

1 2 9 0



UNIVERSIDADE  
**COIMBRA**

Catarina Camacho Caldeira

# DEVELOPMENT AND COMPARATIVE ASSESSMENT OF GENERAL AND ANATOMY-SPECIFIC MODELS FOR SYNTHETIC COMPUTED TOMOGRAPHY GENERATION

Dissertation submitted to the Faculty of Science and Technology of the University of Coimbra for the degree of Master in Biomedical Engineering with specialisation in Clinical Informatics and Bioinformatics, supervised by Dr. Nickolas Papanikolaou, Dr. José Guilherme de Almeida and Dr. João Miguel Castelhano.

September 2025



**50-60%** oncology patients undergo **radiotherapy**

## Computed Tomography (CT)

High spatial resolution

Electron density information  
(from Hounsfield Units)

Dose calculation

## Magnetic Resonance Imaging (MRI)

Superior soft tissue contrast

Facilitates the contouring of  
structures – Gross Tumour  
Volume (GTV) and  
Organs at Risk (OARs)

Reduces ambiguities in inter- and  
intra- observer outlines

?

?

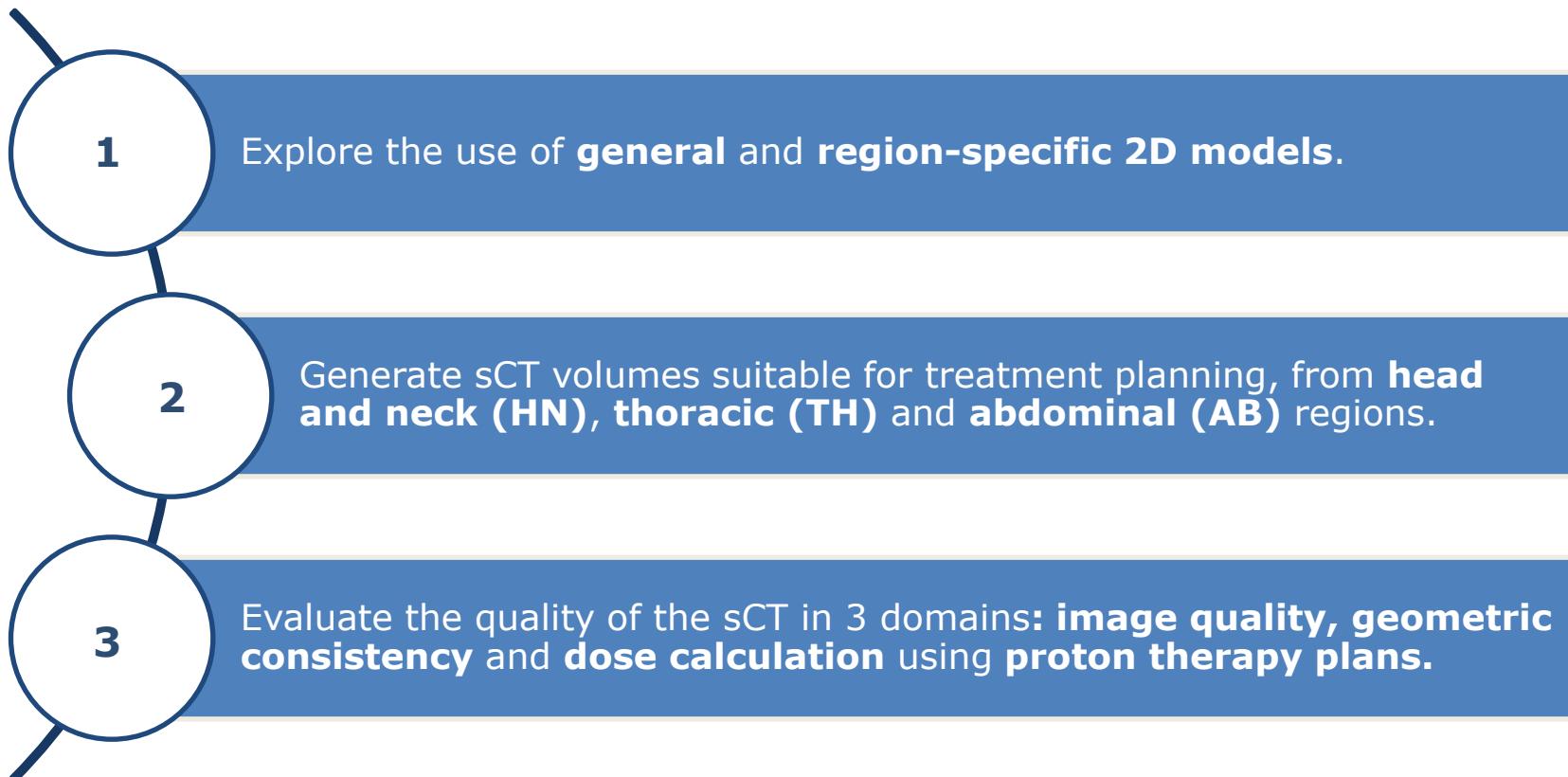
?

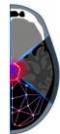
## Problems

- Co-registration complexities
  - Misalignment of MRI into CT coordinated system
  - Expert input to detect and correct potential errors
  - Differences in patient positioning
  - Variations in organ positioning (peristaltic activity) and intestinal gas
- Higher resources and staff demanding
- Patient disadvantages
  - Exposure to additional radiation (from CT)
  - Discomfort
  - Longer time in the hospital

## Solution

>Create a **machine learning model** to **generate** synthetic CT (**sCT**) from **MRI**.  
↳ MR-only treatment planning

- 
- 1 Explore the use of **general** and **region-specific 2D models**.
  - 2 Generate sCT volumes suitable for treatment planning, from **head and neck (HN)**, **thoracic (TH)** and **abdominal (AB)** regions.
  - 3 Evaluate the quality of the sCT in 3 domains: **image quality, geometric consistency** and **dose calculation** using **proton therapy plans**.



