Deep learning-based segmentation of head and neck organs at risk on CBCT images with dosimetric assessment for radiotherapy

Lucía Cubero, Cédric Hémon, Anais Barateau, Joël Castelli, Renaud de Crevoisier, Oscar Acosta, Javier Pascau

Physics in Medicine & Biology

PAPER • OPEN ACCESS

Deep learning-based segmentation of head and neck organs at risk on CBCT images with dosimetric assessment for radiotherapy

To cite this article: Lucía Cubero *et al* 2025 *Phys. Med. Biol.* **70** 075014

IOP Publishing Phys. Med. Biol. 70 (2025) 075014 https://doi.org/10.1088/1361-6560/adbf63

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

Deep learning-based segmentation of head and neck organs at risk on CBCT images with dosimetric assessment for radiotherapy

Lucía Cubero[1,2,*], **Cédric Hémon**[2], **Anaïs Barateau**[2], **Joël Castelli**[2], **Renaud de Crevoisier**[2], **Oscar Acosta**[2] and **Javier Pascau**[1,3]

[1] Departamento de Bioingeniería, Universidad Carlos III de Madrid, Madrid, Spain
[2] Université Rennes, CLCC Eugène Marquis, Inserm, LTSI—UMR 1099, F-35000 Rennes, France
[3] Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
[*] Author to whom any correspondence should be addressed.

E-mail: lcubero@ing.uc3m.es

Keywords: CBCT, OAR, segmentation, head and neck, radiotherapy, deep learning

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
29 October 2024

REVISED
28 February 2025

ACCEPTED FOR PUBLICATION
11 March 2025

PUBLISHED
26 March 2025

Abstract

Objective. Cone beam computed tomography (CBCT) has become an essential tool in head and neck cancer (HNC) radiotherapy (RT) treatment delivery. Automatic segmentation of the organs at risk (OARs) on CBCT can trigger and accelerate treatment replanning but is still a challenge due to the poor soft tissue contrast, artifacts, and limited field-of-view of these images, alongside the lack of large, annotated datasets to train deep learning (DL) models. This study aims to develop a comprehensive framework to segment 25 HN OARs on CBCT to facilitate treatment replanning. *Approach.* The proposed framework was developed in three steps: (i) refining an in-house framework to segment 25 OARs on CT; (ii) training a DL model to segment the same OARs on synthetic CT (sCT) images derived from CBCT using contours propagated from CT as ground truth, integrating high-contrast information from CT and texture features of sCT; and (iii) validating the clinical relevance of sCT segmentations through a dosimetric analysis on an external cohort. *Main results.* Most OARs achieved a dice score coefficient over 70%, with mean average surface distances of 1.30 mm for CT and 1.27 mm for sCT. The dosimetric analysis demonstrated a strong agreement in the mean dose and D2 (%) values, with most OARs showing non-significant differences between automatic CT and sCT segmentations. *Significance.* These results support the feasibility and clinical relevance of using DL models for OAR segmentation on both CT and CBCT for HNC RT.

1. Introduction

Head and neck cancer (HNC) is the sixth most prevalent cancer worldwide, with nearly one million new cases and 500 000 deaths reported annually (Bray *et al* 2024). Radiotherapy (RT) plays a central role in HNC treatment, employed in about 75% of patients as a primary or adjuvant therapy or combined with chemotherapy (Alfouzan 2021). Despite its effectiveness, RT also exposes surrounding healthy tissues—referred to as organs at risk (OARs)—to radiation, which can lead to the development of multiple toxicities. Achieving precise RT delivery is crucial to maximizing tumor control while minimizing damage to OARs, which is particularly challenging in HNC given the anatomical complexity of the region.

Cone beam computed tomography (CBCT) has become an essential tool in RT treatment delivery (Iliopoulos *et al* 2023), primarily assisting in patient positioning and alignment with the planned treatment, reducing the margins of the target volume and the radiation exposure to the OARs. The development of adaptive RT (ART) has further enhanced the clinical value of CBCT (Guberina *et al* 2024). ART is a personalized treatment strategy that adjusts the radiation plan during therapy to account for anatomical changes, such as weight loss and tumor or gland shrinkage, which can alter the dose distribution (Dona Lemus *et al* 2024). CBCT provides intra-fraction imaging of tumors and OARs, allowing clinicians to

identify possible deviations in anatomy and dose distribution that may trigger the need for offline or online treatment replanning. Offline ART, the current standard for most centers, involves recalculating the dose on a new CT scan when significant anatomical or dosimetric changes are detected on CBCT. Online ART, exemplified by Varian's Ethos system, allows real-time treatment replanning directly on CBCT, integrating AI-driven segmentation and advanced dose calculation (MacDonald *et al* 2024, Robar *et al* 2024). In Ethos version 1.0, influencer OARs are first segmented on the CBCT with a neural network, and then the planning CT (pCT) is registered to the CBCT to generate a synthetic CT (sCT) using structure-guided registration to propagate the target and remaining OARs to the sCT (Archambaul *et al* 2020, Kisling *et al* 2022, Guberina *et al* 2024). The latest version of Ethos can be coupled with HyperSight imaging, which generates high-quality CBCT with Hounsfield units comparable to CT, enabling direct AI segmentation of a wider range of OARs and replanning without the need for deformable registration (Yashayaeva *et al* 2024). However, due to the substantial cost of these systems (Kunnen *et al* 2024), widespread adoption will take time, and many institutions will continue using Ethos version 1.0 with its registration-based limitations or other CBCT-based offline workflows. Therefore, there still exists a gap for accurate segmentation of target volumes and OARs on CBCT, particularly for offline ART.

In recent years, deep learning (DL) has achieved state-of-the-art results for OAR segmentation on CT. However, relatively little effort has been made to automate this process on CBCT, mainly due to their lower image quality and the scarcity of large, well-labeled datasets to train DL models (Rabe *et al* 2024). Most research has focused on pelvic imaging, with early approaches training on CT and pseudo-CBCT images (Schreier *et al* 2020, Abbani *et al* 2022) or applying extensive data augmentation to make CT and CBCT scans more alike (Brion *et al* 2021). Other methods trained exclusively on manually segmented CBCT images (Zhang *et al* 2022, Hirashima *et al* 2023) or incorporated synthetic magnetic resonance (MR) images generated from CBCT (Fu *et al* 2020).

For HNC RT, a few DL approaches have also been proposed for automatic OAR segmentation on CBCT. Dai *et al* (2021a) trained a dual pyramid network on CBCT and synthetic MR images, achieving dice score coefficients (DSC) ranging from 0.66 to 0.96 for 12 OARs. Chen *et al* (2021) enhanced CBCT quality with an in-house network, reporting a mean DSC of 0.83 for five OARs between the manually delineated segmentations on the enhanced CBCTs and the rigidly registered CT contours. Ma *et al* (2022) introduced a registration module before training a U-Net on CBCT with registered CT contours, achieving a mean DSC of 0.87 for seven OARs. Liang *et al* (2023) proposed a three-step framework, combining pseudo-labels from CT, influencer volumes, and a small dataset of manually delineated CBCTs, reporting DSCs from 0.72 to 0.95 for 19 OARs. Liu *et al* (2023) developed a generalized CT segmentation model, personalized for each patient by refining it with its manually delineated CT to segment the clinical target volume on sCT images generated from CBCTs, achieving a DSC of 0.91. Finally, Zhao *et al* (2024) proposed the LSTM-UNet, trained sequentially on delineated pCTs while incorporating sCTs derived from CBCT and the predicted, corrected contours at each fraction, reaching DSCs between 0.85 and 0.94 for eight OARs. However, they trained on a simulated database of CBCTs and tested their model on sCTs generated by the Ethos system, which deforms the pCTs to the CBCTs, rather than directly generating sCTs from CBCTs.

Although these studies have demonstrated promising results, they often rely on well-labeled CBCT or MR databases, which are rarely segmented for RT purposes (McGee *et al* 2024). In contrast, larger labeled CT databases are more readily available due to their routine use in RT. To address the existing gap in automatic HN OAR segmentation on CBCT, our study builds on recent approaches (Liu *et al* 2023, Zhao *et al* 2024), combining an extensive, partially-labeled CT database with paired, unlabeled CBCT scans acquired during RT treatment.

The aim of this study was to develop a robust model for segmenting 25 HN OAR on CBCT and validate the results through comprehensive dosimetric analysis. This validation is critical for confirming the clinical utility of segmented CBCT images for dose comparison and treatment replanning, focusing on offline ART applications rather than online ART. This work was structured around three key objectives: (i) to adapt an in-house CT segmentation framework to include 25 HN OARs; (ii) to apply domain adaptation for segmenting these 25 OARs on sCT images generated from CBCT; and (iii) to perform a dosimetric analysis in an external database to ensure the clinical relevance of the sCT segmentations.

