

**FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO**

# **Segmentation of Cardiac Tissue in CT for Coronary Artery Disease Screening**

**Rúben Baeza da Silva**



Mestrado em Engenharia Biomédica

Supervisor: Prof. Dr. João Manuel Patrício Pedrosa

Second Supervisor: Prof. Dr. Francesco Renna

September 25, 2023



# Abstract

Coronary artery disease (CAD) is the leading cause of heart disease and death worldwide. It is caused by the buildup of various substances, such as fats, calcium, and inflammatory cells, in the walls of the coronary arteries, resulting in the formation of plaques. This can lead to myocardial infarction or chest pain. The most commonly non-invasive method for diagnosing CAD involves quantifying coronary artery calcium through CT scans, where elevated calcium levels indicates potential plaque presence. However, the presence of non-calcified plaques can result in poor diagnosis.

Several studies have pointed to a possible relationship between the thickness and volume of epicardial adipose tissue (EAT) and cardiovascular diseases, even claiming that it could be an independent marker for CAD. Therefore, accurately measuring and quantifying EAT through its segmentation is becoming increasingly important. CT imaging is currently the most commonly used method for assessing EAT. The pericardium is particularly difficult to define as it is a thin layer that surrounds the heart, measuring only 2 mm in thickness. Accurately identifying the pericardium is fundamental to distinguish the different types of fat found in the heart, including EAT. However, manual segmentation is a time-consuming process and often leads to disagreement among experts.

This work presents a study towards the fully automatic segmentation of the pericardium from non-contrast CT scans. Through this document, many experiments and comparisons were conducted in order to find the best approach to segment this structure. It was found that a 2.5D U-Net provided with three consecutive axial slices was the model that achieved the most satisfactory results. Besides that, data augmentation techniques and artificial calcifications were proved to optimize and deal better with some variations observed in the test set. A post-processing step was also determined to be beneficial to eliminate disconnected components and inconsistencies in the automatic pericardial segmentation. The final results for the pericardium segmentation yielded Dice Similarity Coefficient (DSC), Hausdorff Distance (HD) and Mean Absolute Distance (MAD) values of  $0.909 \pm 0.024$ ,  $32.443 \pm 13.729$  mm, and  $4.389 \pm 1.291$  mm, respectively, for a private dataset comprising 190 patients.

After segmenting the pericardium, EAT was separated using established Hounsfield Units (HU) fat range criteria within the pericardium. In 20 patients, intra and interobserver variability were assessed to determine the potential of the automated approach to function as an independent reader, potentially reducing variability. The automated method showed promising results in volume quantification metrics, achieving a Pearson correlation coefficient (PCC) of 0.924 and a bias of  $0.98 \text{ cm}^3$ . Segmentation metrics, including DSC, recall, and precision, achieved values of  $0.749 \pm 0.051$ ,  $0.766 \pm 0.069$ , and  $0.748 \pm 0.085$ , respectively. However, the automated model couldn't fully replace independent human readers. It's important to highlight the complexity of pericardium delineation, even among experts.

In conclusion, incorporating contextual information into the U-Net was shown to be primordial for an accurate segmentation. However the main concern remain the lack of high-quality publicly

available data, affecting the generalization of DL models. The considerable value of this study lies in training the model exclusively with public data and testing it on an external dataset of 190 patients. This aspect is notably absent in the existing literature.



# Resumo

CAD ou doença da artéria coronária é a principal doença cardíaca e principal causa de mortalidade do mundo. CAD é provocada pelo acúmulo de várias substâncias (gorduras, cálcio, células inflamatórias) nas paredes das artérias coronárias, formando assim placas. Este fenómeno pode provocar infarte do miocárdio ou dores no peito. O método de diagnóstico para a CAD mais usado, é a quantificação do cálcio nas artérias coronárias a partir de imagens CT. Elevadas concentrações de cálcio são um forte indicador da possível formação de placa. No entanto, podem existir placas não calcificadas, contribuindo para um mau diagnóstico dos pacientes.