## SynthRAD2023

Synthesizing computed tomography for radiotherapy



**180 co-registered CT-MRI pairs**,  
along with dilated body outline masks

Single region: **Pelvis**

**Multi-centric:** Dutch university medical centres

**A** (*120 T1-weighted MRI*) and  
**C** (*60 T2-weighted MRI*)

No external validation



## SynthRAD2025

Synthesizing computed tomography for radiotherapy



**578 co-registered CT-MRI pairs**,  
along with dilated body outline masks

Multi-region: **Abdominal (AB)**, **Thoracic (TH)**, and **Head and Neck (HN)**

**Multi-centric:** European university medical centres

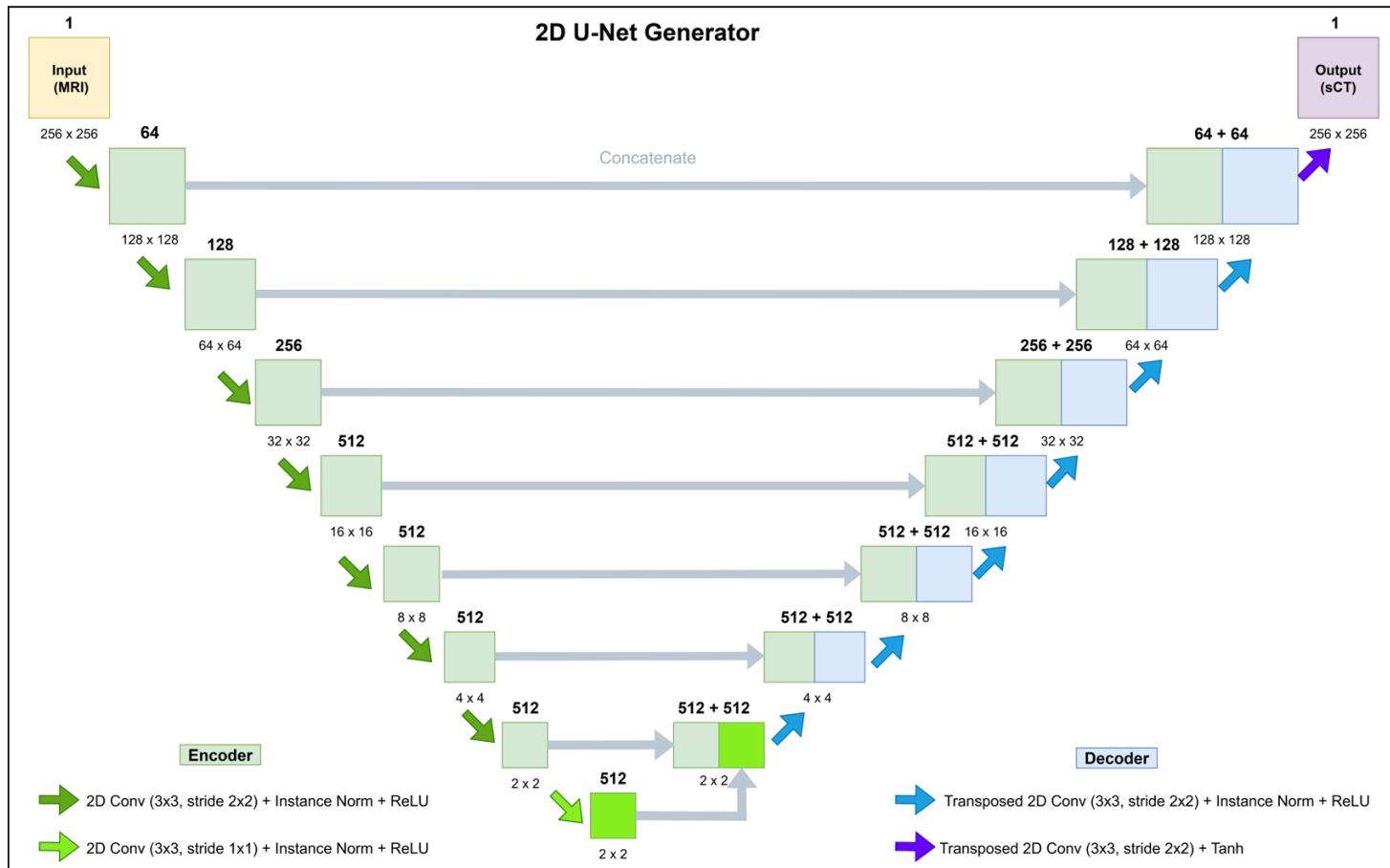
**A** (*247 T1-weighted MRI*),  
**B** (*182 T2/T1-weighted MRI*) and  
**C** (*AB 12 T2-weighted, AB 7 T2/T1-weighted,  
and HN 65 T1-weighted MRI with limited field  
of view*)

External validation: **HN**, centre **D** (*65 T1-weighted MRI*)

## Data augmentation processing:

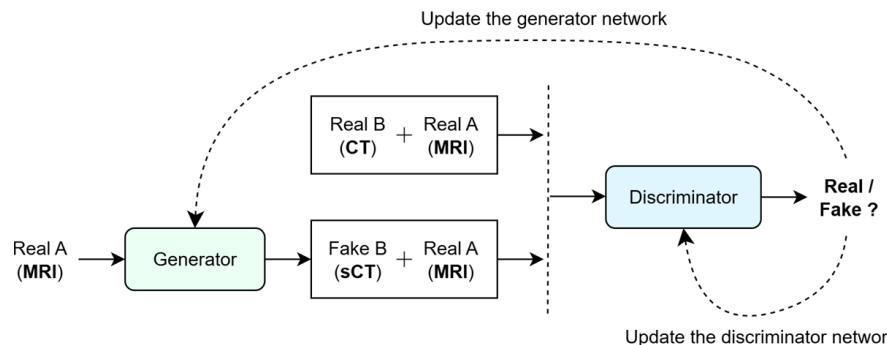
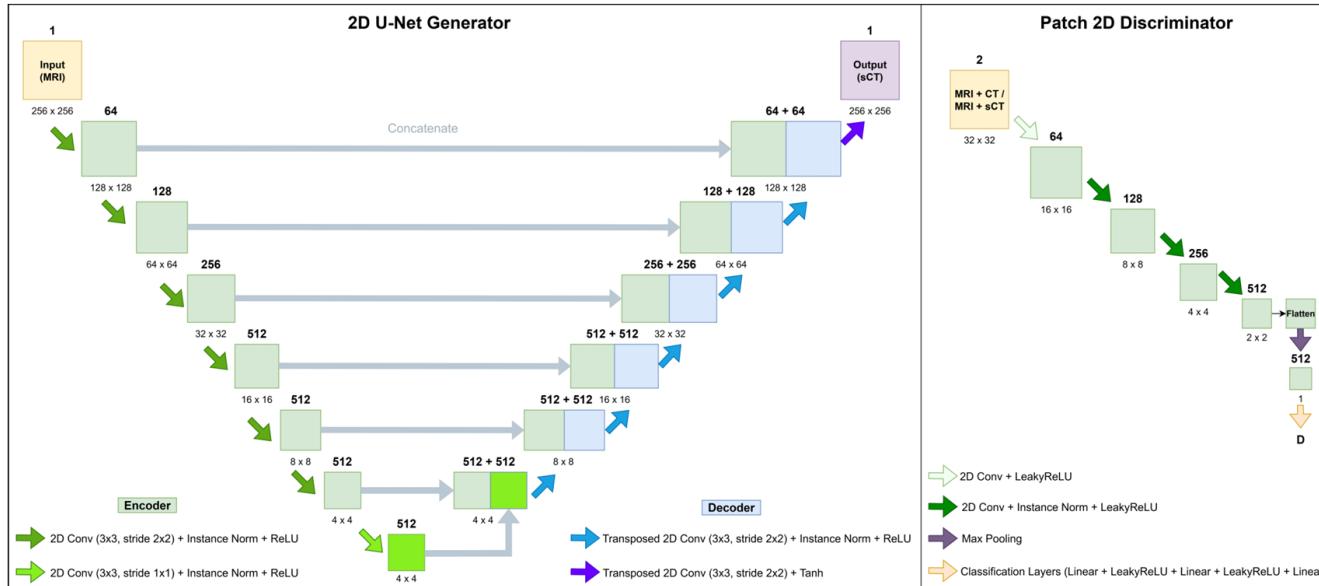
Spatial Transformations (MRI & CT) and Intensity Transformations (MRI)

