



OPEN ACCESS

EDITED BY

Francesca Angelone,
Università del Sannio, Italy

REVIEWED BY

Federica Kiyomi Ciliberti,
Reykjavik University, Iceland
Sara Spaziani,
University of Sannio, Italy

*CORRESPONDENCE

Ciro Benito Raggio
✉ ciro.raggio@kit.edu

RECEIVED 16 February 2026

REVISED 22 May 2026

ACCEPTED 29 May 2026

PUBLISHED 15 June 2026

CITATION

Raggio CB, Zaffino P and Spadea MF
(2026) A privacy-preserving federated
learning framework for generalizable
CBCT to synthetic CT translation in head
and neck.
Front. Digit. Health 8:1812254.
doi: 10.3389/fdgth.2026.1812254

COPYRIGHT

© 2026 Raggio, Zaffino and Spadea. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

A privacy-preserving federated learning framework for generalizable CBCT to synthetic CT translation in head and neck

Ciro Benito Raggio^{1*}, Paolo Zaffino² and
Maria Francesca Spadea¹

¹Institute of Biomedical Engineering, Karlsruhe Institute of Technology, Karlsruhe, Germany,

²Department of Experimental and Clinical Medicine, Magna Graecia University, Catanzaro, Italy

Background: Cone-beam computed tomography (CBCT) has become a widely adopted modality for image-guided radiotherapy (IGRT). However, CBCT is characterized by increased noise, limited soft-tissue contrast, and artifacts. These issues result in unreliable Hounsfield unit (HU) values, which limits electron density estimation for direct dose calculation. These issues have been addressed by deriving synthetic CT (sCT) from CBCT, particularly by adopting deep learning (DL) methods. However, existing DL approaches are hindered by institutional heterogeneity, scanner-dependent variations, and data privacy regulations that prevented multi-center data sharing.

Methods: To overcome these challenges, we propose a cross-silo federated learning (FL) approach for CBCT-to-sCT synthesis in the head and neck region. This approach extends the original FedSynthCT framework to a different image modality and anatomical region. A conditional generative adversarial network (cGAN) was trained using data from three European medical centers within the SynthRAD2025 public challenge dataset while maintaining data privacy at each institution. A combination of the FedAvg and FedProx aggregation strategies, alongside a standardized preprocessing pipeline, was adopted to federate the DL model.

Results: The federated model effectively generalized across participating centers, as evidenced by the mean absolute error (MAE) ranging from 64.38 ± 13.63 to 85.90 ± 7.10 HU, the structural similarity index (SSIM) ranging from 0.88 ± 0.02 to 0.92 ± 0.04 , and the peak signal-to-noise ratio (PSNR) ranging from 32.86 ± 0.94 to 34.91 ± 1.04 dB. Notably, performance on an external validation dataset of 60 patients yielded comparable metrics: a MAE of 75.22 ± 11.81 HU, an SSIM of 0.90 ± 0.03 and a PSNR of 33.52 ± 2.06 , confirming robust cross-center generalization despite differences in imaging protocols and scanner types, without additional training. Furthermore, a visual analysis of the results revealed that the obtained metrics were influenced by registration errors.

Conclusions: Our findings demonstrated the technical feasibility of FL for CBCT-to-sCT synthesis task while preserving data privacy, offering a collaborative solution for developing generalizable models across institutions without requiring data sharing or center-specific models.

KEYWORDS

CBCT, data privacy, data sharing, deep learning, federated learning, head and neck cancer, image-to-image translation, synthetic computed tomography

1 Introduction

Radiotherapy (RT) has seen significant advancements in recent decades, particularly with the integration of image-guided radiotherapy (IGRT). IGRT involves the use of frequent imaging during treatment to account for anatomical changes and improve the accuracy of radiation delivery. Cone-beam computed tomography (CBCT) scanners are often incorporated into the gantry of linear accelerators, facilitating their integration into clinical practice. Consequently, CBCT emerged as one of the most widely adopted IGRT modalities (1–3). The advent of CBCT addressed the limitations of planar imaging, previously dominant in early image guidance, by enabling volumetric imaging directly within the treatment suite, significantly improving treatment precision and clinical workflows (2, 4, 5). This technological evolution enabled improved alignment, visualization of patient anatomy, and real-time verification of target localization, allowing clinicians to manage both inter- and intra-fractional anatomical variability. As a result, CBCT became a pivotal technology in the development of online image-guided adaptive radiotherapy (IGART) approaches (4–7). Despite these advantages, CBCT is hindered by several drawbacks, including increased image noise, reduced soft-tissue contrast, and reconstruction artifacts, which lead to unreliable Hounsfield Unit (HU) values and limit its applicability in electron density estimation for direct dose calculation (1, 5, 6). To overcome these limitations, the concept of synthetic CT (sCT) generation was introduced, where CBCT images are translated into equivalent synthetic CT images with improved dosimetric fidelity. This approach avoids unnecessary radiation exposure and reduces clinical workload by eliminating the need for additional scans (3, 5, 8).

The CBCT-to-sCT conversion task has evolved considerably, with applications across diverse anatomical regions, including head and neck, thorax, pelvis, prostate, abdomen, and pancreas, highlighting both its broad clinical relevance and intrinsic methodological challenges (3, 9). Among the strategies developed for sCT generation, deep learning (DL) approaches have proven especially effective, demonstrating strong capabilities in learning complex, non-linear mappings between image domains and producing high-quality synthetic images suitable for adaptive RT applications (1, 3, 5, 8).

Convolutional Neural Networks (CNNs) have been widely adopted in medical image analysis due to their ability to extract both local and global features, making them well suited for enhancing CBCT image quality in CBCT-to-sCT synthesis tasks. In this context, the U-Net architecture (10), has demonstrated notable efficacy in image-to-image translation tasks by enabling pixel-level predictions while preserving spatial and contextual information (3, 9). Generative Adversarial Networks (GANs) have also been extensively employed, particularly conditional GANs such as Pix2Pix (11), which utilize a generator-discriminator framework and a PatchGAN discriminator to promote local structural consistency and realistic texture synthesis. Cycle-GAN architectures further introduced cycle consistency constraints to preserve anatomical structures during CBCT-CT translation, although training instability has been reported (3, 9). More recent approaches have explored Transformer-based architectures and Denoising Diffusion

Probabilistic Models (DDPMs) for sCT generation. Transformer-based models, such as TransCBCT (12), employ self-attention mechanisms to capture long-range spatial dependencies and complex anatomical relationships, but require large training datasets, substantial computational resources, and are sensitive to image resolution (3, 12). DDPMs, which learn to reverse a progressive noising process, have demonstrated high-fidelity sCT generation when conditioned on CBCT inputs, but are similarly computationally intensive and characterized by slow inference and high hardware requirements (3). Despite their architectural differences, all of the aforementioned models have demonstrated promising performance in producing clinically viable sCTs for radiotherapy planning (3, 9).

The evaluation of sCT generation DL models in predicting accurate images comparable to gold standard planning CTs, also named ground-truth CT, was assessed by computing metrics such as mean absolute error (MAE), mean squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM), computed within patient-specific masks to exclude background regions. Among these metrics, the MAE was often used as loss function, and was referred to as masked MAE or L1 Loss (3, 9). Gradient descent-based optimization has been extensively adopted for the iterative update of model weights. Concurrently, pre-processing methodologies such as image registration, image normalization and data augmentation have been demonstrated to be critical for model performance and stability (3, 8, 9, 13).

Despite the increasing adoption of DL for the sCT generation, its application remains limited by institutional heterogeneity, scanner-dependent variations, and heterogeneous imaging protocols, often necessitating single-site model training and constraining clinical scalability (3, 7, 14). Although multi-center data aggregation could partially address these limitations, it is severely constrained by strict data privacy regulations, such as the General Data Protection Regulation (GDPR)¹ in Europe, and the Health Insurance Portability and Accountability Act (HIPAA)² in the United States, which restrict the sharing and centralization of sensitive medical data (14–16). These limitations emphasize the necessity for collaborative solutions that facilitate access to diverse and extensive datasets without necessitating direct data sharing or centralized storage.

In this context, federated learning (FL) has emerged as a decentralized training paradigm that enables institutions to collaboratively train DL models while keeping data local, thereby maintaining compliance with privacy regulations and enhancing cross-center collaboration (17–19). FL has gained increasing attention in medical imaging, with applications in classification and segmentation tasks (16, 17, 20–23). Beyond the domains of classification and segmentation, federated approaches recently emerged in the context of medical image synthesis. For instance, FedMed-GAN has been developed to synthesize cross-modality brain magnetic resonance images while leveraging simple federated aggregation techniques such as the federated averaging (FedAvg), which averages the locally updated model weights

¹<https://gdpr-info.eu/>.