2. Materials and methods

2.1. Database

263 patients were retrospectively selected from a partially-labeled database (table 1). All patients were adults (aged >18 and <75 years) diagnosed with HNC and treated with external beam RT at the CLCC Eugène Marquis in Rennes, France. The training set included 203 HNC patients, treated with a standard protocol and routine CBCT imaging. This resulted in 385 CT and 963 CBCT scans, from which 196 CT-CBCT pairs

Table 1. Database characteristics for train and test sets.

	TRAIN SET	TEST SET
Patient population	203 HNC with various tumor locations	60 HNC from ARTIX study with oropharynx carcinoma
# CT images	385	230
CT dimensions	512 × 512 × (79–488) voxels	
CT spacing	(0.578–1.367) × (0.578–1.367) × (1–3) mm	
# CBCT images	963	166
CBCT dimensions	407 × 407 × 131 voxels	
CBCT spacing	1 × 1 × 2 mm	
# CT—CBCT pairs	196	166

were selected based on temporal proximity criteria: each CBCT was paired with a CT scan acquired either within three weeks prior to treatment or within one week during RT, to minimize anatomical differences between paired images. For the test set, 60 patients from the ARTIX study (ART to Decrease Xerostomia in Oropharynx Carcinoma) (Castelli *et al* 2023) were selected. These patients had locally advanced carcinoma of the oropharynx limited to T3-T4 and N2-N3, and underwent RT over 35 fractions. 28 patients received weekly replanning alongside CT and CBCT imaging, while the remainder followed a standard acquisition protocol with routine CBCT acquisition. This resulted in 230 CT and 166 CBCT scans, including 166 CT-CBCT pairs. The ARTIX cohort provided a unique dataset with high-quality manual delineations thoroughly reviewed by radiation oncologists, along with a large number of CT-CBCT pairs with minimal anatomical discrepancies.

CT images had standard dimensions of 512 × 512 voxels in the x and y axes, with an average of 195 slices in the z-axis. The voxel spacing averaged 1.156 mm for the x and y axes and 2.06 mm in slice thickness. Each CT scan contained between five and 30 manually delineated OARs, performed according to GORTEC guidelines (Tao *et al* 2020) and thoroughly reviewed by expert radiation oncologists. All CBCT images were acquired with the Elekta Versa HD system and had the same dimensions and voxel spacing. Non-paired CBCT scans were excluded to prevent potential anatomical discrepancies between longitudinal images that could compromise segmentation accuracy (Dona Lemus *et al* 2024). All CT and CBCT scans were acquired as part of the RT workflow, with patients positioned consistently using immobilization masks. Patient data information is summarized in table 1.

2.2. OAR segmentation on CT images

The first objective of this study was to extend a previously developed and validated DL-based framework for segmenting 15 OARs on CT images (Cubero *et al* 2022) to include 25 OARs. This original framework was trained on a partially-labeled, longitudinal database and employed pseudo-contours for unlabeled structures, training on all available data in a multi-class approach. The 25 OARs included in this study were: brain, brainstem, spinal canal, right/left (R/L) eyeballs, R/L optic nerves, R/L inner ears, R/L parotid glands, mandible, R/L temporomandibular joints, oral cavity, lips, R/L submandibular glands, larynx, upper, middle, and lower pharyngeal constrictor muscles (PCM), thyroid gland, esophagus, and esophageal inlet.

To extend the CT segmentation framework, we first preprocessed the CT images by resampling them to the mean spacing of the database (1.156 × 1.156 × 2 mm), cropping them to dimensions of 256 × 256 × 192 voxels, and normalizing the intensity values using Z-score normalization. The training methodology comprised several stages, as detailed in (Cubero *et al* 2022):

1. **OAR-Specific Model Training**: an OAR-specific model was trained for each structure using only the available manual segmentations for that OAR, which ranged from 109 to 370 contours (supplementary material, table S1). This step aimed to generate high-quality initial segmentations, referred to as pseudo-contours.
2. **Pseudo-Contour Generation**: the trained OAR-specific models were then used to generate pseudo-contours for the OARs that were not manually labeled in each CT scan. This step ensured that every CT image had a complete set of OAR segmentations.

3. **Final Segmentation Model Training**: a multi-label 3D full-resolution U-Net, referred to as the final CT segmentation model, was trained using the nnU-Net framework (Isensee *et al* 2021). This model was trained on a combination of ground truth and pseudo-contours, enabling simultaneous segmentation of all OARs. We employed a five-fold cross-validation approach, training the model on 415 CT images for 1250 epochs per fold with a combination of Dice and cross-entropy losses and the Adam optimizer with an initial learning rate of 10^{-4}.

As described in Cubero *et al* (2022), the final CT segmentation model demonstrated higher accuracy compared to OAR-specific models, benefiting from muti-label training that leverages spatial relationships between organs. Therefore, the final CT segmentation model was used to contour OARs on the 230 CT images in the test dataset. These predictions were then evaluated against the ground truth contours using the DSC and average surface distance (ASD).

Finally, to create a fully labeled training set, the final CT segmentation model was also used to generate segmentations for the unlabeled OARs of each CT scan. The accuracy and robustness of these generated segmentations were further validated against available ground-truth labels, with detailed results provided in the supplementary material (table S1).

2.3. OAR segmentation on CBCT images

The second goal of this study was to develop a comprehensive framework for segmenting 25 OARs on CBCT images. This involved two main components: (1) generating sCT images from CBCT scans and (2) training a segmentation model using domain adaptation techniques to transfer knowledge from CT to sCT images. The overall workflow is illustrated in figure 1.

2.3.1. CBCT-to-sCT generation

CBCT-to-sCT translation was performed using a previously proposed method (Hémon *et al* 2023, 2025). This approach employed an unsupervised 2D conditional generative adversarial network comprising a six-block ResNet generator and a PatchGAN discriminator. The generator's architecture included batch normalization and activation functions, with LeakyReLU applied in the ResNet blocks and ReLU in other blocks, except for the final convolution, which had no activation.

The sCT generation problem was addressed through a style transfer approach. The output sCTs were reconstructed by applying the style of CT images to the content of CBCTs using a perceptual loss, specifically the Content and Style Representation by Enhanced Perceptual Synthesis loss, which included:

- Content Loss: Comparing feature maps of the CBCT and sCT images at various network layers.
- Style Loss: Comparing the style of the CBCT and sCT images using the 1-norm between their Gram matrices across multiple layers.

The model processed 2D patches of 168 $\times$ 168 pixels extracted from the CBCT images. During both training and testing, these patches were selected with a stride of half the patch size. Postprocessing involved computing the median of overlapping regions, reversing the preprocessing steps, and clipping pixel values to the range $[-1024, 3000]$.

Importantly, the sCT generation model from Hémon *et al* (2023, 2025) was used as it is, without additional retraining for this study. All 196 training and 166 test CBCT images were transformed into sCT using this pre-trained model. From this point forward, we will focus on segmenting these sCT images.

2.3.2. sCT segmentation

The sCT segmentation framework was trained with both sCT and CT images, allowing the model to leverage high-contrast information from CT while learning to segment the lower-quality sCT images. We trained a 3D full-resolution U-Net within nnU-Net on labeled CT scans (restricted to the CBCT field of view (FOV)) and sCT images with ground truth contours transferred from their paired CTs.

Three factors ensured anatomical and FOV consistency between CT and CBCT pairs: strict temporal proximity criteria, immobilization masks for patient positioning, and non-rigid registration. Among the 362 CT-CBCT pairs used for both training and test, 50% (182 pairs) were acquired within a maximum of two days, while 71% (258 pairs) were obtained within the same week (figure S1, supplementary material).

The complete methodology can be summarized in the following steps:

1. **sCT Cleaning**: the generated sCT images were cleaned to remove any noise outside the patient's body contour, predicted using TotalSegmentator v2 (Wasserthal *et al* 2023). Voxels outside the body contour were set to -1024 Hounsfield units, corresponding to air intensity values.