## Autoencoder (AE)



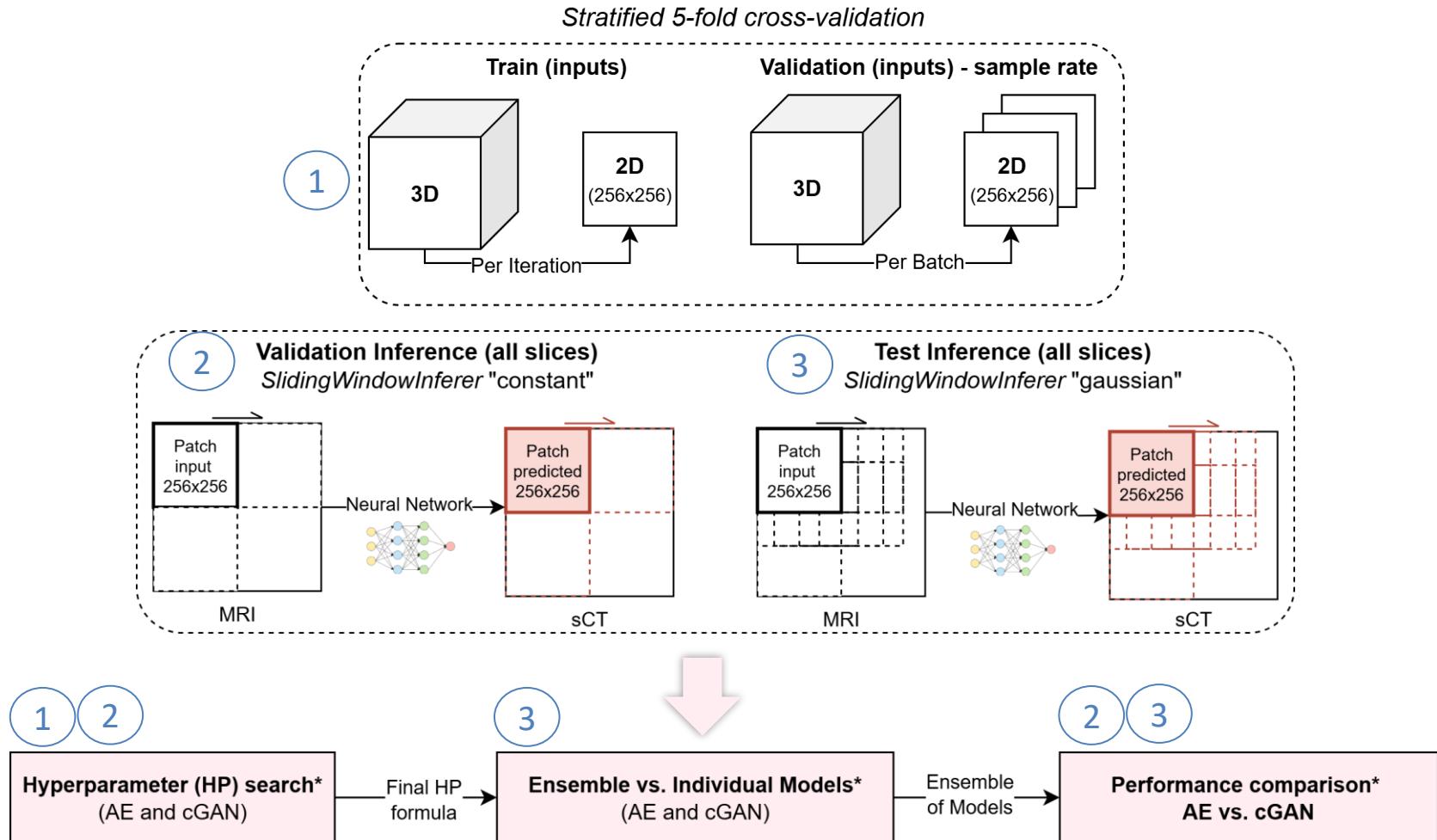
**Figure 1:** AE final architecture.

