Automatic Epicardial Adipose Tissue Segmentation in CT Using 3D U-Net

José Antonio Rangel*, Ravi Vazirani, Morena Raiola, Álvaro Navarro-Guzmán, Javier Sánchez-González, Borja Ibáñez

Centro Nacional de Investigaciones Cardiovasculares Carlos III, Madrid, Spain

Existing automated methods for Epicardial Adipose Tissue (EAT) segmentation in cardiac CT lack systematic validation across heterogeneous adiposity phenotypes and rarely report reproducibility against human observer standards. We present a 3D U-Net trained on 130 calcium scoring CT scans acquired on a Philips Brilliance 16-slice CT scanner from the PESA (Progression of Early Subclinical Atherosclerosis) CNIC-SANTANDER study, stratified by EAT volume into three categories (low, medium, high) across training (100) and validation (30) sets. Images were preprocessed with a Hounsfield Unit window with width 500 and level 50. The network was optimised with AdamW, cosine annealing, and a combined Dice-Focal loss. Augmentation included random affine transformations and Gaussian smoothing. Inference used sliding window with overlap. The model was benchmarked against SwinUNETR and a 3D U-Net with attention gates on the same dataset. Reproducibility was assessed on two independent test set (20 each): one for intra-observer and another one for inter-observer analysis, both providing ICC, Passing–Bablok regression and Bland-Altman.

On the test set, the 3D U-Net achieved the highest Dice (0.795) and lowest volume error (10.7%) among all architectures (Table 1). SwinUNETR underperformed due to the tension between its global attention mechanism and patch-based inference, amplified by the limited training set size. The attention-gate variant showed the lowest Dice (0.750) and highest HD95 (35.17mm), suggesting architectural complexity may hinder convergence with small datasets. Despite being the most compact architecture evaluated, the 3D U-Net achieved reproducibility within human observer range. In the reproducibility study, human inter-observer ICC = 0.979 (0.855; 0.994) with Bias = 2.45cm³. The AI model achieved ICC = 0.987 and Bias = 1.56cm³, falling within the inter-observer range and confirming comparable agreement to human experts. These results position the proposed pipeline as a robust and scalable alternative to manual EAT annotation in large-scale CT studies.

Table 1. Segmentation performance and reproducibility analysis

Model	Dice	IoU	Sens	Spec	HD95	VErr%
3D U-Net	**0.795**	**0.663**	0.804	**0.997**	28.70	**10.70**
SwinUNETR	0.770	0.628	**0.822**	0.996	**24.83**	13.64
3D U-Net + Att.	0.750	0.616	0.759	0.997	35.17	14.21

	Cohort description			Agreement metrics					
	Mean±SD (cm^3)	Med. (cm^3)	Range (cm^3)	ICC (95% CI)	R	Slope	Int. (cm^3)	Bias (cm^3)	LoA (cm^3)
H. intra-obs	89.1±42.8	82.75	34.1–162.6	0.984 (0.961;0.994)	0.999	1.006	−1.09	0.59	−4.1,+5.3
H. inter-obs	87.4±53.6	82.60	25.1–209.1	0.979 (0.855;0.994)	0.989	0.883	7.79	2.45	−16.3,+21.2
AI vs cons.	87.4±53.6	82.60	25.1–209.1	0.987 (0.965;0.994)	0.987	0.989	2.52	1.56	−14.8,+17.9

LoA: Bland-Altman limits. Int.: Passing-Bablok intercept. Bold = best (Table 1).