²<https://www.hhs.gov/hipaa/index.html>.

from multiple clients to obtain a global model without sharing raw data (24, 25).

FedSynthCT-Brain was the first study to demonstrate the successful application of FL to the sCT synthesis task, specifically from brain T1-weighted magnetic resonance imaging (MRI) data, without compromising performance on external datasets (14). This finding suggested potential applications for extending the federated concept to the context of CBCT-to-sCT translation, where the development of robust, center-agnostic and generalizable models is essential for clinical adoption (8).

More in details, while MRI-to-CT translation involves inferring HU values from images which cannot inherently encode them, CBCT measures x-ray attenuation but suffers from severe quality degradation. Thus, the main challenge in the CBCT-to-CT translation task shifts from learning a cross-modality mapping to restoring accurate HU values from noisy, scattered and vendor-dependent reconstructed images. Furthermore, the transition of the model working region, from the brain (introduced by FedSynthCT-Brain) to the head and neck, introduces greater anatomical variability and high-contrast structures, such as dental implants and extensive air cavities. Consequently, this transition results in a task that is substantially different and more sensitive to misalignment and residual registration errors. These factors make the successful deployment of FL in this setting non-straightforward and highlight the need for a dedicated framework to determine whether federated training can remain robust and viable under such adverse and heterogeneous conditions.

Therefore, in this study, we proposed a cross-silo horizontal FL approach for CBCT-to-sCT in the head and neck region, extending the FedSynthCT (14) framework to a different imaging modality and a larger anatomical region. A federated DL model was collaboratively trained using data from different

European centers of the public SynthRAD2025 challenge dataset (26). The approach was validated not only on each federated center but also on an independent dataset outside the federation. This validation process was used to assess model generalization capabilities, thereby demonstrating that the proposed approach can result in more robust and generalizable models while preserving data privacy.

2 Materials and methods

2.1 Datasets

The datasets employed in this work were from the publicly available SynthRAD2025 challenge repository (26, 27). The institutions (also referred to as centers or clients) focus on head and neck (HN) region and were labeled as centers A, B, C, and E. Center D was deliberately excluded from this study due to the limited availability. For consistency, we adopted the same naming convention for the centers as used in the challenge.

As reported, a few thorax scans were included in the HN cohort (26); these outlier cases were excluded from the study. The final number of subjects included per center is reported in Table 1. For each subject, the dataset included a CBCT scan, a CT scan, and a binary segmentation mask delineating the patient body.

Variability in imaging protocols and scanner types among relevant centers has been identified. Notably, CBCT scans from Center E were acquired using a different system, in contrast to the devices employed in centers A, B and C. Center E also exhibited markedly longer exposure times and higher in-plane resolution variability. Regarding the CT images, it was observed that all centers employed standard 120 kVp acquisitions; however, exposure time, slice thickness and pixel spacing varied.

TABLE 1 Overview of the key CBCT and CT original acquisition parameters used at Centers A, B, C, and E according to the dataset documentation (26).

Parameter	Center A	Center B	Center C	Center E
CBCT				
Patient #	60	65	63	63
Scanner	Elekta XVI v5.x	Elekta XVI v5.52	Elekta XVI v5.x	Varian TrueBeam OBI
kVp	100–120	100	120	100–125
Exposure Time [ms]	10–32	10	22	7,500–18,060
Rows/Columns	270	270	270–512 × 270–512	512 × 512
Pixel spacing [mm]	1 × 1	1 × 1	1 × 1	0.5–0.9 × 0.5–0.9
Slice thickness [mm]	1	1	1	2
CT				
Patient #	60	65	63	63
Scanner	Philips Big Bore (90), Siemens Biograph40 (10)	Toshiba Aquilion/LB	Philips Brilliance Big Bore (93), Siemens Biograph40 (7)	Toshiba Aquilion/LB, Siemens Biograph128
kV	120	120	120	120
Exposure Time [ms]	615–1,000	500–1,000	922–1457	500–1,000
Rows/Columns	512	512	512	512
Pixel spacing [mm]	0.7–1.4 × 0.7–1.4	1.1–1.4 × 1.1–1.4	1–1.2 × 1–1.2	1–1.5 × 1–1.5
Slice thickness [mm]	2–3	1–3	2–3	3–5

2.2 Federation setup and pre-processing

The federated training setup included a central server associated with Center A dataset and three clients from Centers B, C, and E. An overview of the proposed FL framework is illustrated in Figure 1. The federated infrastructure was implemented using Flower (28), PyTorch (29) and MONAI (30). The federated training followed a cross-silo, synchronous, and full-participation scheme. At each global round, all clients performed local updates simultaneously, and the server aggregated the global model parameters only after receiving updates from every client. Experiments were conducted on a single-node machine with an NVIDIA A100 (80GB), 16 CPU cores, and 128GB RAM. While all experiments were conducted within a single-node environment for prototyping purposes, the FL framework incorporated all the necessary components to replicate a real-world federated deployment, including encrypted communication using TLS certificates, client reconnection strategies, and the ability to resume training from saved checkpoints.

The Center A dataset was reserved exclusively for external validation, enabling an evaluation of the federated model’s generalization performance. Within each participating client, four validation patients and four test patients were held out, while the remaining cases were used to train local models. This allocation reflects a trade-off between maximizing client training data and maintaining a diagnostic set for monitoring local fitting behavior, with the primary generalization assessment assigned to the external cohort from Center A. Overall, of the 251 patients included in the study, 167 contributed to federated

training ($\approx 66.5\%$ of the total data), 24 were allocated to local validation and test sets across the three clients ($\approx 9.5\%$ of the total data), and 60 served as the independent external test cohort, representing $\approx 24\%$ of the total data.

Since data were released for a public challenge, several pre-processing operations were applied by the authors prior to the publication, such as isotropic resampling, defacing, and rigid registration between modalities, resulting in residual anatomical misalignments between CBCTs and CTs (26). To mitigate the impact of these residual misalignments on model evaluation, an additional deformable registration step was applied exclusively to the test patients.

To ensure consistent spatial resolution across the federation, each client applied a standardized pre-processing pipeline to locally adapt data to a $256 \times 256 \times 256$ image grid dimension. This decentralized approach—which avoided any inter-client data exchange—involved conditional cropping followed by resizing and constant-padding as needed. Furthermore, an intensity processing was applied. A local normalization process was performed on each client for each CBCT scan, employing a LUT-based normalization method proposed by Vicario et al. (31) with a fixed intensity range of $[-800, 2,000]$. Subsequently, CT scans were clipped to the range $[-1,000, 1,500]$ to mitigate the impact of extreme values introduced by metal implants. These ranges were empirically chosen by analyzing the CBCTs and CTs intensity distributions available at each client site, without exchanging any images.

This preliminary harmonization step allowed for the approximation of a common intensity range across clients while preserving the constraints of the federated paradigm.