## Conditional Generative Adversarial Network (cGAN)



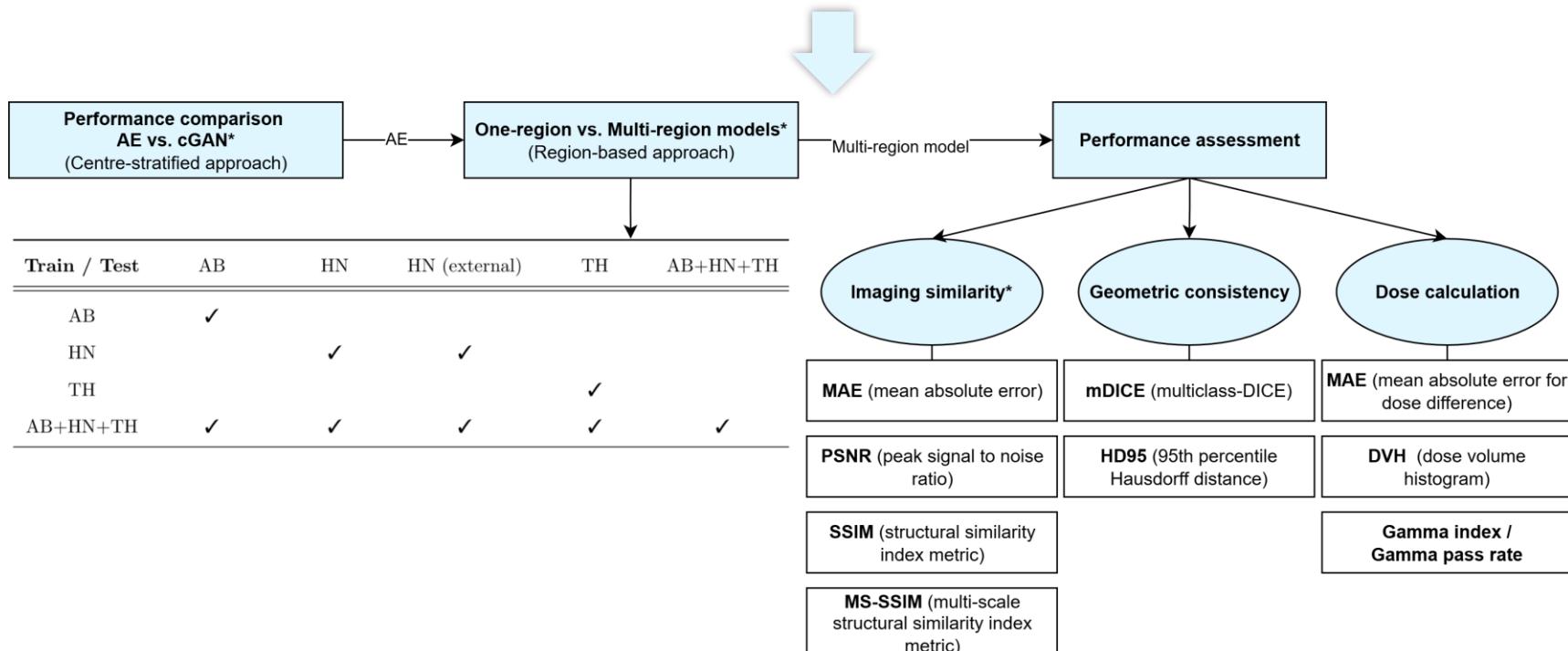
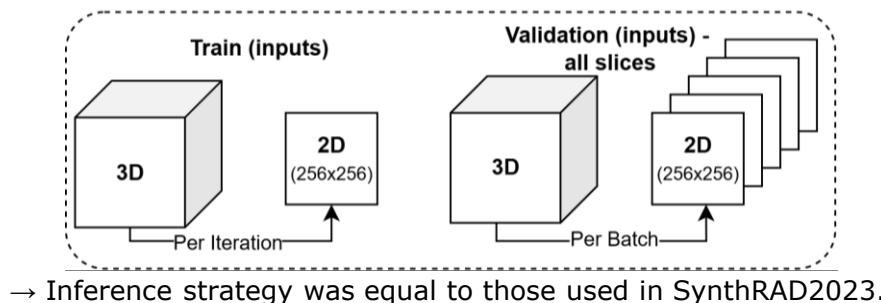
**Figure 2:** Simple cGAN training scheme in the case of MRI-to-CT synthesis task (inferior) and cGAN final architecture (superior).

## I. Small mono-region dataset (SynthRAD2023)

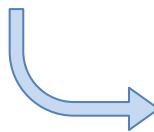


## II. Large multi-region dataset (SynthRAD2025)

(Stratified) 5-fold cross-validation

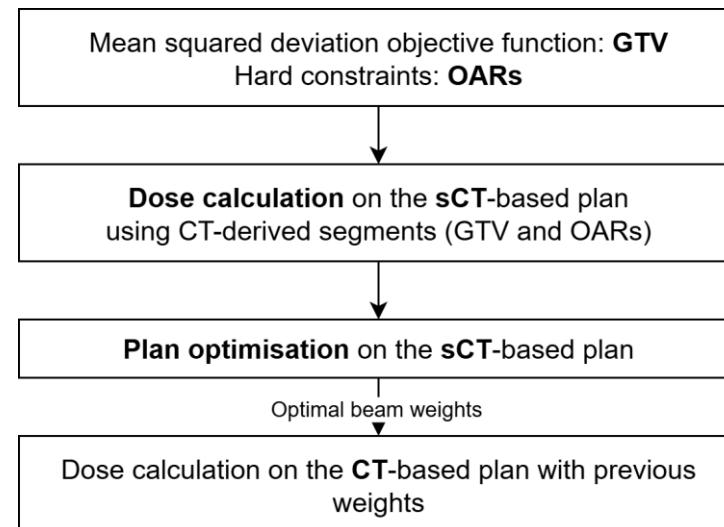


1. Compare the quality of generated **sCTs**: PSNR↑, MAE↓, SSIM↑, MS-SSIM↑
2. Compare the quality of generated **segments of the sCTs** (TotalSegmentator tool): mDICE↑, HD95↓
3. Evaluate the **radiation dose distribution** in the sCT- and CT-based treatment **plans**: MAE↓, DVH↓, Gamma pass rate (2%/2mm)↑