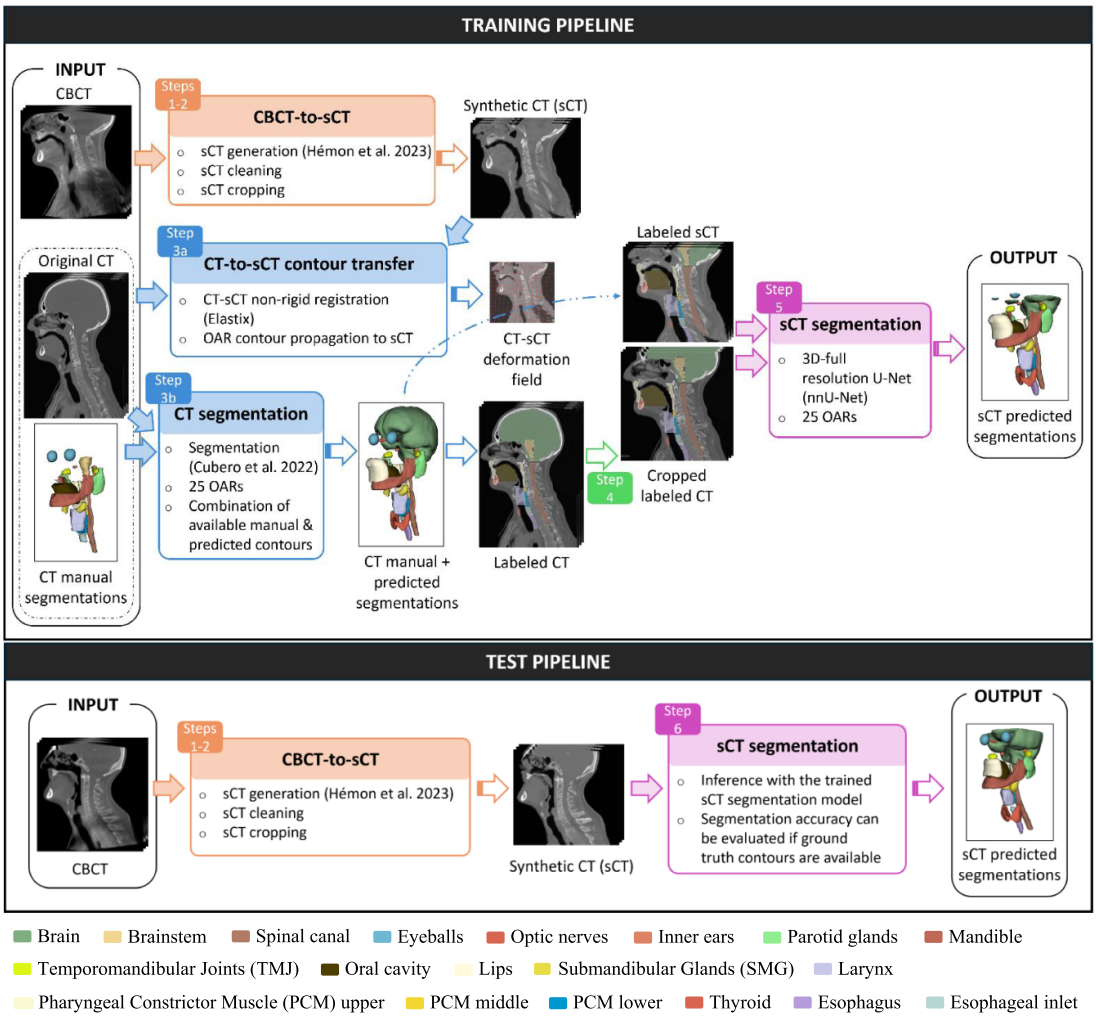

Figure 1. Proposed framework for sCT segmentation. Training Pipeline: Steps 1-2: the CBCT image is transformed into a synthetic CT (sCT), cleaned to remove any noise outside the patient's body, and cropped. Step 3: the CT and sCT images are non-rigidly registered, and the CT contours warped to the sCT. Step 4: the labeled CT image is cropped to a reduced FOV. Step 5: the labeled sCT and labeled cropped CT images are used to train the sCT segmentation network. Test Pipeline: Steps 1-2: the CBCT image is transformed into a sCT, cleaned to remove any noise outside the patient's body, and cropped. Step 6: the trained sCT segmentation network predicts the OARs on the sCT.

2. **sCT Preprocessing**: to address axial truncation artifacts introduced during CBCT acquisition, the sCT images were cropped by two slices superiorly and inferiorly, reducing the total number of slices to 128. Moreover, the sCT images were resampled to 1.156 mm in the x and y axes to align with CT resolution and cropped to 256 voxels in each dimension. Their intensity values were also normalized to match the CT intensity range.
3. **CT-to-sCT Contours Transfer**: non-rigid registration was performed to align sCT images to their paired CT using the Elastix registration toolkit (Klein *et al* 2010). We applied a Recursive B-Spline Transform with a spline order of 3, using Adaptive Stochastic Gradient Descent as the optimizer with up to 2000 iterations and automatic parameter estimation. The cost function used was Advanced Normalized Correlation with a weight of 0.9, combined with a Transform Bending Energy Penalty weighted by 0.1. The transformation was interpolated using a B-Spline Interpolator with an order of 3. The resulting deformation field was used to transfer OAR contours from CT to their respective sCT images using nearest-neighbor interpolation to preserve the discrete segmentation labels. These propagated contours were used as ground truth annotations for the sCT images. A visual evaluation was performed on a subset of 20 cases with time differences exceeding one week to validate the accuracy of the registration. Five representative examples of these cases are provided in figure S2 (supplementary material).
4. **CT Cropping**: to ensure that CT images matched the CBCT FOV, they were cropped to 128 slices. The cropping ensured the inclusion of the complete eyeball contours, often excluded in sCT images due to the limitations of the CBCT FOV. This guaranteed that the model would learn to segment the eyeballs and optic nerves accurately.

5. **Model Training**: a 3D full-resolution U-Net was trained within nnU-Net (Isensee *et al* 2021) with five-fold cross-validation over 2500 epochs, using the Dice Cross-Entropy loss function and Adam optimizer with an initial learning rate of 10^{-4}.
6. **Segmentation Evaluation**: segmentation accuracy was assessed with the DSC and ASD on the 166 test CBCT images by calculating these metrics between the predicted and registered CT contours on the corresponding sCT images.

2.4. Dosimetric evaluation of sCT contours

The third objective of this study was to evaluate the clinical viability of automatic OAR segmentations on sCT images for dose comparison and treatment replanning—an essential yet often overlooked aspect in CBCT segmentation studies. For this purpose, the dose received by each OAR was compared between the contours delineated on CT and sCT, using the same dose distribution calculated solely on the original CT. This analysis was performed to validate the accuracy of CBCT-based segmentations for offline ART applications and retrospective dose accumulation, not for online ART.

The dosimetric evaluation was performed on a subset of 28 patients from the ARTIX cohort, selecting the first paired CT-CBCT scan for each patient. All patients were treated following the ARTIX protocol (Castelli *et al* 2023), receiving a total dose of 70 Gy in 35 fractions (2 Gy/fraction), with dose constraints based on GORTEC guidelines (Tao *et al* 2020).

To ensure a reliable comparison, each sCT was registered to its corresponding CT using Elastix (as described earlier), and the CT was cropped to match the sCT's FOV. DL-generated segmentations were used for both CT and sCT to avoid smoothing inconsistencies between manual and DL contours.

The dosimetric evaluation included the following steps:

1. **Dose Calculation on CT:** the dose distribution used in this study was the clinically approved treatment dose calculated on the CT images in the Treatment Planning System Pinnacle (Philips). Dose volume histograms (DVH) for each OAR were then computed separately for the CT and sCT contours based on the same dose distribution.
2. **Dosimetric Metrics:** for each OAR segmentation in CT and sCT, the mean, maximum doses and D2 (%) were calculated from the DVH, as these metrics are frequently used for dose optimization in HNC RT planning. Additionally, the volume of each OAR was also measured.
3. **Statistical Analysis:** a paired Wilcoxon signed-rank test was performed to compare the mean dose, D2 (%), and volume for each OAR between the CT and sCT segmentations. A p-value < 0.05 was considered statistically significant.

3. Results

3.1. OAR segmentation on CT and sCT images

Figure 2 summarizes the segmentation performance of 25 OARs on CT and sCT images, with detailed metrics available in table S2 (supplementary material). Statistical significance was assessed using the two-sided Mann-Whitney U test.

The CT segmentation framework exhibited high accuracy across most OARs, with DSC values generally exceeding 70% and ASDs under 2 mm. The brain achieved the highest DSC (98.1%) and smallest ASD (0.51 mm), while elongated structures like the spinal canal had a high DSC of 84.2% but a slightly higher ASD of 1.69 mm. Smaller structures, such as the optic nerves and lips, showed lower DSC values (66.2% and 60.1%, respectively). The PCM also presented low DSCs (56.7%, 48.2%, and 59.6% for upper, middle, and lower segments, respectively), though their ASDs remained small (1.78, 1.83, and 1.23 mm).