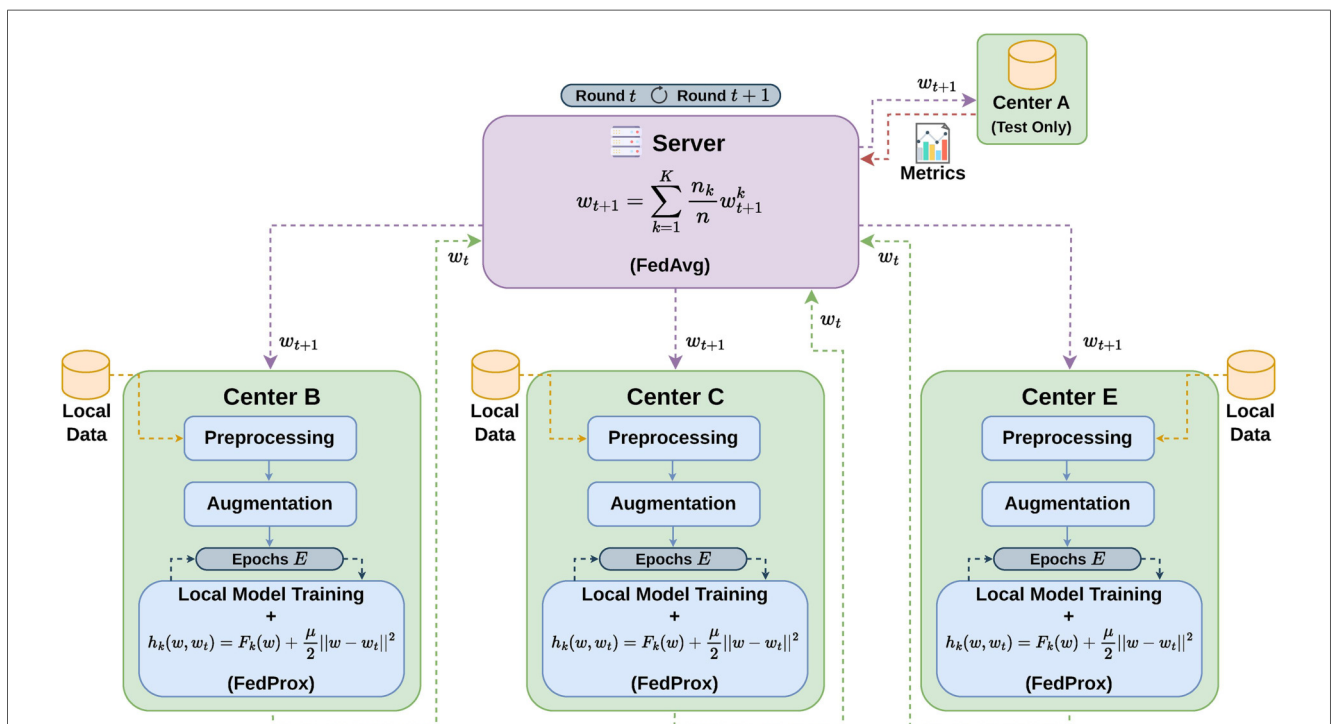


FIGURE 1 Schematic overview of the proposed FL framework. Three clients (Centers B, C, and E) implement a standardized preprocessing and data augmentation pipeline for their local datasets. Each client k performs local model training at round t , introducing the FedProx regularization. The locally updated weights (w_t) are subsequently transmitted to a central server, where they are aggregated via FedAvg. The resulting global model (w_{t+1}) is subsequently redistributed to all clients for the next round. The Center A dataset is exclusively reserved for external validation, thereby enabling an evaluation of the federated model’s generalization performance.

2.3 Deep learning model

The DL architecture employed relied on the cGAN paradigm, and more specifically, on the original Pix2Pix (11) architecture.³ Pix2Pix has been identified as one of the most widely adopted architectures for the CBCT-to-sCT translation task, as indicated in Section 1. The adversarial paradigm exhibited a remarkable capacity in managing imperfectly paired or partially aligned data, a prevalent scenario in clinical practice where precise voxel-wise correspondence between images cannot be assured, even after the registration process. This has led to the identification of Pix2Pix as a suitable and practical solution for our application setting.

The generator was structured as a 2D U-Net with eight encoder-decoder blocks connected via skip connections, and included dropout layers to mitigate overfitting. The discriminator was implemented as a 2D PatchGAN, composed of convolutional blocks combining 2D convolutions, instance normalization, and LeakyReLU activations to assess local realism at the patch level.

The training of the DL model was implemented using the Random-Multi2D sampling approach, to enhance the robustness and generalisability. This approach consisted of presenting to the model different axial, sagittal, and coronal slices in a randomized order, ensuring that the DL model received non-sequential input from multiple anatomical planes (14).

In detail, the discriminator's objective (Equation 1) was achieved by minimizing the sum of two sigmoid cross-entropy losses:

$$\mathcal{L}_D = \frac{1}{2} \times [\text{BCE}(\mathcal{D}(CT, CBCT), 1) + \text{BCE}(\mathcal{D}(sCT, CBCT), 0)] \quad (1)$$

where $\text{BCE}(\cdot)$ was the cross-entropy loss output, $\mathcal{D}(\cdot)$ was the discriminator output, and 1 and 0 were images of ones and zeros indicating real and fake labels, respectively. The first term penalized misclassifications of CTs as fake, while the second term penalized misclassifications of sCTs as real.

The adversarial loss function employed for the optimization of the generator (Equation 2) was combined by two components. The first was a sigmoid cross-entropy loss that enabled the discriminator to classify the generated images as real or fake:

$$\mathcal{L}_{GAN} = \text{BCE}(\mathcal{D}(sCT, CBCT), 1). \quad (2)$$

The second term was the L1 loss (Equation 3), specifically the Masked MAE (see Equation 7)—typically employed for this task as reported in Section 1—between the generated sCT and the ground-truth CT, which enforced structural similarity and penalized large deviations. This loss was also referred to as pixel loss:

$$\mathcal{L}_{\text{pixel}} = \text{MAE}(CT - sCT). \quad (3)$$

The total generator loss (Equation 4) was subsequently computed as:

$$\mathcal{L}_G = \mathcal{L}_{GAN} + \lambda_{\text{pixel}} \cdot \mathcal{L}_{\text{pixel}}, \quad (4)$$

where $\lambda_{\text{pixel}} = 100$ as reported in the original Pix2Pix study (11).

Inference was based on the median voxel-voting mechanism. Specifically, following the generation of the slice-wise predictions along each anatomical plane, the final sCT volume was computed by taking the median across the three orientations (1, 14). Both training and inference strategies were previously validated in the federated scenario within the FedSynthCT-Brain framework (14).

In each local client, the model was trained using a batch size of 32 slices before participating in a federated aggregation round. The Adam optimizer with a learning rate of 10^{-4} was used. An additional investigation was conducted to establish the optimal number of local epochs, comparing 1, 2 and 3 epochs per round.

To further improve robustness and cross-site generalization, a consistent data augmentation pipeline was applied at the client level. Each client independently introduced random geometric and intensity transforms during the training phase, including flipping, rotations, translations, and random rescaling or shifting of intensity values.

2.4 Aggregation strategy and sensitivity analysis

The combination of FedAvg and FedProx (32) was employed as aggregation strategy for this study. The aggregation was equivalent with the approach used in the FedSynthCT-Brain study, as it was identified as the optimal aggregation strategy, yielding results that are equivalent to results obtained by alternative strategies while requiring a reduced number of rounds (14). Nevertheless, in order to determine the optimal proximal coefficient μ and to verify that the hyperparameter choice originally established in the FedSynthCT-Brain study (14) remains appropriate for the CBCT-to-sCT synthesis task, a sensitivity analysis was conducted.

The FedAvg algorithm was implemented on the server side to aggregate the clients' weights, ensuring that the contribution of each client was proportional to the size of its dataset. Therefore, in the Equation 5, K represents the total amount of clients; n_k denotes the number of samples for a single client k ; n is the total amount of samples across the federation, and w_{t+1}^k represents the post-training weights of the client k .

$$w_{t+1} = \sum_{k=1}^K \frac{n_k}{n} w_{t+1}^k \quad (5)$$

The FedProx method was subsequently implemented on the client side, thereby introducing a proximal term, denoted by $\frac{\mu}{2} \|w - w_t\|^2$ to the local objective function (thus, the Masked MAE loss). This term penalized large deviations from the global model weights in order to control local update variations. The Equation 6 provides a formal explanation of the impact of the proximal term on the client model. The original local objective

³<https://github.com/phillipi/pix2pix>.

is denoted by $F_k(w)$, the pre-training global model weight by w_t , the weights during local training by w , and the proximal coefficient by μ . Consequently, FedProx mitigates update divergence and approximates FedAvg when $\mu = 0$, as evidenced in the original FedProx study (32).

$$h_k(w, w_t) = F_k(w) + \frac{\mu}{2} \|w - w_t\|^2 \quad (6)$$

To assess the sensitivity of the aggregation strategy to the proximal coefficient μ , four configurations were evaluated: (i) $\mu = 0$ (FedAvg), (ii) $\mu = 1$, (iii) $\mu = 3$, and (iv) $\mu = 5$. All other hyperparameters (presented in Section 2.3) were held constant across configurations. Each experiment was run for a total of 50 global rounds, ensuring sufficient training time for convergence and enabling fair comparisons across different μ configurations.

2.5 Evaluation

The performance of the federated model was evaluated using standard image similarity metrics—presented in Section 1—on the external test dataset (Center A) and the client test datasets (Centers B, C, and E). Specifically, the following metrics were employed:

- **Masked Mean Absolute Error:**

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |CT_i - sCT_i|, \quad (7)$$

where n is the number of voxels within the region of interest (ROI).

- **Masked Structural Similarity Index Measure (SSIM):**

$$\text{SSIM} = \frac{(2\mu_{sCT}\mu_{CT} + C_1)(2\sigma_{sCT,CT} + C_2)}{(\mu_{sCT}^2 + \mu_{CT}^2 + C_1)(\sigma_{sCT}^2 + \sigma_{CT}^2 + C_2)}, \quad (8)$$

where the mean (μ), variance/covariance (σ), and constants $C_1 = (k_1L)^2$, $C_2 = (k_2L)^2$ are computed over the ROI. Here, L is the dynamic range of the CT image, with $k_1 = 0.01$ and $k_2 = 0.03$.