Vários estudos apontaram uma possível relação entre a espessura e volume do tecido adiposo do epicárdio (EAT) com doenças cardiovasculares, afirmando mesmo que poderia ser um marcador independente para a CAD. Com isto, a sua segmentação é de extrema relevância para possibilitar a sua quantificação. A técnica imagiológica mais usada atualmente para avaliar e segmentar o EAT é CT. O pericárdio é uma das estruturas mais desafiadoras para se delimitar, pois é uma camada fina que reveste o coração, com cerca de apenas 2 mm de espessura. A sua delimitação permite a diferenciação das diferentes gorduras do coração, incluindo EAT. No entanto, a segmentação manual é um processo demasiado lento e gera muita discordância entre os especialistas.

Este trabalho apresenta um estudo para a segmentação automática do pericárdio a partir de imagens CT sem contraste. Ao longo deste documento, muitas experiências foram realizadas para encontrar a melhor abordagem para segmentar essa estrutura. Verificou-se que uma U-Net 2.5D fornecida com três cortes axiais consecutivos foi o modelo que apresentou resultados mais satisfatórios. Além disso, técnicas de aumento de dados e calcificações artificiais mostraram-se eficazes para lidar com algumas variações observadas no conjunto de testes. Uma etapa de pós-processamento foi útil para eliminar artefactos desconectados e inconsistências na segmentação pericárdica. A segmentação do pericárdio apresentou valores para a Dice Similarity Coefficient (DSC), Hausdorff Distance (HD) e Mean Absolute Distance (MAD) de  $0.909 \pm 0.024$ ,  $32,443 \pm 13,729$  mm e  $4,389 \pm 1,291$  mm, respectivamente, para uma base de dados externa que contém 190 pacientes.

Após a segmentação do pericárdio, segmentou-se o EAT através da aplicação da faixa de Hounsfield Units (HU) para a gordura dentro do pericárdio. Em 20 pacientes, a variabilidade intra e interobservador foi avaliada para determinar se a nossa abordagem pode ser usada como uma ferramenta independente. O método automatizado mostrou resultados promissores em métricas de quantificação de volume, alcançando um coeficiente de correlação de Pearson (PCC) de 0.924 e um bias de  $0.98 \text{ cm}^3$ . Nas métricas de segmentação, incluindo DSC, recall e precisão, obteve-se valores de  $0.749 \pm 0.051$ ,  $0.766 \pm 0.069$ , and  $0.748 \pm 0.085$ , respectivamente. No entanto, o modelo automático não pode, ainda, substituir totalmente os especialistas. É importante realçar a complexidade na identificação do pericárdio, mesmo entre especialistas.

Em conclusão, demonstrou-se que a incorporação de informação contextual na U-Net é fundamental para uma segmentação precisa. No entanto, a principal preocupação permanece na falta de dados públicos de alta qualidade, afetando a generalização dos modelos de DL. O valor con-

siderável deste estudo reside no treino exclusivo do modelo com dados públicos e na sua posterior avaliação num conjunto de dados externo composto por 190 pacientes. Este aspeto encontra-se notavelmente ausente na literatura existente.

# Agradecimentos

Primeiramente agradecer aos meus pais que foram fantásticos, um constante suporte em todos os aspetos da minha vida e que obviamente sem eles nada disto seria possível. Agradecer também a todos os meus amigos e colegas que me acompanharam nesta grande aventura nas duas fantásticas cidades de Aveiro e Porto que me acolheram neste percurso académico. Importantes tanto no estudo e em momentos de descontração que, na minha opinião, são essenciais para o sucesso e equilíbrio de uma vida académica.

Um agradecimento especial ao meu orientador João Pedrosa que se demonstrou impecável pela sua acessibilidade e constante ajuda sem os quais este trabalho não seria possível. Não posso deixar de agradecer também ao meu co-orientador Francesco Renna pelo seu apoio e opiniões que ajudaram a dar rumo aos vários obstáculos que surgiram. Como é óbvio não posso deixar de referir os professores e colegas de laboratório pelo acolhimento e confiança depositada.