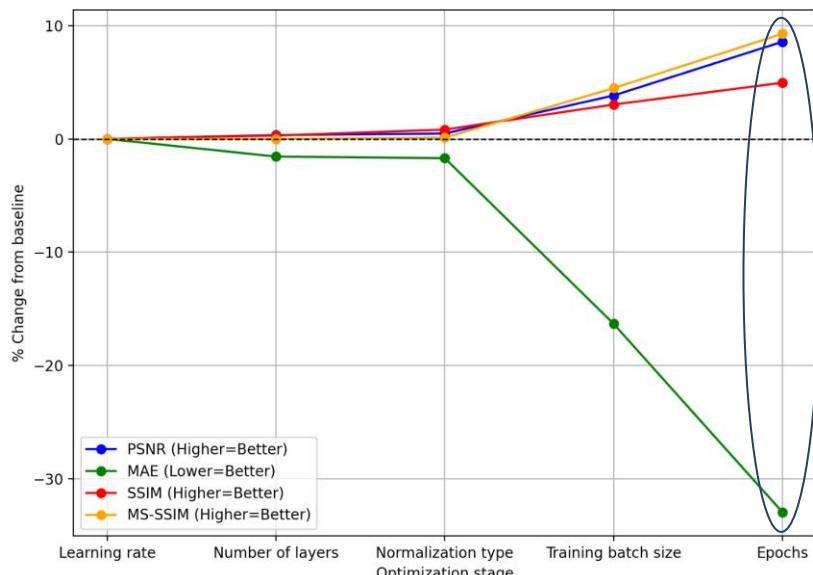


## Treatment plans (IMPT)

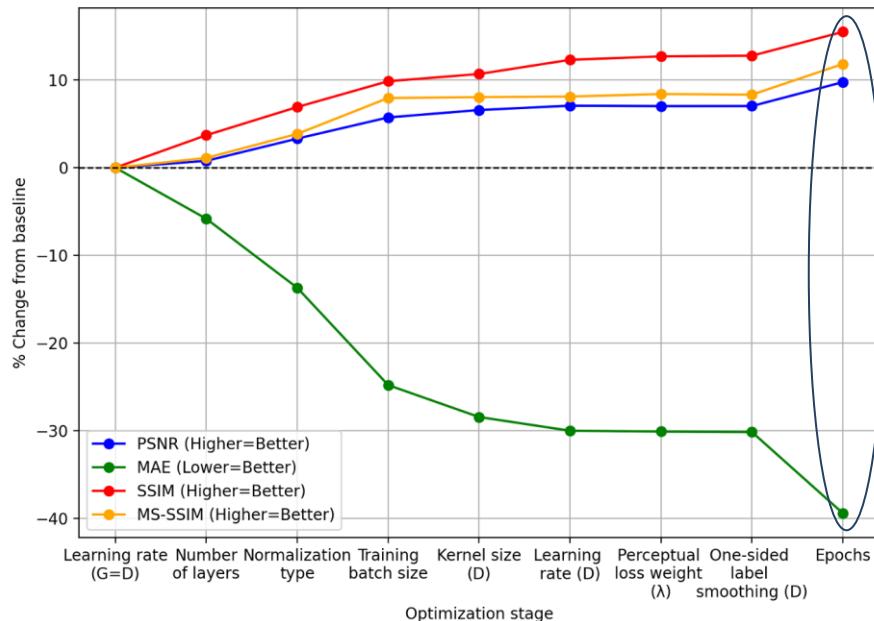
**GTV**: AB – stomach; TH – right upper lobe; HN – tongue, hard palate and soft palate.



### Hyperparameter search: AE and cGAN



**Figure 3: Progression** of the percentage change from baseline in validation metrics performance (PSNR, MAE, SSIM, MS-SSIM) across **hyperparameter optimization stages** for the **AE** model.



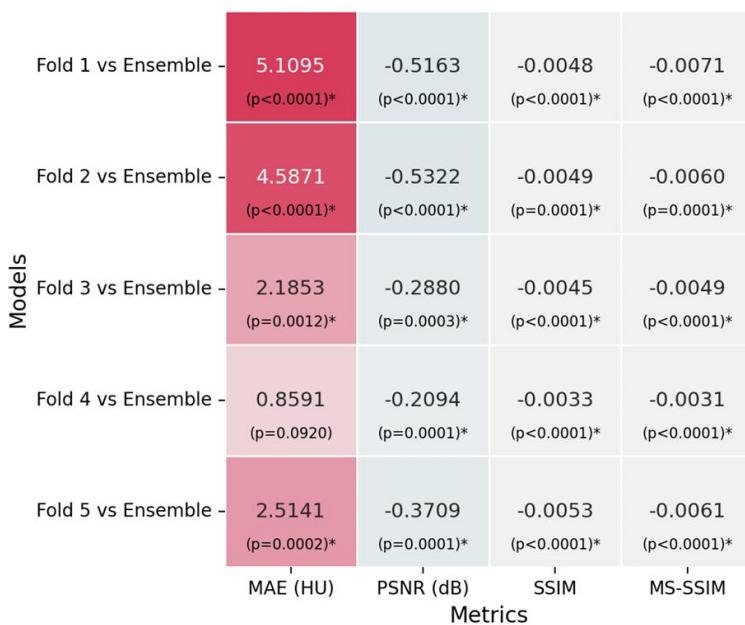
**Figure 4: Progression** of the percentage change from baseline in validation metrics performance (PSNR, MAE, SSIM, MS-SSIM) across **hyperparameter optimization stages** for the **cGAN** model. G-generator, D-discriminator.

Optimizing the **training batch size** resulted in the largest percentage drop among all stages.

# Results & Discussion

## SynthRAD2023

### Testing on 18 patients: Ensemble of models > Individual model



**Figure 5:** Performance comparison of metric differences (Fold-Ensemble) in the best performing **AE** model. **MAE**: Pink (>0): Ensemble performs better; Blue (<0): Model from fold performs better; **PSNR, SSIM, MS-SSIM**: Pink (>0): Model from fold performs better; Blue (<0): Ensemble performs better.



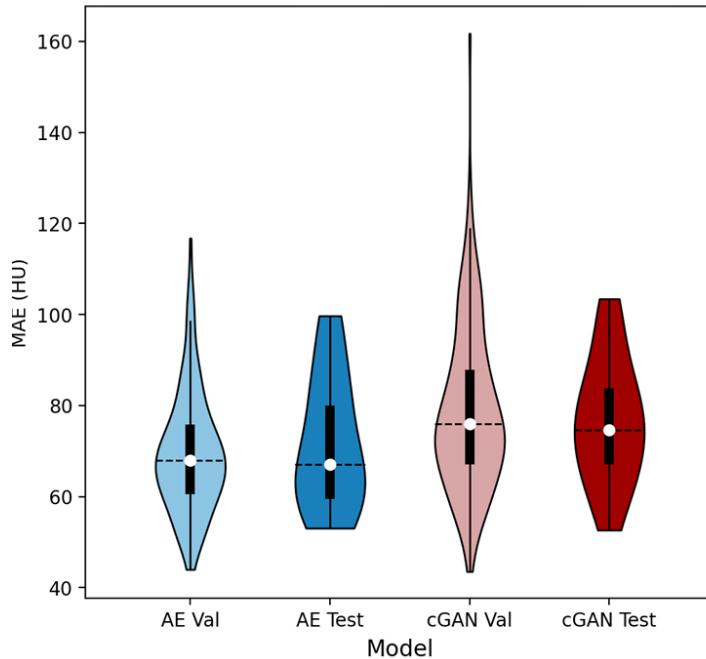
**Figure 6:** Performance comparison of metric differences (Fold-Ensemble) in the best performing **cGAN** model. **MAE**: Pink (>0): Ensemble performs better; Blue (<0): Model from fold performs better; **PSNR, SSIM, MS-SSIM**: Pink (>0): Model from fold performs better; Blue (<0): Ensemble performs better.