- **Masked Peak Signal-to-Noise Ratio (PSNR):**

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_{CT}^2}{\text{MSE}} \right), \quad (9)$$

where MAX_{CT} is the maximum intensity value in the CT image and MSE is the mean squared error between CT and sCT within the ROI.

In order to assess the impact of registration errors on image similarity metrics, a registration quality analysis was conducted on the external test cohort from Center A, which was sufficiently large to enable a correlation analysis. For each patient, the Normalized Mutual Information (33) (NMI) between the CT and CBCT volumes was computed both before and after the additional deformable registration step. Thus, the

NMI was used as an intensity-based surrogate for registration quality. The patient-specific variation in NMI was defined as:

$$\Delta\text{NMI} = \text{NMI}_{\text{after}} - \text{NMI}_{\text{before}} \quad (10)$$

This variation was then correlated with the corresponding change in generation error:

$$\Delta\text{MAE} = \text{MAE}_{\text{after}} - \text{MAE}_{\text{before}} \quad (11)$$

where $\text{MAE}_{\text{after}}$ and $\text{MAE}_{\text{before}}$ denote the MAE (see Equation 7), obtained by using as input the CBCT volumes after and before the deformable registration step, respectively.

3 Results

The performance of the proposed federated cGAN model across multiple clinical centers was assessed through both quantitative and qualitative analyses, as reported in Table 2 and Figures 3–6, respectively.

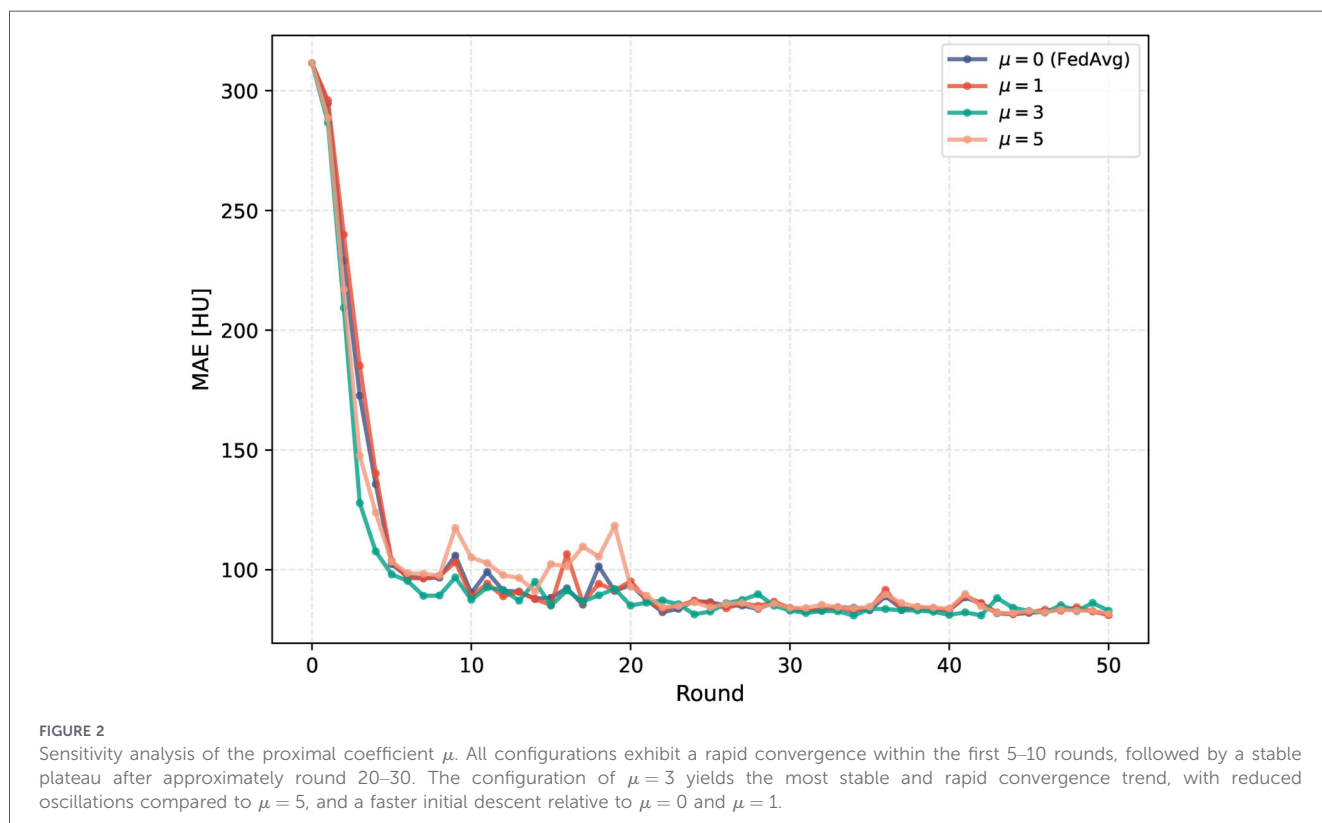
The convergence curves obtained across different μ configurations ($\mu = 0, 1, 3, 5$) are presented in Figure 2. In all configurations, the global MAE demonstrated a rapid decrease during the initial 5–10 global rounds, from ≈ 310 HU to values in the range of 90–130 HU, followed by a stable plateau beyond round 30–35, with round-to-round fluctuations remaining below ± 5 HU. The quantitative metrics across all configurations demonstrated high consistency (Table 2), with marginal variations in MAE, SSIM (Equation 8) and PSNR (Equation 9). Among the tested configurations, $\mu = 3$ exhibited the most stable convergence trajectory, with reduced oscillations during the intermediate rounds compared to $\mu = 5$, and a faster initial descent relative to $\mu = 0$ (FedAvg) and $\mu = 1$. Therefore, the $\mu = 3$ configuration was selected on the basis of its convergence properties, rather than on marginal differences in image quality metrics.

Subsequently, an additional investigation on the optimal number of local epochs was conducted (see Section 2.3). The convergence curves and quantitative metrics are provided in Supplementary Figure S1 and Supplementary Table S1, respectively. The federated training with a single local epoch led to a notable decrease in the stability of the global model convergence, exhibiting increased oscillations across rounds. As the number of local epochs per round increased (2 and 3), a proportional increase in computation time was observed, along with a progressively faster early descent and smaller oscillations in the intermediate rounds. Furthermore, a comparable outcome was observed in configurations with 2 and 3 local epochs, where convergence plateau was reached around round 30. However, no substantial differences were observed in the image similarity metrics. Indeed, MAE differences across centers ranged between 1–4 HU, with no single configuration consistently outperforming the others across all centers (see Supplementary Table S1). Consequently, given that increasing the number of local epochs failed to yield a substantial improvement in performance and instead led to an increase in the computational load on local clients, while a single epoch introduced instability in global convergence, adopting two local epochs was

TABLE 2 Quantitative performance across all centers for the four proximal coefficient μ configurations evaluated in the sensitivity analysis.

Center	Configuration	μ	MAE [HU]	SSIM	PSNR [dB]
Center A (external)	FedAvg	0	75.17 ± 12.54	0.90 ± 0.04	33.63 ± 2.18
	FedAvg + FedProx	1	75.57 ± 12.61	0.90 ± 0.04	33.61 ± 2.14
	FedAvg + FedProx	3	75.22 ± 11.81	0.90 ± 0.03	33.52 ± 2.06
	FedAvg + FedProx	5	75.80 ± 12.51	0.90 ± 0.04	33.59 ± 2.13
Center B	FedAvg	0	81.92 ± 5.89	0.88 ± 0.02	33.09 ± 0.93
	FedAvg + FedProx	1	82.22 ± 6.11	0.88 ± 0.02	33.07 ± 0.93
	FedAvg + FedProx	3	85.90 ± 7.10	0.88 ± 0.02	32.86 ± 0.94
	FedAvg + FedProx	5	82.15 ± 6.30	0.89 ± 0.02	33.06 ± 0.88
Center C	FedAvg	0	64.75 ± 14.38	0.92 ± 0.04	34.38 ± 1.38
	FedAvg + FedProx	1	65.02 ± 14.43	0.92 ± 0.04	34.38 ± 1.35
	FedAvg + FedProx	3	64.38 ± 13.63	0.92 ± 0.04	34.48 ± 1.39
	FedAvg + FedProx	5	65.86 ± 14.68	0.92 ± 0.04	34.27 ± 1.37
Center E	FedAvg	0	75.55 ± 13.40	0.91 ± 0.02	34.74 ± 1.03
	FedAvg + FedProx	1	74.71 ± 12.93	0.92 ± 0.01	34.81 ± 0.98
	FedAvg + FedProx	3	74.22 ± 13.23	0.92 ± 0.01	34.91 ± 1.04
	FedAvg + FedProx	5	74.14 ± 12.49	0.92 ± 0.01	34.90 ± 0.98

Results are reported as mean ± standard deviation. Center A was used exclusively for external testing, while Centers B, C, and E participated in the federated training.



determined to be an optimal trade-off, balancing the stability of the global convergence trend, the performance of the federated model, and the computational cost on clients.

