DCT and 4D-sCT volumes overlayed with corresponding dose distributions. The clinically optimized plan on the 0EX-CT was recalculated on the 30EX/30IN-CT phases and on the 4D-sCT volumes. Patient P1 of the test set is illustrated as an example.

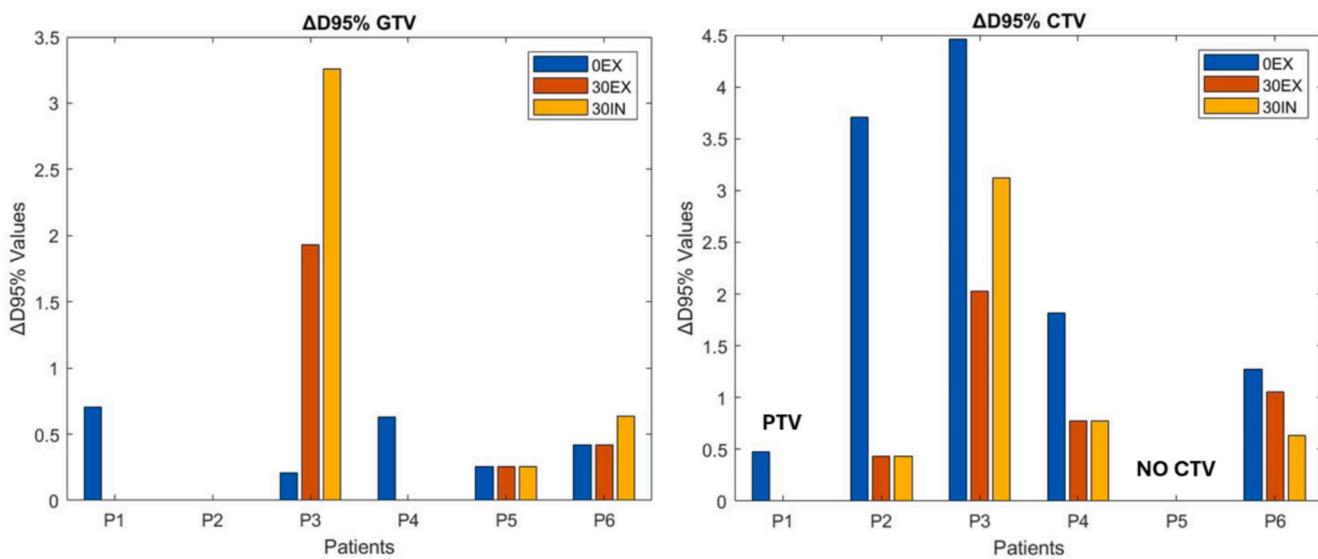


Fig. 4. Relative D95% difference on GTV (left) and CTV (right) between the 4DCT dose plans and the recalculated 4D-sCT ones, per respiratory phase for all patients in the 4DMRI-only scenario. For P1, PTV values are shown on the right graph, whereas for P5 only GTV was considered for comparison; Patients with no visible barplot had zero relative errors.

4D-sCT that might have better replicated the 4DCT anatomy. In contrast in the 4DMRI-only case, the v4DT1-derived masks aided the replication of the MRI anatomy, thus resulting in higher similarity metric values,

especially in the air channel, which was highly inter-fractionally variable within the abdominal area. Nonetheless, our network can favorably compare with other studies in literature, utilizing U-nets, cycleGANs or