The **ensemble masks the overfitting** of an individual model.

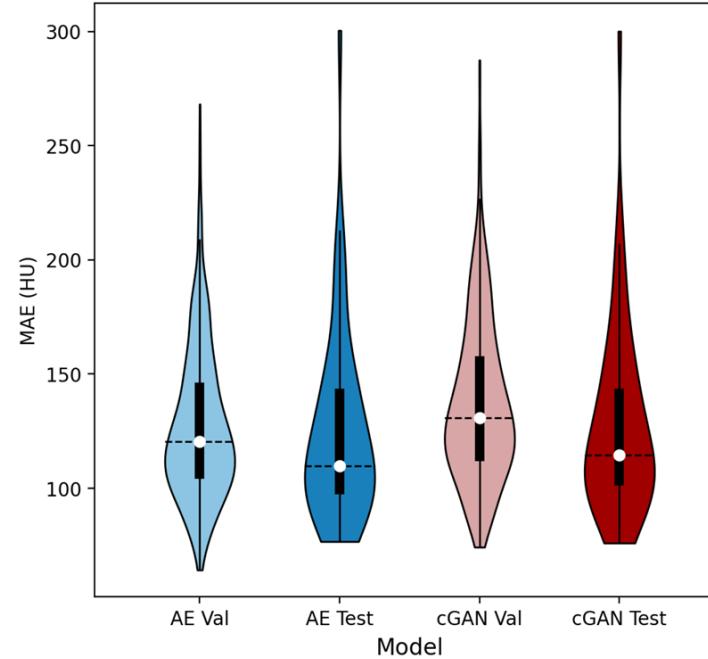
# Results & Discussion

## SynthRAD2023 vs. SynthRAD2025

### Validation and Testing (ensemble of models): AE > cGAN



**Figure 7:** Violin plots comparing the MAE between AE and cGAN best models for **SynthRAD2023**. MAE↓:  $p_{\text{val}} = 1.27 \times 10^{-13}$ ;  $p_{\text{test}} < 0.01$  (paired sample t-test showed statistically significant differences when p-value<0.05).



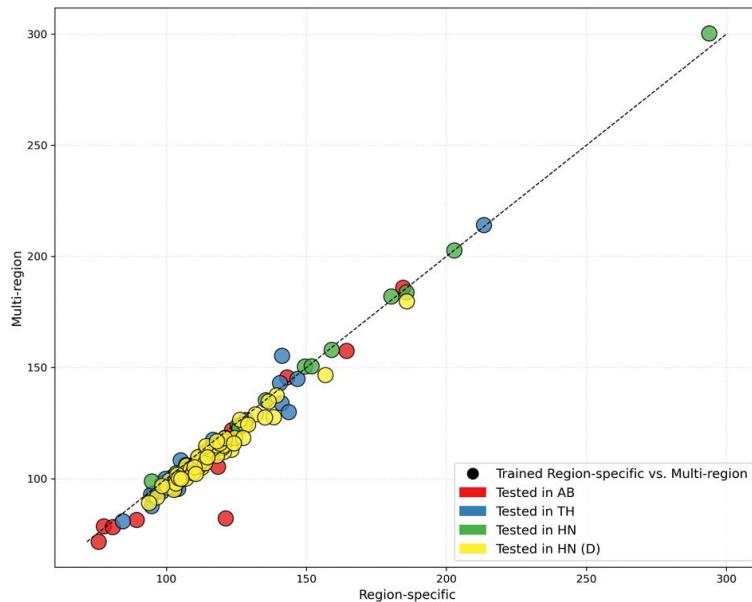
**Figure 8:** Violin plots comparing the MAE between AE and cGAN best models for **SynthRAD2025** (centre-stratified approach). MAE↓:  $p_{\text{val}} = 6.22 \times 10^{-37}$ ;  $p_{\text{test}} = 1.25 \times 10^{-5}$  (paired sample t-test showed statistically significant differences when p-value<0.05).

The **dataset** or the **HP** tuning could have impacted the results.

# Results & Discussion

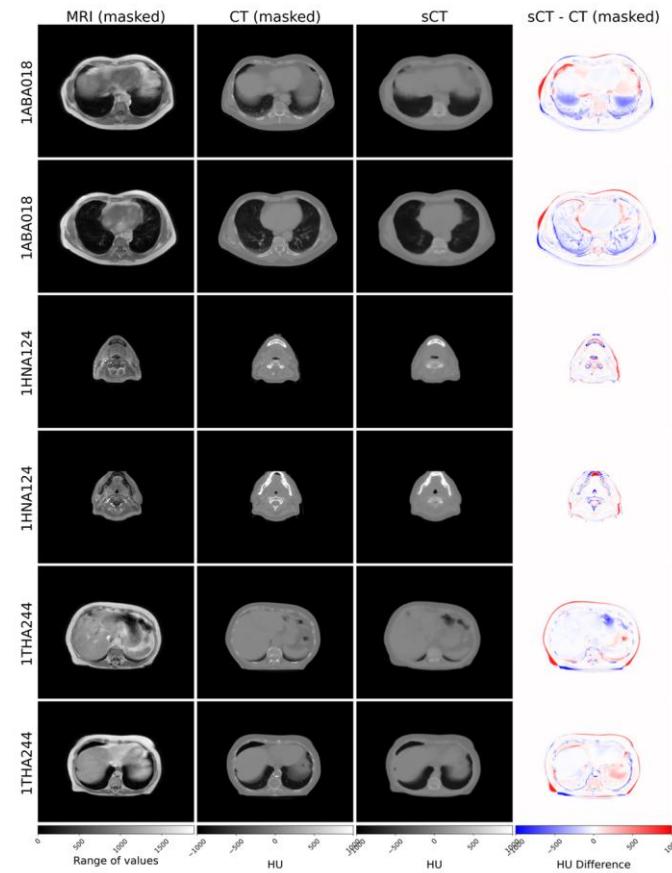
## SynthRAD2025

### Ensemble of models (AE): Multi-region model > One-region model



**Figure 9:** Comparison between the test **MAE** results of models trained in a **specific region versus a model trained in all regions**. The majority of points are below the reference line,  $y < x$ , which indicates **better performance of the multi-region model**.