Therefore, all the subsequent per-center results reported herein refer to the $\mu = 3$ and 2 local epochs configuration.

As detailed in Section 2.2, Center A dataset was excluded from the federated training process and served as an external test set, including 60 test cases that provided a more robust and less biased estimation of the model’s generalization capabilities. Centers B, C, and E contributed data to train the federated

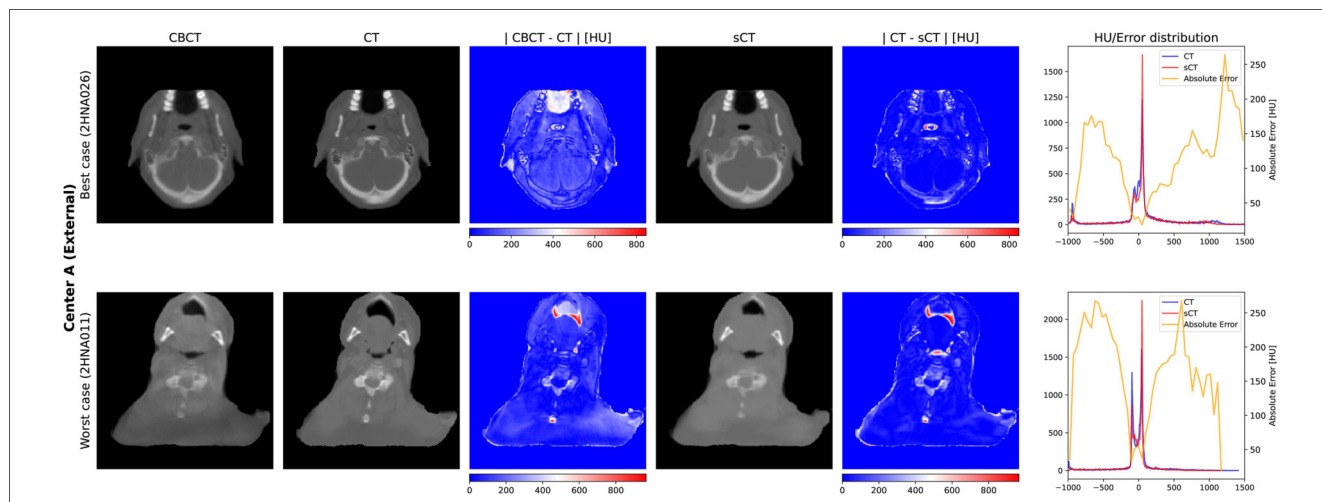


FIGURE 3 Visual evaluation of the best and worst sCT cases from Center A, used exclusively as an external test set. For each case, the mid-axial slice of the original CBCT, the ground-truth CT, the generated sCT, and the absolute difference maps for CBCT-to-CT and sCT-to-CT (error map) are shown. Additionally, intensity distributions of the sCT and CT for the displayed slice are presented, overlaid with the pixel-wise MAE. These distributions highlight the conversion accuracy across different tissue density ranges. Notably, regions with very high absolute differences appear in both CBCT-to-CT and error maps, suggesting the presence of residual misregistration between CBCT and CT scans.

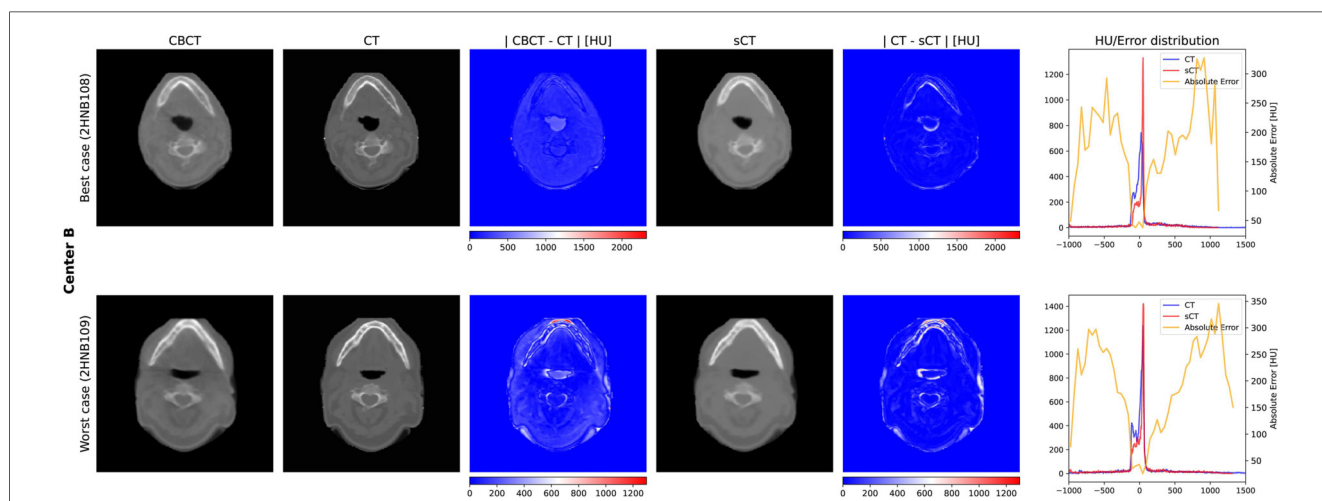


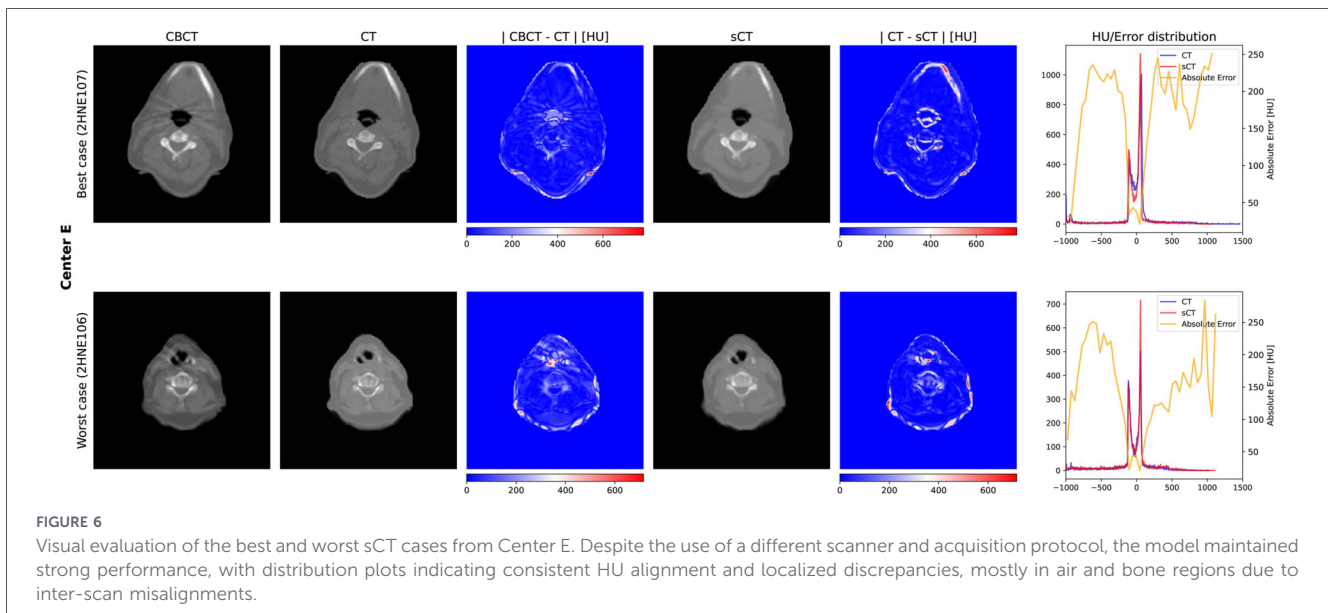
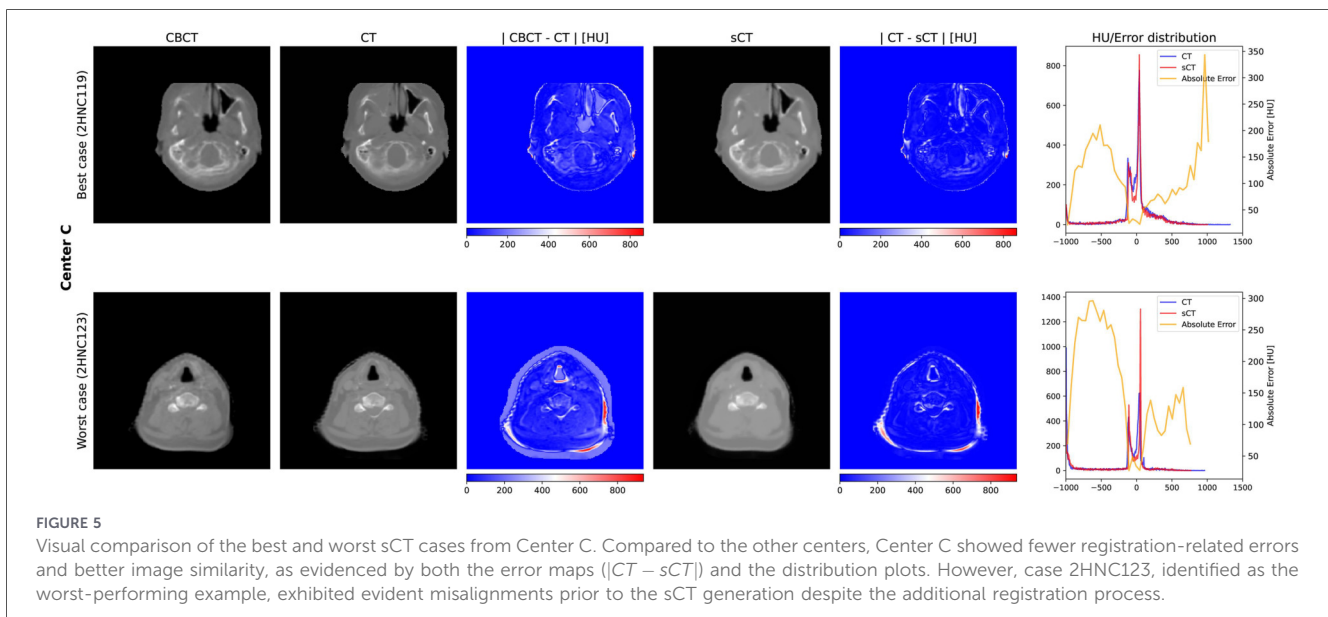
FIGURE 4 Qualitative results for the best and worst sCT cases from Center B, which showed lower image similarity metrics among the training centers. Widespread registration artifacts between CBCT and CT scans contributed to the visible errors. This was especially evident in the oral cavity of the 2HNB109 patient, and are reflected in both $|CBCT - CT|$ figure and error map.