For sCT segmentation, the model maintained overall accuracy, although with slight reductions compared to CT. For instance, the larynx, and upper and middle PCM, showed ASDs exceeding 2 mm. Additionally, OARs often affected by the limited FOV of the DSC of the brain decreased to 94.4% ($p < 0.001$), with a marginally higher ASD of 0.61 mm ($p < 0.01$). Several OARs, including the oral cavity, sCT images, such as the eyeballs and optic nerves, presented notable drops in DSC (eyeballs decreased from 86.2% to 71.3% ($p < 0.001$), and optic nerves from 66.2% to 53.5% ($p < 0.001$)). In contrast, the esophagus, also frequently cropped in sCT, showed improved segmentation accuracy (DSC increased from 67.9% to 72.2% ($p < 0.01$)).

Figure 3 depicts two patient cases, each showcasing paired CT and CBCT images with their respective manual and predicted segmentations. Patient 1 demonstrates high segmentation performance across all OARs. In contrast, Patient 2 illustrates a more challenging scenario due to his slight forward inclination resulting in certain OARs being partially outside the FOV, the presence of truncation artifacts in the sCT (visible in the third row), and a larger tumor extending into the nasopharyngeal region. While our training

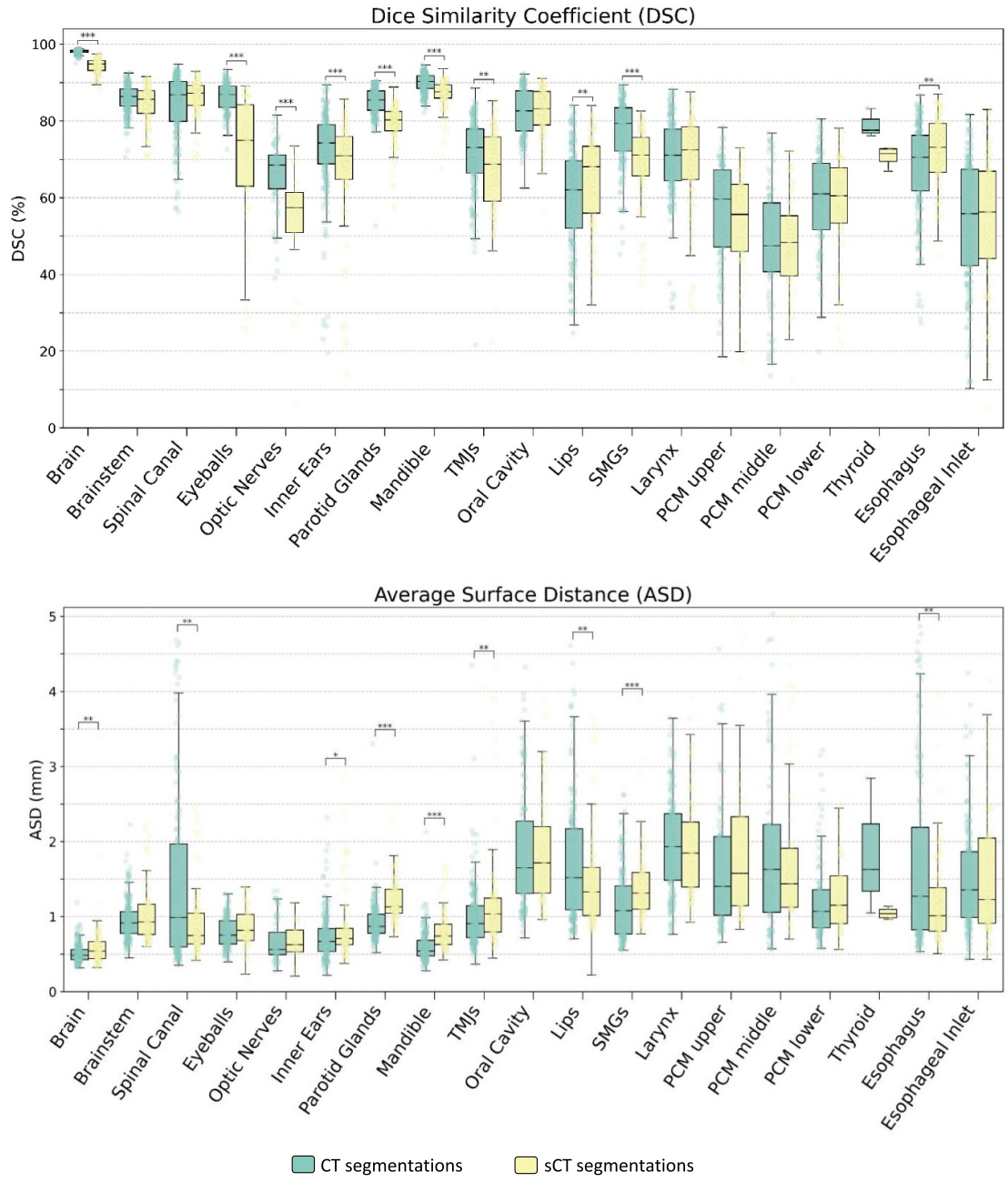

Figure 2. OAR segmentation evaluation. The dice score coefficient (DSC) (%) and average surface distance (ASD) (mm) are depicted for each OAR for CT (green) and sCT (yellow) segmentation models in the test set. The number of test samples for each OAR varies, with a median of 221 CT segmentations (from 230 available CTs) and 164 sCT segmentations (from 166 available sCTs). TMJs: Temporomandibular Joints; SMGs: Submandibular Glands; PCM: Pharyngeal Constrictor Muscle. Statistical significance is indicated as: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***).

database consists of advanced HN tumors, this especially large tumor may modify the shape and size of adjacent OARs, complicating segmentation and making the case more distinct from the training data.

3.2. Dosimetric evaluation of sCT segmentations

Table 2 compares dosimetric values—mean, maximum dose and D2 (%)—between automatic OAR segmentations on paired CT and sCT images in the ARTIX cohort.

Most OARs showed minimal differences in mean, maximum doses and D2, with no significant differences ($p > 0.05$). A notable exception is the spinal canal, where the mean, maximum doses and D2 significantly increased from 29 ± 2 Gy on CT to 30 ± 6 Gy on sCT ($p = 0.002$), from 43 ± 2 Gy to 44 ± 2 Gy ($p = 0.047$), and from 36.2 ± 2.1% to 37.6 ± 6.1% ($p = 0.045$), respectively. The brainstem and lips also presented significant differences in D2, increasing in the brainstem from 30.1 ± 6.9% on CT to 31.3 ± 7.6% on sCT and decreasing in the lips from 32.2 ± 10.0% to 30.4 ± 11.6%, alongside a mean dose difference of