- The **TH** set contained **AB** images and **vice versa**.
  - MAE showed significant **gains**.



**Figure 10:** Slices examples of **the best mean ranked patients** per region using the **multi-region model**. 1ABA018: PSNR=28.60 dB, MAE=78.59 HU, SSIM=0.83, MS-SSIM=0.91; 1HNA124: PSNR=26.11 dB, MAE=98.86 HU, SSIM=0.80, MS-SSIM=0.89; 1THA244: PSNR=27.48 dB, MAE=80.94 HU, SSIM=0.81, MS-SSIM=0.90.

# Results & Discussion

## Image quality metrics

**Table 3: Mean test results** across all patients per region for masked 3D volumes in HU. Metrics were calculated on **3D images in HU**, limited to the **region of interest** and using a **consistent data range** of 4000 HU.

Region	PSNR (dB,↑)	MAE (HU,↓)	SSIM (↑)	MS-SSIM (↑)	
<b>SynthRAD2023 test set</b>					
Pelvis	$28.09 \pm 1.51$	$71.69 \pm 14.28$	$0.84 \pm 0.03$	$0.86 \pm 0.03$	<a href="#">18 cases (12-A, 6-C)</a>
<b>SynthRAD2025 test set</b>					
AB	$25.69 \pm 1.99$	$106.99 \pm 30.93$	$0.75 \pm 0.06$	$0.79 \pm 0.09$	<a href="#">16 cases (6-A, 9-B, 1-C)</a>
TH	$25.06 \pm 1.66$	$118.46 \pm 31.27$	$0.72 \pm 0.06$	$0.81 \pm 0.07$	<a href="#">18 cases (9-A, 9-B)</a>
HN	$23.58 \pm 2.04$	$149.40 \pm 50.99$	$0.67 \pm 0.09$	$0.80 \pm 0.09$	<a href="#">15 cases (9-A, 6-C)</a>

- **HN** images from centre **C** were acquired in a **limited field of view**, which explains the mean **MAE** being the **worst**.
- **MAE(TH) > MAE(AB)**, because of the presence of **bones**, which are difficult for the model to generate (less contrast in MRI).

# Results & Discussion

## Geometric consistency metrics

### Multi-region model: HN images present best mean mDICE and HD95 (bone)

#### Expected segments in both sCT and CT

- AB & TH: segments included left and right kidneys, liver, stomach, lung lobes, vertebrae (1 sacral, 5 lumbar, 12 thoracic and 7 cervical), heart, spinal cord, ribs (12 on the left side and 12 on the right side), sternum.
- HN: segments included Oesophagus, trachea, thyroid, vertebrae (7 cervical and 12 thoracic), spinal cord, brain, skull.

#### With the obtained segments in both sCT and CT

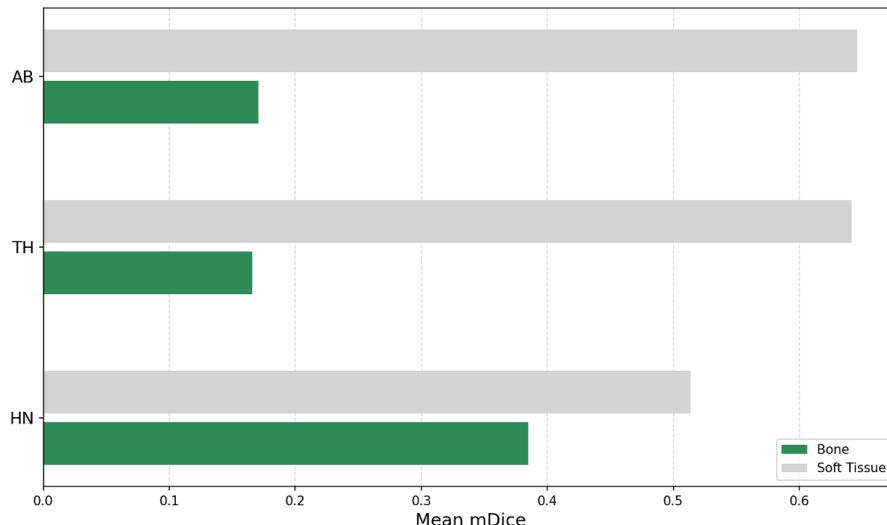


Figure 11: Mean multi-class Dice results averaged across all bone and soft tissue segments and patients within each region.

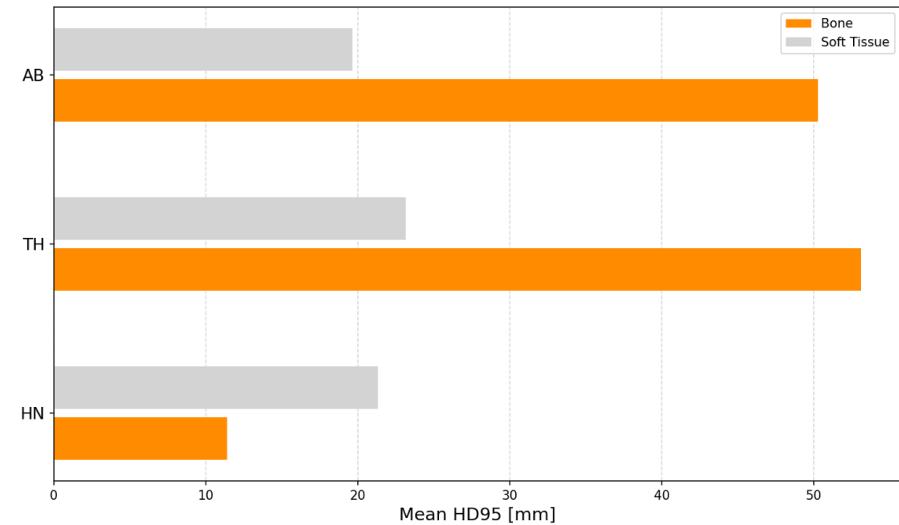


Figure 12: HD95 results averaged across all bone and soft tissue segments and patients within each region.