model, with each center reserving four patients for local evaluation. As shown in Table 2 (for $\mu = 3$), despite not participating in the federated training, Center A achieved performance metrics comparable to those of the training centers, with a MAE of 75.22 ± 11.81 HU, an SSIM of 0.90 ± 0.03 , and a PSNR of 33.52 ± 2.06 dB. A qualitative evaluation of the best and worst test cases from Center A (Figure 3) showed that the generated sCTs aligned closely with the ground-truth CTs. The error maps, computed as $|CT - sCT|$, indicated that most deviations were within ± 100 HU. The absolute difference between the original CBCT and the ground-truth CT, reported as $|CBCT - CT|$ in Figure 3, highlighted spatial discrepancies mainly related to anatomical variations and registration mismatches.

Among the federated clients, Center B obtained a MAE = 85.90 ± 7.10 HU, an SSIM = 0.88 ± 0.02 , and a PSNR = 32.86 ± 0.94 dB. Qualitative results (Figure 4) demonstrated that the generated sCTs were anatomically consistent with the ground-truth CTs, particularly in the representation of major structures.

Center C yielded the best overall quantitative performance, with the lowest MAE (64.38 ± 13.63 HU), highest SSIM (0.92 ± 0.04), and highest PSNR (34.48 ± 1.39 dB). The worst-case example (2HNC123) shown in Figure 5 showed that the dominant discrepancies were associated with misalignments between CBCT and CT rather than pure generation errors.

Center E achieved results consistent with the overall federation despite images being acquired using a different scanner and



acquisition protocol (Table 1). The federated model achieved a $MAE = 74.22 \pm 13.23$ HU, an $SSIM = 0.92 \pm 0.01$, and a $PSNR = 34.91 \pm 1.04$ dB (Table 2). Furthermore, quantitative results confirmed the quality of the generated sCTs, as shown in Figure 6.

The results of the registration quality analysis described in Section 2.5 are presented in Figure 7. A significant negative correlation between ΔNMI (Equation 10) and ΔMAE (Equation 11) was observed in the Center A cohort, with a Pearson coefficient of $r = -0.63$ ($p = 5.95 \times 10^{-8}$) and a Spearman coefficient of $\rho = -0.61$ (2.35×10^{-7}).

4 Discussion

In this work, we investigated whether a cross-silo FL framework can support robust CBCT-to-sCT synthesis in the

head and neck region under realistic multi-center constraints. Specifically, we extended the FedSynthCT (14) study to a different imaging modality and a more anatomically heterogeneous region, evaluating its performance across three federated European centers and an independent external validation cohort from the SynthRAD2025 dataset (27). By analyzing both quantitative metrics and qualitative image characteristics, we aimed to determine whether federated training can preserve generalization capability despite inter-institutional variability in scanners, acquisition protocols, and anatomical presentation, without requiring data sharing or center-specific retraining.

A sensitivity analysis was conducted on the proximal coefficient μ , which confirmed the suitability of $\mu = 3$ for the current task in accordance with the findings of FedSynthCT-Brain (14), despite the acknowledged differences in imaging modality, anatomical region, and scanner heterogeneity. As

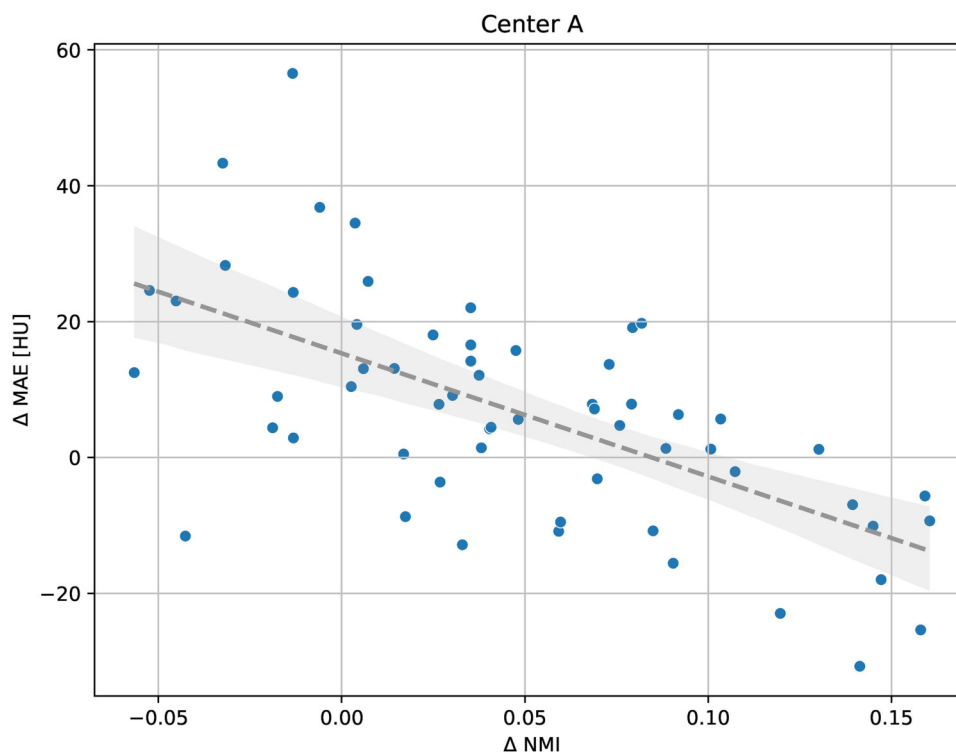


FIGURE 7

Correlation between Δ NMI and Δ MAE (HU) for the Center A test cohort before and after the deformable image registration. A significant negative correlation was observed (Pearson $r = -0.63$, 5.95×10^{-8} ; Spearman $\rho = -0.61$, $p = 2.35 \times 10^{-7}$). This finding indicates that enhancements in deformable registration, consequently leading to enhanced intensity correspondence between CT and CBCT, are associated with a decrease in MAE. The deterioration of NMI corresponds to MAE increases of ≈ 20 – 60 HU (upper-left area). In contrast, cases for which NMI improved substantially Δ NMI $\approx 10\%$ – 15% exhibit a consistent MAE reduction (lower-right area).

demonstrated in Figure 2, the convergence behavior exhibited that moderate regularization ($\mu = 3$) accelerated and stabilized convergence in comparison to FedAvg ($\mu = 0$), $\mu = 1$ and $\mu = 5$. Indeed, for $\mu < 3$ the convergence was slower and exhibited greater fluctuations in the intermediate rounds (approximately rounds 9–20), while more severe regularization ($\mu = 5$) introduced instability in the intermediate rounds, delaying optimization. Consequently, this finding indicates that proximal regularization is beneficial, but leads to over-constraining local updates and hinders global convergence when severe. Notably, the minimal variations in the final image quality metrics across all μ configurations (Table 2), suggest that the convergence benefit of $\mu = 3$ occurs without compromising final performance. Furthermore, the consistent plateau behavior observed across all configurations demonstrated that 50 rounds were sufficient to ensure stable convergence of the federated model.