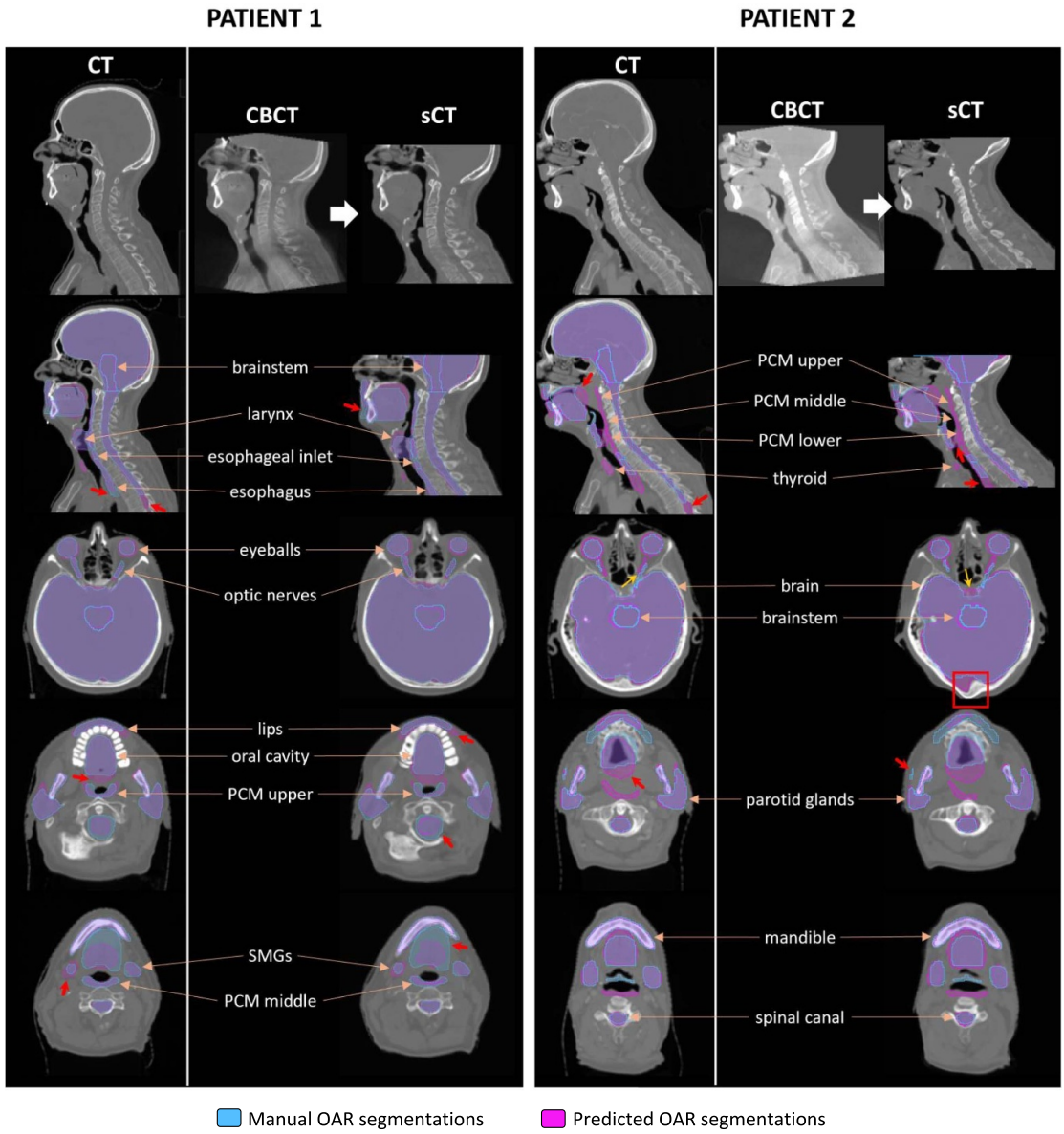

Figure 3. Visualization of manual (blue) and predicted (pink) OAR segmentations for two patients. Top row: CT, CBCT, and generated synthetic CT (sCT) images. Below: midsagittal and three different axial slices with manual and predicted segmentations on CT and sCT, enabling direct comparison. Violet indicates perfect contour overlap. Red arrows highlight regions of disagreement between segmentations, while the red square indicates an image artifact in CBCT. Importantly, manual segmentations are not available for some OARs, such as the upper, middle, and lower PCM in Patient 2; therefore, only predicted segmentations are shown for these structures. Moreover, not all the OARs present in each slice are pointed out to ease visualization. SMGs: Submandibular Glands; PCM: Pharyngeal Constrictor Muscle.

1 Gy in both OARs. The brainstem also demonstrated a significant increase in the maximum dose from 35 ± 10 Gy on CT to 38 ± 11 Gy on sCT. Additionally, the mandible showed a significant reduction in mean dose on sCT (42 ± 8 Gy vs. 40 ± 9 Gy, $p < 0.001$).

OAR volume differences (table S3, supplementary material) between CT and sCT segmentations were generally non-significant. Notable exceptions included reductions in the upper PCM and contralateral parotid gland volumes on sCT (12.4 ± 2.0 cc vs. 11.4 ± 2.1 cc for the upper PCM, $p = 0.015$; and 25.3 ± 8.6 cc vs. 22.8 ± 8.3 cc for the parotid gland, $p = 0.009$), although these changes did not impact the mean, maximum doses or D2. Conversely, the brainstem exhibited a significant increase in volume from CT to sCT (22.2 ± 4.5 cc vs. 24.0 ± 6.1 cc, $p = 0.010$).

Table 2. Dosimetric comparison between paired automatic CT and sCT OAR segmentations in the ARTIX cohort. OARs are listed in approximate order by mean dose, with higher-dose OARs generally at the top and lower-dose OARs toward the bottom. TMJ: Temporomandibular Joints; SMG: Submandibular Glands; PCM: Pharyngeal Constrictor Muscle; ipsi.: ipsilateral; contra.: contralateral.

	Mean dose (Gy)			Maximum dose (Gy)			D2 (%)		
OAR	CT	sCT	p-value	CT	sCT	*p*-value	CT	sCT	*p*-value
SMG ipsi.	66 ± 2	66 ± 3	0.76	72 ± 1	72 ± 1	0.58	66.8 ± 2.2	66.5 ± 2.3	0.70
SMG contra.	59 ± 7	57 ± 8	0.96	67 ± 4	67 ± 5	0.86	61.2 ± 4.9	60.3 ± 5.9	0.68
PCM upper	64 ± 4	64 ± 4	0.53	72 ± 2	71 ± 3	0.75	66.2 ± 3.3	66.1 ± 3.3	0.40
PCM middle	64 ± 4	63 ± 5	0.17	69 ± 3	69 ± 3	0.78	65.0 ± 3.7	64.8 ± 3.7	0.21
PCM lower	57 ± 9	57 ± 9	0.82	67 ± 3	67 ± 3	0.31	59.0 ± 7.1	59.1 ± 7.2	0.52
Larynx	54 ± 10	54 ± 10	0.08	68 ± 4	69 ± 4	**0.004**	58.7 ± 8.1	59.3 ± 7.8	**0.010**
Oral cavity	51 ± 10	50 ± 11	0.38	71 ± 5	70 ± 6	0.99	63.1 ± 6.5	62.6 ± 6.8	0.63
Esophageal inlet	50 ± 10	51 ± 9	0.98	58 ± 6	59 ± 5	0.55	51.7 ± 8.7	51.8 ± 8.4	0.71
Thyroid	50 ± 9	50 ± 9	0.33	66 ± 2	66 ± 3	0.53	58.5 ± 3.3	58.2 ± 3.5	0.10
Mandible	42 ± 8	40 ± 9	**<0.001**	69 ± 5	69 ± 5	0.69	57.1 ± 6.6	56.5 ± 7.2	0.57
Spinal canal	29 ± 2	30 ± 6	**0.002**	43 ± 2	44 ± 2	**0.047**	36.2 ± 2.1	37.6 ± 6.1	**0.045**
Esophagus	28 ± 10	28 ± 10	0.37	55 ± 6	54 ± 7	0.08	44.7 ± 9.2	44.3 ± 9.6	0.38
Parotid ipsi.	26 ± 9	26 ± 10	0.63	67 ± 4	67 ± 5	0.97	55.4 ± 5.1	54.6 ± 6.5	0.55
Parotid contra.	19 ± 5	18 ± 5	0.43	63 ± 6	62 ± 7	0.38	48.9 ± 5.8	47.4 ± 7.1	0.09
TMJ ipsi.	25 ± 12	25 ± 12	0.63	48 ± 22	46 ± 22	0.08	39.5 ± 16.0	39.2 ± 16.1	0.15
TMJ contra.	16 ± 9	15 ± 9	0.78	32 ± 22	31 ± 22	0.70	25.7 ± 16.8	25.1 ± 16.8	0.58
Lips	22 ± 6	21 ± 7	**0.012**	43 ± 13	42 ± 13	0.06	32.2 ± 10.0	30.4 ± 11.6	**0.044**
Inner ear ipsi.	19 ± 10	20 ± 13	0.40	30 ± 16	30 ± 17	0.91	27.0 ± 12.6	27.4 ± 13.9	0.43
Inner ear contra.	13 ± 9	13 ± 9	0.51	21 ± 15	19 ± 14	0.65	19.9 ± 12.7	18.9 ± 12.5	0.41
Brainstem	13 ± 6	14 ± 6	**0.038**	35 ± 10	38 ± 11	**0.016**	30.1 ± 6.9	31.3 ± 7.6	**0.001**
Brain	9 ± 5	9 ± 5	**0.011**	53 ± 12	54 ± 13	0.97	33.7 ± 8.3	34.3 ± 8.5	0.15
Eyeball ipsi.	3 ± 2	3 ± 2	0.17	5 ± 3	6 ± 3	**0.036**	4.9 ± 3.6	5.1 ± 3.6	0.14
Eyeball contra.	2 ± 1	2 ± 1	**0.008**	5 ± 2	5 ± 3	**0.028**	3.7 ± 2.2	4.0 ± 2.3	**0.007**
Optic nerve ipsi.	3 ± 1	3 ± 1	0.42	3 ± 1	3 ± 1	0.49	3.6 ± 1.3	3.9 ± 2.0	0.21
Optic nerve contra.	2 ± 1	2 ± 1	0.67	3 ± 1	3 ± 1	0.56	3.2 ± 1.4	3.2 ± 1.4	0.92

4. Discussion

This study presents a comprehensive approach for segmenting 25 OARs on head and neck CBCT images, aiming to enhance RT treatment replanning. The methodology was developed in three key steps: refining an in-house framework to segment 25 OARs on CT; adapting it to segment the same OARs on sCT images generated from CBCT; and validating the accuracy and clinical relevance of sCT segmentations through a dosimetric analysis.