**Bone** generation is more **difficult** in **TH** region, due to more bone density.

# Results & Discussion

## Dose metrics

### Multi-region model: T1-weighted MRI sequence achieve superior performance

**Table 4:** Mean dose metrics acquired per region and centre from the proton treatment plan.

Region	Centre	$MAE_{\text{target dose}}$ (Gy, ↓)	$DVH$ (↓)	$\gamma_{2\%/2\text{mm}}$ (% , ↑)
AB	A	0.03±0.03	0.27±0.15	93.79±3.69
AB	B	0.02±0.01	0.38±0.34	91.89±8.52
AB	C	0.06±0.00	0.60±0.00	90.20±0.00
Mean	-	<b>0.03±0.02</b>	<b>0.35±0.29</b>	<b>92.50±6.86</b>
TH	A	0.06±0.05	0.56±0.32	70.94±14.72
TH	B	0.14±0.22	0.67±0.46	67.98±10.53
Mean	-	<b>0.10±0.17</b>	<b>0.62±0.40</b>	<b>69.37±12.76</b>
HN	A	0.01±0.00	0.30±0.09	98.74±0.94
HN	C	0.01±0.00	0.44±0.42	98.35±1.32
Mean	-	<b>0.01±0.00</b>	<b>0.35±0.29</b>	<b>98.58±1.12</b>

Exception for the MAE, centre B, AB region.

The case 1THA028 is an outlier and is not considered in this TH analysis.

- Lower mean gamma pass rate value in TH region is justified by the number of bones, especially the bones irregular shapes and heterogeneous composition of bone tissue, which affects dose modulation.
- Centre A achieved superior performance, due to the larger number of images used in training and testing.



## Main findings

**Better AE validation and testing performance** in both SynthRAD2023 and SynthRAD2025

More **robust** predictions with the **ensemble of models** in both AE and cGAN

**Multi-region model outperformed** region-specific models

**Dose calculations** computed on treatment plans based on **sCT** were **consistent** with those based on **CT**



## Limitations

The model **fails** to produce **high-resolution** images indistinguishable to the human eye

The **AE model** is **fallible**, particularly when the input (MRI) was acquired with a **limited field of view**

The **sCTs do not include** certain **anatomical structures**

## ► Future Work

1. Developing models capable of addressing **variations** among multiple centres, vendors, and MRI acquisitions in a limited field of view.
2. Exploring different **architectures** (e.g. MC-DDPM).



Champalimaud  
Foundation

**Thank you for your attention!**

---

*Physics Department, University of Coimbra*  
25/09/2025

# Results & Discussion

## Geometric consistency metrics

Structures **not** segmented by the TotalSegmentator tool - **AB**:

→ Vertebrae C1, C2, C3, C4, C5 and C6

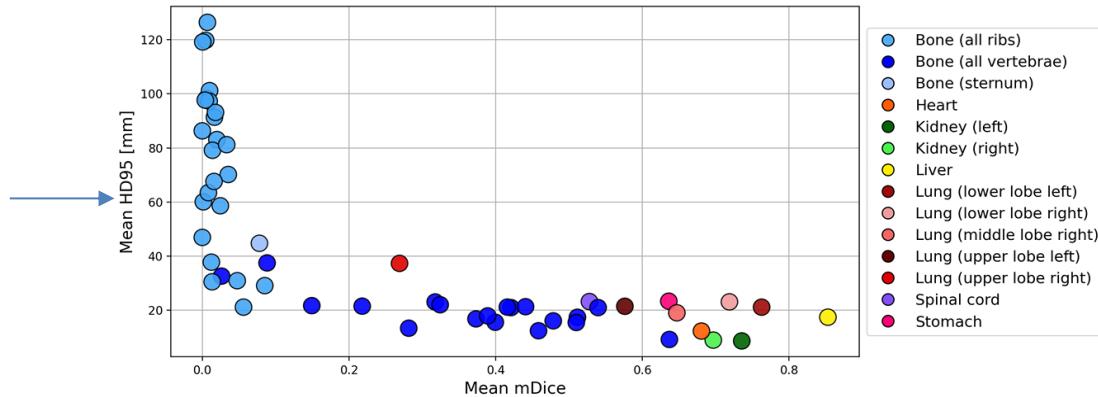


Figure 13: Mean metrics for each segment across all patients from **AB** region.

Structures **not** segmented by the TotalSegmentator tool - **TH**:

→ Vertebrae C1, C2, C3, C4, S1 and left 12th rib

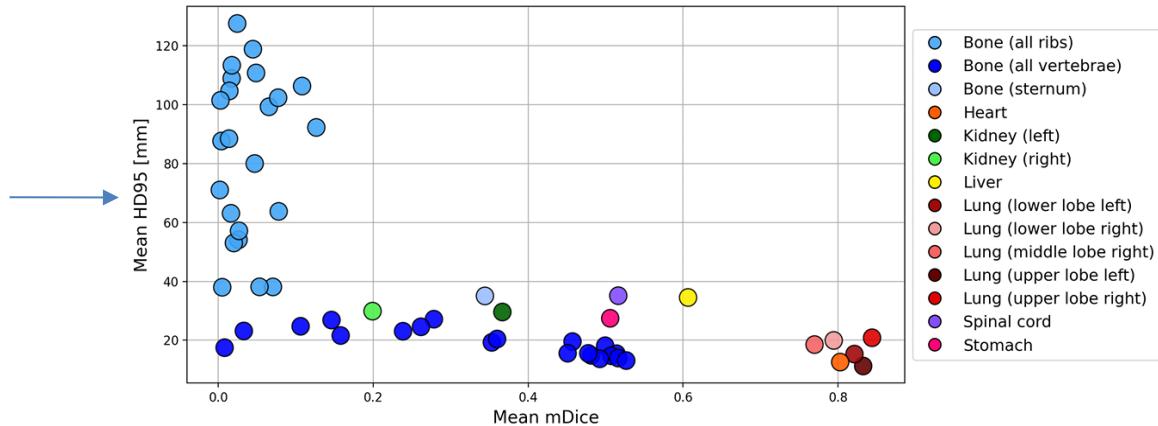


Figure 14: Mean metrics for each segment across all patients from **TH** region.