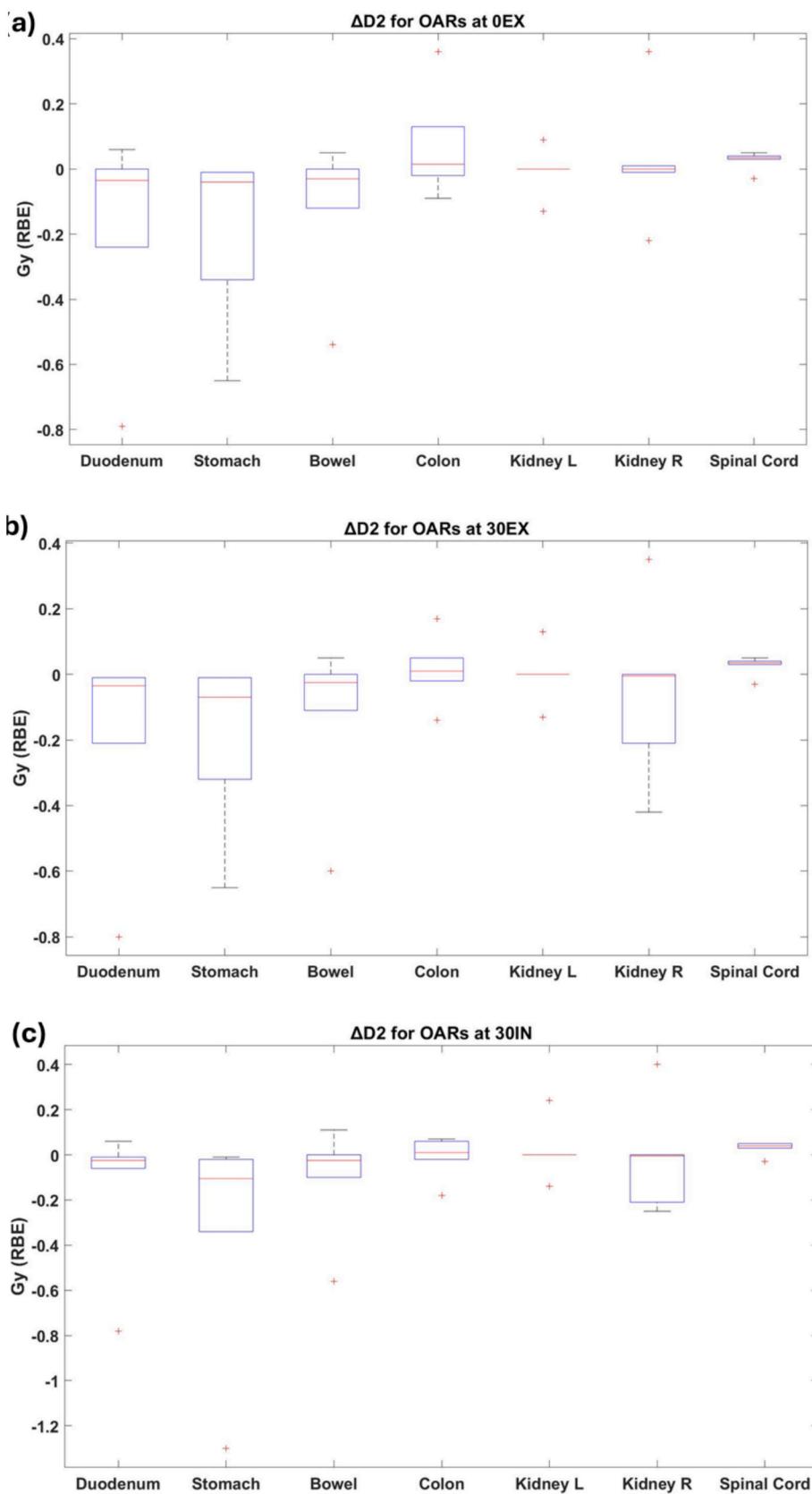


Fig. 5. D2 differences between 4DCT and 4D-sCT for OARs in all patients for the 4DMRI-only scenario. The results are organized in boxplots for each separate phase, namely the end-exhale (a), 30% exhale (b) and 30% inhale (c).