The first contribution of this study is the segmentation of 25 OARs on CT images, offering a broader anatomical scope than the existing literature. The geometric evaluation demonstrated high segmentation accuracy, with DSCs generally exceeding 70% and ASDs below 2 mm. For instance, the parotid glands, crucial in HNC RT, achieved a DSC of 85.1 $\pm$ 3.8%, outperforming existing methods focused solely on these glands ((Kakkos *et al* 2024), 82.7 $\pm$ 1.0%). Smaller structures like the optic nerves and lips showed lower DSCs (66.2% and 60.1%, respectively), reflecting the challenges of segmenting small volumes and the limitations of the DSC as an evaluation metric (Taha and Hanbury 2015).

We also compared our segmentation performance to the DAHANCA study (Nielsen *et al* 2024), which assessed interobserver variability across 15 expert radiation oncologists contouring 17 HN OARs from 26 patients with similar HNC profiles to ours. Our model exceeded expert variability for some OARs, like the brain (table S4, supplementary material). While the proposed CT segmentation framework achieved lower DSCs for smaller OARs like the lips and upper, middle, and lower PCM (60%, 57%, 48%, and 60%, respectively), these values remain comparable to or better than expert interobserver variability, which often falls below 56% for these structures. This highlights the robustness of the model, which aligns with or surpasses expert-level accuracy.

Despite being trained with partially-labeled data obtained from the clinical treatment plans of over 200 HNC patients, the CT segmentation framework yielded robust results. Our previous work (Cubero *et al* 2022) demonstrated that partially-labeled data from clinical pipelines can still provide accurate segmentations, and studies suggest that increasing the number of patients beyond 200 does not improve segmentation performance for HN OARs (Henderson *et al* 2023), supporting the validity of our approach.

The second contribution of this study is developing a framework to segment 25 OARs on sCT images generated from CBCT. As expected, segmentation accuracy on sCT was slightly lower than on CT due to the inherent lower quality and variability of CBCT and, consequently, sCT images (Iliopoulos *et al* 2023). Nevertheless, key OARs like the brain and spinal canal maintained high segmentation performance, with the brain's DSC decreasing from 98.1% on CT to 94.4% ($p < 0.001$) on sCT. OARs often cropped in CBCT due to its limited FOV, such as the eyeballs and optic nerves, experienced more pronounced reductions in accuracy, with DSCs dropping from 86.2% to 70.8% ($p < 0.001$) and 66.2% to 52.9% ($p < 0.001$), respectively. These decreases are attributed to truncation artifacts in sCT, where empty regions in the upper slices are not segmented by the model, in contrast to manually propagated contours. In contrast, the esophagus showed improved performance, with the DSC increasing from 67.9% on CT to 70.5% ($p < 0.01$) on sCT. This improvement likely results from the reduced variability in the lower esophagus' segmentation boundaries, which are fully visible and consistently labeled within the smaller FOV of CBCT, which minimizes segmentation errors typically encountered in CT. Finally, the thyroid gland presented a higher DSC on CT but a smaller ASD on sCT, likely due to the reduced contrast in CBCT, which resulted in smaller segmented volumes.

Our approach demonstrates several advantages over the few published studies on the topic. Notably, it offers comprehensive segmentation of 25 OARs, surpassing the narrower scope of previous works, which included only seven to 19 OARs (Dai *et al* 2021a, Ma *et al* 2022, Liang *et al* 2023, Zhao *et al* 2024). Segmenting a broader set of OARs is crucial for minimizing radiation exposure to critical structures during treatment planning. Furthermore, this study is among the first to train directly with both CT and sCT images for OAR segmentation, while most studies have relied on manually delineated datasets or CBCT enhancement techniques (Chen *et al* 2021, Liu *et al* 2023). We have compared our model trained on both CT and sCT images to models trained on CT or sCT alone. The results, presented in table S5 (supplementary material), demonstrate the superior segmentation performance of the proposed model for all OARs, especially soft tissue organs. The DSC superiority with the proposed model was statistically significant ($p < 0.05$) for 16 out of 19 OARs when compared to the CT-only model and for 10 out of 19 OARs when compared to the sCT-only model. This aligns with recent findings (Radici *et al* 2024, Zhao *et al* 2024) that using CT and sCT as image inputs enhances segmentation accuracy.

Regarding the geometric evaluation, our results exhibited some variation compared to existing studies (table S6, supplementary material). DSCs for the brainstem (85%), spinal canal (86%), and mandible (87%) aligned closely with those reported by Dai *et al* (2021a) (87%, 84%, and 86%) and Liang *et al* (2023) (91%, 89%, and 91%). In contrast, DSCs for the parotid glands (79%) and SMGs (69%) were notably lower than

those reported by Dai *et al* (2021a) (88% for parotids), Ma *et al* (2022) (88% and 79%), Liang *et al* (2023) (90% and 85%), and Zhao *et al* (2024) (94% and 86%).

It is important to acknowledge that these comparisons are influenced by factors such as differences in manual segmentation protocols and the smaller test datasets used in these studies: 30 CBCT images (Dai *et al* 2021a), eight (Ma *et al* 2022), nine (Liang *et al* 2023), and ten (Zhao *et al* 2024). In contrast, this study analyzed 166 CBCT scans. Additionally, training datasets and methodologies differ: (Dai *et al* 2021a) trained on MR-delineated contours, potentially benefiting from MR's superior soft tissue contrast, while (Zhao *et al* 2024) trained on a simulated database and tested on sCTs generated from pCTs rather than actual CBCTs, which could explain their superior results due to the higher quality of CT-based synthetic images.

The third major contribution of this study is the dosimetric analysis of OAR segmentations on sCT images, designed to assess the impact of the observed geometric discrepancies in the dose distribution and clinical viability. Automatic segmentations on paired CT and CBCT images from 28 patients in the ARTIX cohort were compared, with scans acquired within the same week of treatment, using the same dose distribution calculated on CT.

Most OARs showed non-significant differences in the mean, maximum doses and D2. Notably, the spinal canal's mean, maximum doses and D2 increased significantly from 29 $\pm$ 2 Gy on CT to 30 $\pm$ 6 Gy on sCT, from 43 $\pm$ 2 Gy to 44 $\pm$ 2 Gy and from 36.2 $\pm$ 2.1% to 37.6 $\pm$ 6.1%, respectively. The brainstem and lips also exhibited significant differences in D2, with increases in the brainstem from 30.1 $\pm$ 6.9% on CT to 31.3 $\pm$ 7.6% on sCT and decreases in the lips from 32.2 $\pm$ 10.0% to 30.4 $\pm$ 11.6%, alongside a mean dose difference of 1 Gy in both OARs. Importantly, the brainstem also experienced a significant increase in the maximum dose, from 35 $\pm$ 10 Gy on CT to 38 $\pm$ 11 Gy on sCT. Additionally, the mandible showed a significant reduction in mean dose on sCT (42 $\pm$ 8 Gy vs. 40 $\pm$ 9 Gy). Volume differences were generally non-significant, except for a few exceptions. Two OARs had significant reductions in volume that did not impact the mean dose or D2, which were the upper PCM (12.4 $\pm$ 2.0 cc on CT vs. 11.4 $\pm$ 2.1 cc on sCT) and contralateral parotid gland (25.3 $\pm$ 8.6 cc vs. 22.8 $\pm$ 8.3 cc). The only OAR presenting both significant dosimetric and volumetric differences was the brainstem, with a volume increase from 22.2 $\pm$ 4.5 cc on CT to 24.0 $\pm$ 6.1 cc on sCT.

These findings indicate promising results for clinical application. Among the 25 OARs, only two exhibited mean dose differences exceeding 1 Gy, a variation deemed clinically acceptable given that most HN OARs have mean dose recommendations above 20 Gy (Bentzen *et al* 2010). More specifically, these OARs are the mandible, with a mean dose on CT of 42 Gy, and the contralateral SMG, with a mean dose of 59 Gy. Furthermore, even OARs with lower DSCs on sCT, like the lips (64.7%) and the upper, middle, and lower PCM (56.7%, 48.2%, and 59.6%), demonstrated favorable dosimetric outcomes. This suggests that, for most OARs, geometric inaccuracies in segmentation had a relatively minor impact on the dose.