Rúben Baeza da Silva



*“Those who are content sit still and do nothing,  
those who are discontented are the only benefactors in the world.”*

Walter S. Landor



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Context . . . . .	1
1.2	Goals . . . . .	2
1.3	Document Structure . . . . .	2
<b>2</b>	<b>Background</b>	<b>3</b>
2.1	Heart Anatomy . . . . .	3
2.2	Coronary Artery Disease . . . . .	4
2.2.1	The Role of EAT in CAD . . . . .	7
2.3	Imaging Techniques for Cardiac Fat Measurement . . . . .	8
2.3.1	Echocardiography . . . . .	8
2.3.2	Cardiac Magnetic Resonance . . . . .	8
2.3.3	Computed Tomography (CT) . . . . .	9
2.4	Towards Automatic Pericardial Fat Segmentation . . . . .	10
<b>3</b>	<b>Automatic Pericardial Fat Segmentation on Non-contrast CT Images</b>	<b>11</b>
3.1	State of the Art . . . . .	11
3.1.1	Atlas-based Techniques . . . . .	11
3.1.2	Machine Learning Techniques . . . . .	13
3.1.3	Deep Learning Techniques . . . . .	17
3.2	Evaluation Metrics . . . . .	24
3.2.1	Accuracy . . . . .	24
3.2.2	Precision . . . . .	24
3.2.3	Recall . . . . .	24
3.2.4	Dice Similarity Coefficient (DSC) . . . . .	24
3.2.5	Hausdorff Distance (HD) . . . . .	25
3.2.6	Mean Absolute Distance (MAD) . . . . .	25
3.2.7	Pearson Correlation Coefficient (PCC) . . . . .	25
3.3	Conclusion . . . . .	26
<b>4</b>	<b>A First Approach for Pericardial Segmentation</b>	<b>29</b>
4.1	Methods . . . . .	29
4.1.1	Datasets . . . . .	29
4.1.2	Image Registration . . . . .	32
4.1.3	Pericardial Segmentation . . . . .	33
4.2	Experiments . . . . .	35
4.3	Results and Discussion . . . . .	36
4.3.1	Image Registration . . . . .	36

4.3.2	Pericardial Segmentation . . . . .	36
4.4	Final Outcomes . . . . .	43
<b>5</b>	<b>Data Augmentation and Artificial Calcifications</b>	<b>45</b>
5.1	Methods . . . . .	45
5.1.1	Data Augmentation . . . . .	45
5.1.2	Artificial Calcifications . . . . .	46
5.1.3	Training Setup . . . . .	48
5.2	Experiments . . . . .	48
5.3	Results and Discussion . . . . .	48
5.4	Final Outcomes . . . . .	54
<b>6</b>	<b>Three Dimensional Approaches</b>	<b>55</b>
6.1	Methods . . . . .	55
6.1.1	2.5D U-Net . . . . .	55
6.1.2	Preprocessing . . . . .	56
6.1.3	Post-processing . . . . .	58
6.2	Experiments . . . . .	59
6.3	Results and Discussion . . . . .	59
6.3.1	2D Model vs 2.5D Model . . . . .	59
6.3.2	Preprocessing . . . . .	63
6.3.3	Post-processing . . . . .	66
6.4	Final Outcomes . . . . .	70
<b>7</b>	<b>EAT Segmentation and Quantification</b>	<b>71</b>
7.1	Methods . . . . .	71
7.1.1	EAT Segmentation . . . . .	71
7.1.2	Evaluation Metrics . . . . .	72
7.2	Results and Discussion . . . . .	73
7.3	Final Outcomes . . . . .	77
<b>8</b>	<b>Conclusion</b>	<b>79</b>
	<b>References</b>	<b>81</b>