cGANs. Parrella et al. [22], made a thorough comparison with multiple studies, demonstrating consistency and agreement of his results. Both in the 4DCT-based and 4DMRI-only scenario, our MAE and RMSE values outperformed Parrella's network, with an improvement seen in PSNR and NCC, while SSIM values were comparable. The reason for this improvement could be the expansion of our training dataset. Indeed, even though a similar number of patients was used in both studies, additional volumes corresponding to different respiratory phases were introduced in the network's training, increasing the training dataset by 87 %. Our network also favorable compares to recent studies on abdominal cases, with [35] reporting a MAE of 56 ± 7 (HU) using the HighRes3DNet and [36] having $MAE = 70.10 \pm 18.97$ (HU) using a cycle-consistent cycleGAN. Garcia Hernandez et al., [21] presented outstanding results utilizing a U-net and a GAN, with MAE values of 28 ± 14.7 (HU) and 25.9 ± 13.4 (HU), respectively, exploiting however a larger patient dataset, thus having more heterogenous data in their network.

The motion analysis showed comparable GTV and liver SI motion among 4DCT, 4D-sCT and v4DT1 datasets in both 4D-sCT generation scenarios. It is notable that for the 4DMRI-only case, the motion range is more similar to the v4DT1 while for the 4DCT-based, motion resembles the 4DCT, as a result of the corresponding masks used to derive 4D-sCT volumes.

Concerning dose analysis, our 4DMRI-only results were severely affected by the inter-acquisition variation between 4DCT and 4D-sCT and predominantly by air pocket dispositions. Gamma pass rates were in the range of ~ 94–98 for all respiratory phases with a (3 %, 3mm) criterion, except for P2 and P6 that resulted in ~ 86 and ~ 77. When using a more conservative gamma criterion of (2 %, 2mm), gamma pass rates degraded (Table S3.1), although it should be noticed that original plans are optimized through a robust optimization of 3 mm set-up, 3 % range uncertainty, as performed in the clinic. In general, gamma pass rates among literature are found to have values above 99 %, however studies so far have focused only on photon and proton therapy applications, making their findings not directly comparable to our application in CIRT, which is also more sensitive to variations along the beam path with respect to the other two treatment modalities.

DVH analysis on the GTV/CTV presented relative differences within clinical tolerances, namely below 5 % of the prescribed dose for all patients, with our results being aligned with other studies [22]. For OARs, errors were reported in the gastrointestinal organs, which were affected by the inter-acquisition motion and air cavities between the 4DCT and 4D-sCT (S5 of [Supplementary Materials](#)). In general, our findings dictate that compliance for OARs constraints is difficult, with other studies also highlighting this issue. Parrella et al. [22] presented an error of 15.18 Gy(RBE) in the colon, while Knausl et al. [23] reported errors up to 28 % in case of 3D applications. For patient P2 errors caused by inter-acquisition motion of the inner organs (Fig. S5.2), suggest that in-room imaging is required [28,37]. It should be also noted that the analysis was conducted with just one single beam, as this was the clinical procedure in the absence of a gantry in CNAO [28]. To further evaluate the performance of our method we simulated a 2-beams plan on the worst-case scenario (P3 – selected in terms of highest ΔD_2 variation in the stomach and colon) (Fig. S3.5). Dose variations between 4DCT and 4D-sCT remained less than 0.7 Gy(RBE) (Table S3.2 and Fig. S3.6), while gamma analysis with a (3 %, 3 mm) criterion, resulted in ~ 95 % for each respiratory phase (Table S3.3).

The 4DCT-based analysis showed improved dosimetric results, with higher gamma rates and DVH metrics on GTV and CTV within clinical tolerances. Regarding OARs, 4DCT-based results were consistent with Florkow et al., [18] whose results were in the range of [-2.7, 3.7]%. However, higher errors (more than 4 %) were still observed for certain patients mainly in the gastrointestinal organs, like the duodenum, stomach, bowel and colon (Fig. S6.1–S6.2).