Furthermore, minor segmentation differences in high-dose gradient regions can lead to significant dosimetric discrepancies, particularly in patients with large oropharyngeal tumors deforming the surrounding anatomy and complicating soft tissue segmentation. For instance, as illustrated in figure 4, Patient 2 displays notable geometric variations, yet the brain's minimal dose exposure means these discrepancies have negligible effects. Conversely, the oral cavity, situated in a high-dose area, is more susceptible to these variations. Overall, the results support the clinical viability of the proposed framework for OAR segmentation on sCT images.

To the best of our knowledge, this study is the first to validate dosimetrically HN OAR segmentations between CT and sCT generated from CBCT images. Unlike previous studies—on other anatomical regions—that recalculated the dose on sCT (Dai *et al* 2021b, Jiang *et al* 2024), this study used the same dose distribution calculated on CT, as our final aim is to validate CBCT segmentation for offline replanning and toxicity studies on CT. Our results underscore the feasibility of using CBCT and sCT for accurately monitoring anatomical changes—such as weight loss or gland and tumor shrinkage—during RT delivery, triggering offline ART based on CBCT segmentations, and dose accumulation for toxicity evaluation on the pCT. A key strength of our study is the ability to segment CBCT images where no CT is available or when CT and CBCT acquisitions differ too significantly, addressing a critical gap in clinical practice.

While the primary focus of this study is offline ART, the proposed approach also holds significant potential for online ART with straightforward replanning on sCT, as explored in other studies (Chang *et al* 2024) and commercial solutions (Ethos with integrated HyperSight technology, Varian). Even though MR-guided RT is gaining traction, its widespread implementation remains limited by complexity, cost, and need for specialized resources (McGee *et al* 2024, Ocanto *et al* 2024). Until these challenges are overcome and the clinical advantages of MR-guided RT are fully proven, IMRT will likely continue to rely primarily on CT imaging for treatment planning and CBCT imaging for fraction monitoring. Therefore, ensuring accurate OAR segmentation on CBCT is essential for advancing in traditional CT-based IMRT and pushing the development and implementation of both offline and online ART (Dona Lemus *et al* 2024).

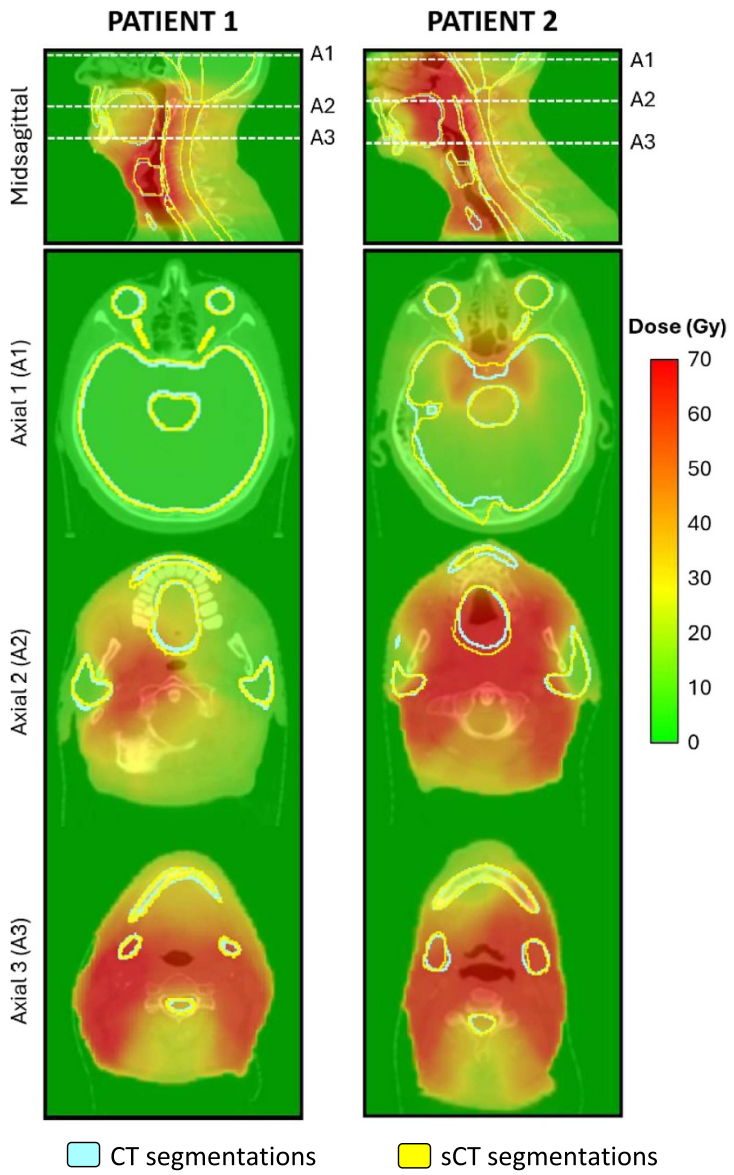

Figure 4. Visualization of dose distribution in two patients. The first row displays the midsagittal slice, followed by three axial slices for each patient, showing automatic CT (blue) and synthetic CT (sCT) segmentations (yellow) overlaid on the CT image and dose distribution. While Patient 1's dose distribution remains concentrated in the larynx, Patient 2 shows a broader high-dose region along with more noticeable segmentation differences.

Despite the promising results, this study has several limitations. The decrease in segmentation accuracy for certain OARs on sCT, particularly small or cropped structures, highlights the need for further refinement of our models, even if the dosimetric evaluation remains favorable. Moreover, the use of propagated CT contours as ground truth introduces inherent uncertainties, though a visual assessment (figure S2) confirmed good anatomical alignment after registration. Importantly, this study is currently limited to head and neck CBCT images, where relatively low intra-fraction variability facilitates the propagation of contours and the FOV primarily affects the superior-inferior direction. For other anatomical regions with higher intra-fraction variability, such as the abdomen or pelvis, our method may not be directly applicable. Additionally, lateral FOV limitations may pose further challenges, requiring careful consideration of the methodology to address differences in FOV and structure coverage. Furthermore, while we relied on a previously validated sCT generation model (Hémon *et al* 2023, 2025), further validation of sCT quality and potential hallucination effect could enhance confidence in segmentation performance and clinical utility.

In conclusion, this study demonstrates the feasibility and clinical relevance of using DL models for OAR segmentation on sCT images derived from CBCT for HNC RT. By achieving high segmentation accuracy and reliable dosimetric performance, our framework offers a promising pathway for improving ART, ultimately

enhancing treatment precision and reducing radiation exposure to healthy tissues. Future research should focus on validating these techniques in clinical adaptive workflows.

Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information.

Acknowledgment

Research supported by Projects AC20/00102 (Ministerio de Ciencia, Innovación y Universidades, Instituto de Salud Carlos III, Asociación Española Contra el Cáncer and European Regional Development Fund 'Una manera de hacer Europa'), project PerPlanRT (under the frame of ERA PerMed), TED2021-129392B-I00, TED2021-132200B-I00, PID2023-149604OB-I00 (European Union 'Next Generation EU'/PRTR and MCIN/AEI/10.13039/501 100 011 033).

ORCID iDs

Lucía Cubero https://orcid.org/0000-0003-3134-1433
Cédric Hémon https://orcid.org/0009-0003-6669-5108

References

Abbani N *et al* 2022 Deep learning-based segmentation in prostate radiation therapy using Monte Carlo simulated cone-beam computed tomography *Med. Phys.* **49** 6930–44

Alfouzan A F 2021 Radiation therapy in head and neck cancer *Saudi Med. J.* **42** 247–54

Archambaul Y, Boylan C, Bullock D, Morgas T, Peltola J, Ruokokoski E, Genghi A, Haas B, Suhonen P and Thompson S 2020 Making on-line adaptive radiotherapy possible using artificial intelligence and machine learning for efficient daily re-planning *Med. Phys. Int. J.* **8** 77–86 (available at: http://mpijournal.org/pdf/2020-02/MPI-2020-02-p077.pdf)

Bentzen S M, Constine L S, Deasy J O, Eisbruch A, Jackson A, Marks L B, Ten Haken R K and Yorke E D 2010 Quantitative analyses of normal tissue effects in the clinic (QUANTEC): an introduction to the scientific issues *Int. J. Radiat. Oncol. Biol. Phys.* **76** S3–9

Bray F, Laversanne M, Sung H, Ferlay J, Siegel R L, Soerjomataram I and Jemal A 2024 Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries *Ca A Cancer J. Clin.* **74** 229–63

Brion E, Léger J, Barragán-Montero A M, Meert N, Lee J A and Macq B 2021 Domain adversarial networks and intensity-based data augmentation for male pelvic organ segmentation in cone beam CT *Comput. Biol. Med.* **131** 104269