# List of Figures

2.1	Illustration of the layers of the heart. . . . .	3
2.2	Illustration of the mechanism of CAD. . . . .	5
2.3	CAG before (A) and after (B) Coronary Angioplasty to restore blood flow to the heart. . . . .	5
2.4	CACS exam without (left) and with (right) calcium in the arteries. . . . .	6
2.5	CCTA exam with the visualization of the coronary arteries and respective severity of stenosis. . . . .	6
2.6	Epicardial versus paracardial fat thickness. Paracardial fat (within yellow arrows and yellow dashed shape), epicardial fat (within red arrows and red dashed shape). . . . .	8
2.7	The quantification of EAT and MAT by CMR. The pericardium was identified. The EAT and MAT are identified in blue and green, respectively. . . . .	9
2.8	Segmented EAT in Non-Contrast (left) and Contrast-Enhanced (right) CT Images. . . . .	9
2.9	Illustration of the pericardium, EAT and paracardial fat in a CT image. . . . .	10
3.1	Illustration of the registration process. . . . .	13
3.2	An illustration of using active contours to fill in missing sections of a pericardium line, depicted in white. The contour is displayed in blue, starting as a perfect circle around the image. The contour is gradually altered in each iteration to better conform to the image. Note that this is only the rational about geodesic active contour and not the exact method used bu Ding et al. . . . .	14
3.3	Creation of the atlas by Rodrigues et al. . . . .	15
3.4	Workflow for the proposed methodology of Priya et al. . . . .	16
3.5	CNN achitecture proposed by Commandeur et al. . . . .	18
3.6	U-Net architecture presented by Ronneberger et al. . . . .	19
3.7	Architecture of a pyramid pool module. . . . .	20
3.8	U-Net architecture proposed by Siriapisith et al. with the AG and DSV modules implemented in the standard U-Net. . . . .	21
3.9	The full structure of the automated EAT segmentation training process and the preprocessing steps performed by Hoori et al. . . . .	22
3.10	The architecture of the "EAT-Net" and "Crop-net" performed by Molnar et al. . . . .	23
3.11	MRA-U-Net architecture proposed by Qu et al. . . . .	23
4.1	Cardiac Fat dataset. Left: DICOM images; right: DICOM in fat HU range and manual segmentations of cardiac fat. Red indicates EAT, green indicates MAT. . . . .	30
4.2	OSIC Dataset. DICOM images with manual pericardium segmentation in red. . . . .	31
4.3	CHVNGE Dataset. DICOM images with manual pericardium segmentation in red. . . . .	31
4.4	Example of the matching of key points between the two images using BRFFMatcher. . . . .	32

4.5	Convex hull pericardium labelling example. Left: EAT mask; middle: generated pericardium label using convex hull; right: EAT (red) and Pericardial (yellow) labels registered and overlapped with DICOM images. . . . .	33
4.6	CT image shown in ranges [-1000, 1000] HU (left) and [-125, 225] HU (right) showing pericardium layer enhancement. . . . .	34
4.7	U-Net architecture used for pericardial segmentation. . . . .	34
4.8	Illustration of the method for alternating between datasets for each batch during training. . . . .	35
4.9	Two examples ((a) and (b)) before (left) and after (right) image registration. . . .	36
4.10	Results for the three datasets for each image size, where the three metrics are displayed using boxplots. The mean is depicted as a point with the corresponding value written above the maximum value of each boxplot. . . . .	37
4.11	Examples of automatic pericardium segmentation (PS) for the two input sizes in the public datasets. . . . .	38
4.12	Examples of automatic pericardium segmentation (PS) for the two input sizes in the CHVNGE datasets. . . . .	39
4.13	Results of the loss function and HU range for the three datasets, where the three metrics are displayed using boxplots. Note: blue and green color for the boxplots represents the [-1000, 1000] HU and [-125, 225] HU range, respectively. . . . .	40
4.14	Examples of automatic pericardium segmentation (PS) for the two different HU ranges in the public datasets. . . . .	42
4.15	Examples of automatic pericardium segmentation (PS) for two patients from CHVNGE dataset, considering the two different HU ranges. . . . .	43
5.1	Example of the applied data augmentation transformations on an illustrative CT slice. . . . .	46
5.2	Example of a few calcifications in the CHVNGE dataset. . . . .	46
5.3	Example of artificial calcifications and the effect of manipulating parameters of the Gaussian. . . . .	47
5.4	The results of each model for the three datasets, where the three metrics are displayed using boxplots, with the mean depicted as a point with the corresponding value written on the upper side. Note that $S$ = Standard model; $DA$ = DA model; $DA + AC$ = DA+AC model . . . . .	49
5.5	Examples of the effect of data augmentation on the automatic pericardium segmentation for the public datasets. . . . .	50
5.6	Examples of the effect of data augmentation on the automatic pericardium segmentation for the CHVNGE dataset. . . . .	51
5.7	Examples of the effect of each one of the techniques and their combination, on two patients from the public datasets. . . . .	52
5.8	Examples of the effect of the combination of the artificial calcifications with data augmentation on two patients from CHVNGE dataset. . . . .	53
6.1	Example of the input and output of the 2.5D model. . . . .	56
6.2	Illustration of the strategy of slice classification with the 2/2.5D model for pericardium segmentation. . . . .	57
6.3	Illustration of the strategy of 3D segmentation for preprocessing with the 2/2.5D model for pericardium segmentation. . . . .	58
6.4	Example of a segmentation in 3D of a heart. Left: manual segmentation. Right: automatic segmentation from the 2D model. . . . .	59