The local epochs analysis revealed that a single local epoch introduced instability and slowed convergence (see Supplementary Figure S1). The configurations with 2 and 3 local epochs yielded fewer oscillations across rounds. However, no substantial improvement in image quality metrics was observed, with MAE differences ranging from 1 to 4 HU across centers and no consistent hierarchy among configurations. Moreover, minimizing local updates in the context of FL is typically advantageous to mitigate client drift (24). Although FedProx partially constrains local divergence, its regularizing effect could

diminish as the number of local epochs increases (32). Consequently, 2 local epochs were selected as the configuration that optimally balanced convergence stability, model performance, and computational cost on local clients.

External validation on Center A enabled further investigation into the generalization capability of the federated model in a fully independent setting. The observed performance metrics were comparable to those obtained across the participating centers, despite the absence of additional training or fine-tuning. These findings suggest stable model behavior on previously unseen data. Furthermore, qualitative analysis indicated that residual spatial discrepancies (Figure 3) between CBCT and planning CT were predominantly associated with registration inaccuracies and anatomical variations occurring during treatment. For instance, anatomical differences in the oral cavity were observed in patient 2HNA011 (Figure 3); these misalignments were propagated to the generated sCT, affecting its apparent visual quality and negatively influencing the quantitative metrics reported in Table 2.

Across the federated centers, performance differences were primarily attributed to image misalignments, despite the inclusion of an additional deformable registration step for the evaluation data. The lower performance at Center B relative to Centers C, E, and the external Center A was mainly attributable to a higher prevalence of misalignments. In contrast, the superior performance at Center C reflected a reduced influence of registration errors, resulting in higher image similarity

metrics. Taken together, these observations suggest that registration quality, rather than the inherent capability of the generative model, represented a primary limiting factor in the evaluation. To provide further analytical insights, an additional investigation was conducted on the Center A test cohort using NMI as an intensity-based surrogate for registration quality (see Section 2.5). Although NMI lacks direct geometric quantification of spatial misalignment, it measures intensity correspondence between CT and CBCT volumes. A significant negative correlation was identified between Δ NMI and Δ MAE. Furthermore, Figure 7 supports the hypothesis that residual misalignment systematically affects sCT evaluation. In detail, when deformable registration substantially improved intensity correspondence (Δ NMI \approx 10%–15%), a corresponding MAE reduction was observed. In contrast, when NMI deteriorated, MAE degraded by approximately 20–60 HU (upper-left area of Figure 7).

However, the robustness of the federated model was further reflected by the results obtained at Center E, where data were acquired using a different scanner and acquisition protocol. Despite this domain shift, the model maintained performance levels consistent with the federation, highlighting its capacity to adapt across heterogeneous scanners. A global analysis across all centers, focusing on anatomical structures and HU distributions, demonstrated that the model preserved tissue-specific intensity patterns (Figures 3–6). The largest errors were consistently observed in tissues occupying smaller volumes in the head and neck region, such as air cavities and dense bone, which were underrepresented in the training data relative to soft tissues. These regions are intrinsically more challenging to predict accurately and, due to their extreme HU values, disproportionately affected image similarity metrics.

A limitation of this study concerns the quantitative evaluation strategy. As demonstrated by the NMI-MAE correlation analysis, residual misalignment systematically affected the MAE. However, its independent contribution could not be isolated precisely. Furthermore, the local test sets comprised four patients per client, with the purpose of assessing the federated model fitting at the client level. This methodological approach inherently introduces higher variability in performance estimates. Nevertheless, the findings in Table 2 and Figures 3–6 demonstrated consistent performance of the federated model across both local and external testing, indicating its capacity to maintain comparable outcomes for the clients and the independent center.

In addition, although the integration of FL-generated sCTs into RT treatment planning represents a potential clinical application of this work, its direct evaluation was not feasible in this study due to the absence of dosimetric and planning data in the SynthRAD2025 dataset, which contains imaging data only. This scenario also reflects a realistic clinical setting in which sensitive RT data are not shared, further underlining the relevance of FL as a framework for privacy-preserving multi-institutional model development. Thus, future research may focus on evaluating the clinical applicability of sCTs generated through FL. Furthermore, FL does not inherently guarantee complete privacy protection. Potential information leakage from shared model updates has been reported in the literature, including through gradient and model inversion attacks (34, 35). Mitigation strategies such as secure aggregation or differential

privacy can further reduce these risks (34, 35). While these strategies have not been explicitly examined in this work, they remain an important area for future research in federated sCT generation.

From a methodological perspective, federated and centralized DL represent distinct optimization paradigms rather than interchangeable training strategies. As previously observed in the FedSynthCT-Brain (14) study, methodological configurations that are effective in centralized settings may not necessarily translate directly to federated environments, where statistical heterogeneity, client-specific data distributions, and aggregation constraints fundamentally shape the learning dynamics. Centralized training optimizes performance under full data availability and implicitly benefits from pooled distributions, whereas FL aims to achieve stable cross-institutional performance while preserving data privacy and respecting structural data-sharing constraints. FL also introduces additional communication and synchronization overhead compared to centralized training, which may increase overall training time, although this cost is often offset by the practical benefits of decentralized data access and improved feasibility in multi-institutional settings. Within this framework, moderate center-specific performance trade-offs should be interpreted in light of the federated objective: promoting generalization across heterogeneous institutions rather than maximizing performance on any single site. Federated models may also serve as robust shared initializations that can be further adapted through optional local fine-tuning when institution-specific optimization is required. Accordingly, FL should be regarded not as a substitute for centralized training, but as a complementary collaborative framework tailored to clinical scenarios in which data centralization is impractical or restricted.

Our findings suggest potential directions for future research. Within the GAN paradigm, more targeted training strategies may improve the synthesis of anatomically underrepresented structures. The multi-discriminator framework proposed by Kumar et al. (36) could represent a viable approach. In parallel, task-specific loss formulations, such as those introduced by Liang et al. (37) to improve HU fidelity, noise suppression, and dosimetric accuracy, could be integrated into the FL aggregation scheme. However, this would introduce additional challenges related to client-level heterogeneity and personalization. Furthermore, beyond GAN-based architectures, the investigation of diffusion-based models and hybrid CNN-transformer approaches (38) within a federated framework represents a promising direction for improving both image quality and structural consistency in FL-generated sCTs.

In conclusion, this study demonstrated the feasibility and effectiveness of a cross-silo FL approach for CBCT-to-sCT synthesis in the head and neck region, using the publicly available SynthRAD2025 challenge dataset (26). The federated model generalized effectively across the participating centers and on an external validation cohort, maintaining consistent performance despite differences in imaging protocols, scanner types, and acquisition parameters. This mitigates the impact of institutional heterogeneity, a common limitation in single-center model deployments. Furthermore, these results align with requirements outlined in recent related work (8), which

emphasized that a model should ideally be applicable across different centers without the need for conditional tuning.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://zenodo.org/doi/10.5281/zenodo.15373853>.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

CBR: Software, Visualization, Data curation, Writing – original draft, Investigation, Conceptualization, Validation, Methodology, Writing – review & editing, Formal analysis. PZ: Visualization, Data curation, Writing – review & editing. MS: Writing – review & editing, Data curation, Visualization, Conceptualization, Validation.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