Castelli J *et al* 2023 Weekly adaptive radiotherapy vs standard intensity-modulated radiotherapy for improving salivary function in patients with head and neck cancer. A phase 3 randomized clinical trial *JAMA Oncol.* **9** 1056–64

Chang Y, Liang Y, Wu H, Li L, Yang B, Jiang L, Ren Q and Pei X 2024 Adaptive assessment based on fractional CBCT images for cervical cancer *J. Appl. Clin. Med. Phys.* **25** e14462

Chen W, Li Y, Yuan N, Qi J, Dyer B A, Sensoy L, Benedict S H, Shang L, Rao S and Rong Y 2021 Clinical enhancement in AI-based post-processed fast-scan low-dose CBCT for head and neck adaptive radiotherapy *Front. Artif. Intell.* **3** 614384

Cubero L, Castelli J, Simon A, De Crevoisier R, Acosta O and Pascau J 2022 Deep learning-based segmentation of head and neck organs-at-risk with clinical partially labeled data *Entropy* **24** 1661

Dai X, Lei Y, Wang T, Dhabaan A H, McDonald M, Beitler J J, Curran W J, Zhou J, Liu T and Yang X 2021a Head-and-neck organs-at-risk auto-delineation using dual pyramid networks for CBCT-guided adaptive radiotherapy *Phys. Med. Biol.* **66** 045021

Dai Z *et al* 2021b Geometric and dosimetric evaluation of deep learning-based automatic delineation on CBCT-synthesized CT and planning CT for breast cancer adaptive radiotherapy: a multi-institutional study *Front. Oncol.* **11** 725507

Dona Lemus O M, Cao M, Cai B, Cummings M and Zheng D 2024 Adaptive radiotherapy: next-generation radiotherapy *Cancers* **16** 1206

Fu Y, Lei Y, Wang T, Tian S, Patel P, Jani A B, Curran W J, Liu T and Yang X 2020 Pelvic multi-organ segmentation on cone-beam CT for prostate adaptive radiotherapy *Med. Phys.* **47** 3415–22

Guberina M *et al* 2024 Prospects for online adaptive radiation therapy (ART) for head and neck cancer *Radiat. Oncol.* **19** 4

Hémon C, Boussot V, Texier B, Dillenseger J-L and Nunes J-C 2023 Guiding unsupervised CBCT-to-CT synthesis using content and style representation by an enhanced perceptual synthesis (CREPs) loss SynthRAD2023 Challenge *MICCAI 2023* (*Vancouver, Canada*) (available at: https://hal.science/hal-04272509/)

Hémon C, Cubero L, Boussot V, Martin R-A, Texier B, Castelli J, de Crevoisier R, Barateau A, Lafond C and Nunes J-C 2025 Modeling dose uncertainty in cone-beam computed tomography: predictive approach for deep learning-based synthetic computed tomography generation *Phys. Imaging Radiat. Oncol.* **33** 100704

Henderson E G A, van Herk M and Osorio E M V 2023 The Impact of training dataset size and ensemble inference strategies on head and neck auto-segmentation *2023 IEEE 20th Int. Symp. on Biomedical Imaging (ISBI)* (IEEE) pp 1–4 (available at: https://ieeexplore.ieee.org/document/10230826/)

Hirashima H, Nakamura M, Imanishi K, Nakao M and Mizowaki T 2023 Evaluation of generalization ability for deep learning-based auto-segmentation accuracy in limited field of view CBCT of male pelvic region *J. Appl. Clin. Med. Phys.* **24** e13912

Iliopoulos P, Simopoulou F, Simopoulos V, Kyrgias G and Theodorou K 2023 Review on cone beam computed tomography (CBCT) dose in patients undergoing image guided radiotherapy (IGRT) *Advances in Dosimetry and New Trends in Radiopharmaceuticals* (IntechOpen) (https://doi.org/10.5772/intechopen.1002683)

Isensee F *et al* 2021 nnU-Net: self-adapting framework for U-net-based medical image segmentation *Nat. Methods* **18** 203–11

Jiang J, Min Seo Choi C, Deasy J O, Rimner A, Thor M and Veeraraghavan H 2024 Artificial intelligence-based automated segmentation and radiotherapy dose mapping for thoracic normal tissues *Phys. Imaging Radiat. Oncol.* **29** 100542

Kakkos I, Vagenas T P, Zygogianni A and Matsopoulos G K 2024 Towards automation in radiotherapy planning: a deep learning approach for the delineation of parotid glands in head and neck cancer *Bioengineering* **11** 214

Kisling K, Keiper T D, Branco D, Kim G G, Moore K L and Ray X 2022 Clinical commissioning of an adaptive radiotherapy platform: results and recommendations *J. Appl. Clin. Med. Phys.* **23** e13801

Klein S, Staring M, Murphy K, Viergever M A and Pluim J 2010 elastix: a toolbox for intensity-based medical image registration *IEEE Trans. Med. Imaging* **29** 196–205

Kunnen B *et al* 2024 The added value of a new high-performance ring-gantry CBCT imaging system for prostate cancer patients *Radiother. Oncol.* **200** 110458

Liang X, Morgan H, Bai T, Dohopolski M, Nguyen D and Jiang S 2023 Deep learning based direct segmentation assisted by deformable image registration for cone-beam CT based auto-segmentation for adaptive radiotherapy *Phys. Med. Biol.* **68** 045012

Liu Y *et al* 2023 Efficient segmentation using domain adaptation for MRI-guided and CBCT-guided online adaptive radiotherapy *Radiother. Oncol.* **188** 109871

Ma L *et al* 2022 Registration-guided deep learning image segmentation for cone beam CT–based online adaptive radiotherapy *Med. Phys.* **49** 5304–16

MacDonald R L, Fallone C, Chytyk-Praznik K, Robar J and Cherpak A 2024 The feasibility of CT simulation-free adaptive radiation therapy *J. Appl. Clin. Med. Phys.* **25** e14438

McGee K P *et al* 2024 The use of magnetic resonance imaging in radiation therapy treatment simulation and planning *Magn. Reson. Imaging* **60** 1786–805

Nielsen C P *et al* 2024 Interobserver variation in organs at risk contouring in head and neck cancer according to the DAHANCA guidelines *Radiother. Oncol.* **197** 110337

Ocanto A *et al* 2024 MR-LINAC, a new partner in radiation oncology: current landscape *Cancers* **16** 270

Rabe M, Kurz C, Thummerer A and Landry G 2024 Artificial intelligence for treatment delivery: image-guided radiotherapy *Strahlenther Onkol* (available at: https://link.springer.com/10.1007/s00066-024-02277-9)

Radici L *et al* 2024 Clinical evaluation of a deep learning CBCT auto-segmentation software for prostate adaptive radiation therapy *Clin. Transl. Radiat. Oncol.* **47** 100796

Robar J L, Cherpak A, MacDonald R L, Yashayaeva A, McAloney D, McMaster N, Zhan K, Cwajna S, Patil N and Dahn H 2024 Novel technology allowing cone beam computed tomography in 6 seconds: a patient study of comparative image quality *Pract. Radiat. Oncol.* **14** 277–86

Schreier J, Genghi A, Laaksonen H, Morgas T and Haas B 2020 Clinical evaluation of a full-image deep segmentation algorithm for the male pelvis on cone-beam CT and CT *Radiother. Oncol.* **145** 1–6

Taha A A and Hanbury A 2015 Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool *BMC Med. Imaging* **15** 29

Tao Y *et al* 2020 Concurrent cisplatin and dose escalation with intensity-modulated radiotherapy (IMRT) versus conventional radiotherapy for locally advanced head and neck squamous cell carcinomas (HNSCC): GORTEC 2004-01 randomized phase III trial *Radiother. Oncol.* **150** 18–25

Wasserthal J *et al* 2023 TotalSegmentator: robust segmentation of 104 anatomic structures in CT images *Radiology* **5** e230024

Yashayaeva A, MacDonald R L and Cherpak A 2024 Body contour adaptation for weight-loss and bolus for head and neck radiotherapy on Ethos version 2.0 and hypersight: synthetic CT versus direct calculation *J. Appl. Clin. Med. Phys.* **n/a** e14587

Zhang C *et al* 2022 Automatic segmentation for plan-of-the-day selection in CBCT-guided adaptive radiation therapy of cervical cancer *Phys. Med. Biol.* **67** 245020

Zhao H, Liang X, Meng B, Dohopolski M, Choi B, Cai B, Lin M-H, Bai T, Nguyen D and Jiang S 2024 Progressive auto-segmentation for cone-beam computed tomography-based online adaptive radiotherapy *Phys. Imaging Radiat. Oncol.* **31** 100610