Nevertheless, in both the 4DCT-based and 4DMRI-only scenario, our statistical analysis using the TOST equivalence test, resulted in P-values

lower than 0.05 for all the relevant organs ([Table S4](#)), demonstrating the equivalence in DVH values between 4D-sCT and 4DCT.

5. Conclusion

This work demonstrates that our 3-channel network can generate accurate 4D-sCT of the abdominal site, allowing CIRT dose calculations for abdominal cases. In this context, our approach can be put forward to support treatment planning evaluation and robustness in CIRT. Future studies will focus on expanding and testing our network on additional patients' data, incorporating diverse patient cases and benchmarking against other DL-based networks. Currently the use 3-channels has shown to better account for the anatomical complexity of the abdomen; however, we plan to simplify the net to a 1-channel implementation, making the procedure streamlined and avoiding the segmentation on MRI which can be inaccurate.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by Marie Skłodowska-Curie grant RAPTOR—Real-Time Adaptive Particle Therapy of Cancer, grant agreement No. 955956 and partially supported by AIRC (Associazione Italiana per la Ricerca sul Cancro), grant number IG2020-24946. The authors would like to thank Luca Anemoni and Margherita Sofia Cadeo for data collection.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of CENTRO NAZIONALE DI ADROTERAPIA ONCOLOGICA (CNAO) (protocol code CNAO 37-2019 4D-MRI, 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejmp.2025.104963>.

References

- [1] Durante M, Orechia R, Loeffler JS. Charged-particle therapy in cancer: clinical uses and future perspectives. *Nat Rev Clin Oncol* 2017;14(8):483–95. <https://doi.org/10.1038/nrclinonc.2017.30>.
- [2] Liermann J, Shinoto M, Syed M, Debus J, Herfarth K, Naumann P. Carbon ion radiotherapy in pancreatic cancer: a review of clinical data. *Radiother Oncol* Jun. 2020;147:145–50. <https://doi.org/10.1016/j.radonc.2020.05.012>.
- [3] Bert C, Durante M. Motion in radiotherapy: particle therapy. *Phys Med Biol* Jul. 2011;56(16):R113. <https://doi.org/10.1088/0031-9155/56/16/R01>.
- [4] Mori S, Knopf AC, Umegaki K. Motion management in particle therapy. *Med Phys* 2018;45(11):e994–1010. <https://doi.org/10.1002/mp.12679>.
- [5] C. Paganelli et al., 'MRI-guidance for motion management in external beam radiotherapy: Current status and future challenges', Nov. 20, 2018, *Institute of Physics Publishing*, doi: 10.1088/1361-6560/aaebcf.
- [6] Gupta D, Kim M, Vineberg KA, Balter JM. Generation of synthetic CT images from MRI for treatment planning and patient positioning using a 3-channel U-net trained on sagittal images. *Front Oncol* 2019;9(SEP):482352. <https://doi.org/10.3389/FONC.2019.00964>.
- [7] Hsu SH, Han Z, Leeman JE, Hu YH, Mak RH, Sudhyadhom A. Synthetic CT generation for MRI-guided adaptive radiotherapy in prostate cancer. *Front Oncol* Sep. 2022;12. <https://doi.org/10.3389/FONC.2022.969463>.
- [8] Wang CC, et al. Magnetic resonance-based synthetic computed tomography using generative adversarial networks for intracranial tumor radiotherapy treatment planning. *J Pers Med* 2022;12(3). <https://doi.org/10.3390/JPM12030361/S1>.
- [9] Bird D, Speight R, Andersson S, Wingqvist J, Al-Qaisieh B. Deep learning MRI-only synthetic-CT generation for pelvis, brain and head and neck cancers. *Radiother Oncol* Feb. 2024;191:110052. <https://doi.org/10.1016/j.radonc.2023.110052>.