References

- Spadea MF, Maspero M, Zaffino P, Seco J. Deep learning based synthetic-CT generation in radiotherapy and PET: a review. *Med Phys.* (2021) 48:6537–66. doi: 10.1002/mp.15150
- Boldrini L, D'Aviero A, De Felice F, Desideri I, Grassi R, Greco C, et al. Artificial intelligence applied to image-guided radiation therapy (IGRT): a systematic review by the Young Group of the Italian Association of Radiotherapy and Clinical Oncology (yAIRO). *Radiol Med.* (2023) 129:133–51. doi: 10.1007/s11547-023-01708-4
- Altalib A, McGregor S, Li C, Perelli A. Synthetic CT image generation from CBCT: a systematic review. *IEEE Trans Radiat Plasma Med Sci.* (2025) 9(6): 1. doi: 10.1109/TRPMS.2025.3533749
- Landry G, Hua C. Current state and future applications of radiological image guidance for particle therapy. *Med Phys.* (2018) 45:1086–95. doi: 10.1002/mp.12744
- Rabe M, Kurz C, Thummerer A, Landry G. Artificial intelligence for treatment delivery: image-guided radiotherapy. *Strahlenther Onkol.* (2025) 201:283–97. doi: 10.1007/s00066-024-02277-9
- Riou O, Prunaretty J, Michalet M. Personalizing radiotherapy with adaptive radiotherapy: Interest and challenges. *Cancer/Radiothérapie.* (2024) 28:603–9. doi: 10.1016/j.canrad.2024.07.007
- Mastella E, Calderoni F, Manco L, Ferioli M, Medoro S, Turra A, et al. A systematic review of the role of artificial intelligence in automating computed tomography-based adaptive radiotherapy for head and neck cancer. *Phys Imaging Radiat Oncol.* (2025) 33:100731. doi: 10.1016/j.phro.2025.100731
- Huijben EM, Terpstra ML, Galapon AJ, Pai S, Thummerer A, Koopmans P, et al. Generating synthetic computed tomography for radiotherapy: SynthRAD2023 challenge report. *Med Image Anal.* (2024) 97:103276. doi: 10.1016/j.media.2024.103276
- Rusanov B, Hassan GM, Reynolds M, Sabet M, Kendrick J, Rowshanfarzad P, et al. Deep learning methods for enhancing cone-beam CT image quality toward adaptive radiation therapy: a systematic review. *Med Phys.* (2022) 49:6019–54. doi: 10.1002/mp.15840
- Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing (2015). p. 234–41.
- Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. *arXiv [Preprint]* (2018). doi: 10.48550/arXiv.1611.07004
- Chen X, Liu Y, Yang B, Zhu J, Yuan S, Xie X, et al. A more effective CT synthesizer using transformers for cone-beam CT-guided adaptive radiotherapy. *Front Oncol.* (2022) 12:988800. doi: 10.3389/fonc.2022.988800
- La Greca Saint-Estevan A, Dal Bello R, Lapaeva M, Fankhauser L, Pouymayou B, Konukoglu E, et al. Synthetic computed tomography for low-field magnetic resonance-only radiotherapy in head-and-neck cancer using residual vision transformers. *Phys Imaging Radiat Oncol.* (2023) 27:100471. doi: 10.1016/j.phro.2023.100471
- Raggio CB, Zabaleta MK, Skupien N, Blanck O, Cicone F, Cascini GL, et al. FedSynthCT-Brain: a federated learning framework for multi-institutional brain MRI-to-CT synthesis. *Comput Biol Med.* (2025) 192:110160. doi: 10.1016/j.combiomed.2025.110160

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fdgth.2026.1812254/full#supplementary-material>

15. Kaisis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat Mach Intell.* (2020) 2:305–11. doi: 10.1038/s42256-020-0186-1
16. Guan H, Yap PT, Bozoki A, Liu M. Federated learning for medical image analysis: a survey. *Pattern Recognit.* (2024) 151:110424. doi: 10.1016/j.patcog.2024.110424
17. Sandhu SS, Gorji HT, Tavakolian P, Tavakolian K, Akhbardeh A. Medical imaging applications of federated learning. *Diagnostics.* (2023) 13:3140. doi: 10.3390/diagnostics13193140
18. Dalmaz O, Mirza MU, Elmas G, Ozbey M, Dar SU, Ceyani E, et al. One model to unite them all: personalized federated learning of multi-contrast MRI synthesis. *Med Image Anal.* (2024) 94:103121. doi: 10.1016/j.media.2024.103121
19. Hernandez-Cruz N, Saha P, Sarker MMK, Noble JA. Review of federated learning and machine learning-based methods for medical image analysis. *Big Data Cogn Comput.* (2024) 8:99. doi: 10.3390/bdcc8090099
20. Sheller MJ, Reina GA, Edwards B, Martin J, Bakas S. Multi-Institutional deep learning modeling without sharing patient data: a feasibility study on brain tumor segmentation. *Brainlesion.* (2019) 11383:92–104. doi: 10.1007/978-3-030-11723-8_9
21. Li W, Milletari F, Xu D, Rieke N, Hancox J, Zhu W, et al. Privacy-preserving federated brain tumour segmentation. In: Suk HI, Liu M, Yan P, Lian C, editors. *Machine Learning in Medical Imaging.* Cham: Springer International Publishing (2019). p. 133–41.
22. Li D, Kar A, Ravikumar N, Frangi AF, Fidler S. Fed-Sim: federated simulation for medical imaging. *arXiv [Preprint]* (2020). doi: 10.48550/arXiv.2009.00668
23. Chang Q, Qu H, Zhang Y, Sabuncu M, Chen C, Zhang T, et al. Synthetic learning: learn from distributed asynchronous discriminator GAN without sharing medical image data. *arXiv [Preprint]* (2020). doi: 10.48550/arXiv.2006.00080
24. McMahan B, Moore E, Ramage D, Hampson S, Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: Singh A, Zhu J, editors. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research*, Vol. 54. PMLR (2017). p. 1273–82.
25. Wang J, Xie G, Huang Y, Lyu J, Zheng F, Zheng Y, et al. FedMed-GAN: federated domain translation on unsupervised cross-modality brain image synthesis. *Neurocomputing.* (2023) 546:126282. doi: 10.1016/j.neucom.2023.126282
26. Thummerer A, van der Bijl E, Galapon AJ, Kamp F, Savenije M, Muijs C, et al. SynthRAD2025 grand challenge dataset: generating synthetic CTs for radiotherapy from head to abdomen. *Med Phys.* (2025a) 52:e17981. doi: 10.1002/mp.17981
27. Thummerer A, van der Bijl E, Galapon AJ, Kamp F, Maspero M. Data from: SynthRAD2025 grand challenge dataset: training (2025). doi: 10.5281/ZENODO.15373853
28. Beutel DJ, Topal T, Mathur A, Qiu X, Fernandez-Marques J, Gao Y, et al. Flower: a friendly federated learning research framework. *arXiv [Preprint]* (2022). doi: 10.48550/arXiv.2007.14390
29. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library.* Red Hook: Curran Associates Inc. (2019).
30. Cardoso MJ, Li W, Brown R, Ma N, Kerfoot E, Wang Y, et al. MONAI: an open-source framework for deep learning in healthcare. *arXiv [Preprint]* (2022). doi: 10.48550/arXiv.2211.02701
31. Vicario CM, Kordon F, Denzinger F, Barbari JSE, Privalov M, Franke J, et al. Normalization techniques for CNN based analysis of surgical cone beam CT volumes. In: *Medical Imaging 2022: Image Processing* (2022). p. 85. doi: 10.1117/12.2606973
32. Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated optimization in heterogeneous networks. *arXiv [Preprint]* (2020). doi: 10.48550/arXiv.1812.06127
33. Studholme C, Hill D, Hawkes D. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit.* (1999) 32:71–86. doi: 10.1016/S0031-3203(98)00091-0
34. Hu K, Gong S, Zhang Q, Seng C, Xia M, Jiang S. An overview of implementing security and privacy in federated learning. *Artif Intell Rev.* (2024) 57:204. doi: 10.1007/s10462-024-10846-8
35. Feng Y, Guo Y, Hou Y, Wu Y, Lao M, Yu T, et al. A survey of security threats in federated learning. *Complex Intell Syst.* (2025) 11:165. doi: 10.1007/s40747-024-01664-0
36. Kumar V, Sharma M, Jehadeesan R, Venkatraman B, Sheet D. Simulating cross-modal medical images using multi-task adversarial learning of a deep convolutional neural network. *Int J Imaging Syst Technol.* (2024) 34:e23113. doi: 10.1002/ima.23113
37. Liang Z, Wei H, Liu G, Cheng M, Gao J, Li S, et al. Leveraging GAN-based CBCT-to-CT translation models for enhanced image quality and accurate photon and proton dose calculation in adaptive radiotherapy. *J Radiat Res Appl Sci.* (2024) 17:100809. doi: 10.1016/j.jrras.2023.100809
38. Hu C, Cao N, Li X, He Y, Zhou H. CBCT-to-CT synthesis using a hybrid U-Net diffusion model based on transformers and information bottleneck theory. *Sci Rep.* (2025) 15:10816. doi: 10.1038/s41598-025-92094-6