- [10] Moteghaed NY, Fatemi A, Mostaar A. Pseudo-computed tomography generation from noisy magnetic resonance imaging with deep learning algorithm. *Front Biomed Technol* Dec. 2024;11(1):113–21. <https://doi.org/10.18502/fbt.v1i1.14518>.
- [11] Dayarathna S, Islam KT, Uribe S, Yang G, Hayat M, Chen Z. Deep learning based synthesis of MRI, CT and PET: Review and analysis. *Med Image Anal* Feb. 2024;92:103046. <https://doi.org/10.1016/j.media.2023.103046>.
- [12] Galapon AV, Thummerer A, Langendijk JA, Wagenaar D, Both S. Feasibility of Monte Carlo dropout-based uncertainty maps to evaluate deep learning-based synthetic CTs for adaptive proton therapy. *Med Phys* 2023;1–11. <https://doi.org/10.1002/MP.16838>.
- [13] Boulanger M, et al. Deep learning methods to generate synthetic CT from MRI in radiotherapy: a literature review. *Phys Med Sep.* 2021;89:265–81. <https://doi.org/10.1016/j.ejmp.2021.07.027>.
- [14] Villegas F, et al. Challenges and opportunities in the development and clinical implementation of artificial intelligence based synthetic computed tomography for magnetic resonance only radiotherapy. *Radiother Oncol Sep.* 2024;198. <https://doi.org/10.1016/J.RADONC.2024.110387>.
- [15] Cusumano D, et al. A deep learning approach to generate synthetic CT in low field MR-guided adaptive radiotherapy for abdominal and pelvic cases. *Radiother Oncol Dec.* 2020;153:205–12. <https://doi.org/10.1016/J.RADONC.2020.10.018>.
- [16] Fu J, Singhrao K, Cao M, Yu V, Santhanam AP, Yang Y, et al. Generation of abdominal synthetic CTs from 0.35T MR images using generative adversarial networks for MR-only liver radiotherapy. *Biomed Phys Eng Express* 2020;6(1):015033. <https://doi.org/10.1088/2057-1976/AB6E1F>.
- [17] Xu K, et al. Multichannel residual conditional GAN-leveraged abdominal pseudo-CT generation via dixon MR images. *IEEE Access* 2019;7:163823–30. <https://doi.org/10.1109/ACCESS.2019.2951924>.
- [18] Florkow MC, et al. Deep learning-enabled MRI-only photon and proton therapy treatment planning for paediatric abdominal tumours. *Radiother Oncol Dec.* 2020;153:220–7. <https://doi.org/10.1016/J.RADONC.2020.09.056>.
- [19] Olberg S, et al. Abdominal synthetic CT reconstruction with intensity projection prior for MRI-only adaptive radiotherapy. *Phys Med Biol* 2021;66(20):Oct. <https://doi.org/10.1088/1361-6560/ac279e>.
- [20] Nousiainen K, Santurio GV, Lundahl N, Cronholm R, Siversson C, Edmund JM. Evaluation of MRI-only based online adaptive radiotherapy of abdominal region on MR-linac. *J Appl Clin Med Phys Mar.* 2023;24(3):e13838. <https://doi.org/10.1002/ACM2.13838>.
- [21] Garcia Hernandez A, Fau P, Wojak J, Mailleux H, Benkreira M, Rapacchi S, et al. Synthetic computed tomography generation for abdominal adaptive radiotherapy using low-field magnetic resonance imaging. *Phys Imaging Radiat Oncol* 2023;25:100425. <https://doi.org/10.1016/j.phro.2023.100425>.
- [22] Parrella G, Vai A, Nakas A, Garau N, Meschini G, Camagni F, et al. Synthetic CT in carbon ion radiotherapy of the abdominal site. *Bioengineering* 2023;10(2):250. <https://doi.org/10.3390/bioengineering10020250>.
- [23] Knäusl B, et al. Possibilities and challenges when using synthetic computed tomography in an adaptive carbon-ion treatment workflow. *Z Med Phys* May 2023;33(2):146. <https://doi.org/10.1016/J.ZEMEDI.2022.05.003>.
- [24] Rossi M, Cerveri P. Comparison of supervised and unsupervised approaches for the generation of synthetic CT from cone-beam CT. *Diagnostics (Basel)* 2021;11(8):Aug. <https://doi.org/10.3390/DIAGNOSTICS11081435>.
- [25] Meschini G, et al. A clustering approach to 4D MRI retrospective sorting for the investigation of different surrogates. *Phys Med Feb.* 2019;58:107–13. <https://doi.org/10.1016/j.ejmp.2019.02.003>.
- [26] Meschini G, et al. Virtual 4DCT from 4DMRI for the management of respiratory motion in carbon ion therapy of abdominal tumors. *Med Phys Mar.* 2020;47(3):909–16. <https://doi.org/10.1002/mp.13992>.
- [27] Annunziata S, Rabe M, Vai A, Molinelli S, Nakas A, Meschini G, et al. Virtual 4DCT generated from 4DMRI in gated particle therapy: phantom validation and application to lung cancer patients. *Phys Med Biol* 2023;68(14):145004. <https://doi.org/10.1088/1361-6560/ACDEC5>.
- [28] Molinelli S, et al. The role of multiple anatomical scenarios in plan optimization for carbon ion radiotherapy of pancreatic cancer: Inter-fraction robustness in CIRT for pancreatic cancer. *Radiother Oncol Nov.* 2022;176:1–8. <https://doi.org/10.1016/j.radonc.2022.09.005>.
- [29] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, ‘Image-to-Image Translation with Conditional Adversarial Networks’, Nov. 2016, [Online]. Available: <http://arxiv.org/abs/1611.07004>.
- [30] Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BMW. 4D XCAT phantom for multimodality imaging research. *Med Phys Sep.* 2010;37(9):4902–15. <https://doi.org/10.1118/1.3480985>.
- [31] Paganelli C, Summers P, Gianoli C, Bellomi M, Baroni G, Riboldi M. A tool for validating MRI-guided strategies: a digital breathing CT/MRI phantom of the abdominal site. *Med Biol Eng Comput Nov.* 2017;55(11):2001–14. <https://doi.org/10.1007/S11517-017-1646-6/FIGURES/8>.
- [32] Paganelli C, Kipritidis J, Lee D, Baroni G, Keall P, Riboldi M. Image-based retrospective 4D MRI in external beam radiotherapy: a comparative study with a digital phantom. *Med Phys Jul.* 2018;45(7):3161–72. <https://doi.org/10.1002/mp.12965>.
- [33] Paganelli C, Summers P, Bellomi M, Baroni G, Riboldi M. Liver 4DMRI: A retrospective image-based sorting method. *Med Phys Aug.* 2015;42(8):4814–21. <https://doi.org/10.1118/1.4927252>.
- [34] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, ‘Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks’, *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2017–October, pp. 2242–2251, Mar. 2017, doi: 10.1109/ICCV.2017.244.
- [35] Nousiainen K, Santurio GV, Lundahl N, Cronholm R, Siversson C, Edmund JM. Evaluation of MRI-only based online adaptive radiotherapy of the abdominal region on MR-linac. *J Appl Clin Med Phys* 2023;24(3):e13838. <https://doi.org/10.1002/acm2.13838>.
- [36] Lapaeva M, et al. Synthetic computed tomography for low-field magnetic resonance-guided radiotherapy in the abdomen. *Phys Imaging Radiat Oncol Oct.* 2022;24:173–9. <https://doi.org/10.1016/J.PHRO.2022.11.011>.
- [37] Belotti G, Rossi M, Pella A, Cerveri P, Baroni G. A new system for in-room image guidance in particle therapy at CNAO. *Phys Med Oct.* 2023;114. <https://doi.org/10.1016/J.EJMP.2023.103162>.