6.5	Illustration of the post-processing steps including: the selection of the biggest component component, application of the "Axial Fill" in consecutive slices and application of the Convex hull for each slice. . . . .	60
6.6	The results of the 2D and 2.5D for the three datasets, where the three metrics are displayed using boxplots, with the mean depicted as a point with the corresponding value written on the upper side. . . . .	61
6.7	Examples of automatic pericardium segmentations using the 2D and 2.5D model for the public datasets. . . . .	62
6.8	Examples of automatic pericardium segmentations using the 2D and 2.5D model for the CHVNGE dataset. . . . .	63
6.9	The results of the 2.5D model along with preprocessing approaches for the three datasets, where the three metrics are displayed using boxplots, with the mean depicted as a point with the corresponding value written on the upper side. . . . .	64
6.10	Examples of the effect of both preprocessing approaches on the automated pericardium segmentation for the public datasets. . . . .	65
6.11	Examples of the effect of both preprocessing approaches on the automated pericardium segmentation for the CHVNGE dataset. . . . .	66
6.12	The outcomes of the 2.5D model and 3D + 2.5D model, combined with post-processing, are presented for the three datasets. Boxplots are utilized to visualize the three metrics, with the mean represented as a point accompanied by the corresponding value indicated on the upper side. . . . .	67
6.13	Examples of the impact of the post-processing approach on the automatic pericardium segmentation for the public datasets. . . . .	68
6.14	Examples of the impact of the post-processing approach on the automatic pericardium segmentation for the private dataset. . . . .	69
7.1	Process of segmenting the EAT from CT scans. . . . .	72
7.2	Volume quantifications from two measurements from Reader 1 to access intravariability. The gray line depicted in the Scatter plot illustrates the ideal scenario where the two measurements align perfectly, resulting in a PCC of 1. . . . .	73
7.3	Volume quantifications from the Reader 2 versus the manual quantifications from the Reader 1 to access intervariability. The gray line depicted in the Scatter plot illustrates the ideal scenario where the two measurements align perfectly, resulting in a PCC of 1. . . . .	74
7.4	Automated volume quantifications versus the manual quantifications from the Reader 1. The gray line depicted in the Scatter plot illustrates the ideal scenario where the two measurements align perfectly, resulting in a PCC of 1. . . . .	75
7.5	Examples of EAT segmentation performed by the two readers and the automatic method. . . . .	76



# List of Tables

3.1	Summary of fully-automatic algorithms for segmentation of pericardial fats from non-contrast CT scans and respective results. DL= deep learning; EM=elliptical modelling; GAC= geodesic active contours; ML= machine learning; N= Number of patients; P= pericardium. . . . .	12
6.1	Results of MAED for all the three test sets. Results shown in mean $\pm$ standard deviation mm. . . . .	63
7.1	Results of EAT segmentation metrics for intrareader, interreader and automatic approach. Results shown in mean $\pm$ standard deviation (std). . . . .	75



# Abbreviations and Acronyms

Adam	Adaptive Moment Estimation Algorithm
ACS	Attention to clinical scenario
AC	Artificial Calcifications
AI	Artificial intelligence
AG	Attention Gate
ASPP	Atrous Spatial Pyramid Pooling
AUC	Average Area Under the ROC Curve
BRIEF	Binary Robust Independent Elementary Features
CAD	Coronary Artery Disease
CAG	Coronary Angiography
CACS	Coronary Artery Calcium Scan
CCTA	Coronary Computed Tomography Angiography
CHVNGE	Centro Hospitalar Vila Nova de Gaia e Espinho
CNN	Convolutional Neural Network
COT	Compose thesis
CLAHE	Contrast Limited Adaptive Histogram Equalization
CMR	Cardiac Magnetic Resonance
CVD	Cardiovascular disease
CT	Computerized Tomography
DA	Data Augmentation
DL	Deep learning
DSC	Dice similarity coefficient
DSV	Deep supervision
EAS	Perform EAT segmentation
EAT	Epicardial Adipose Tissue
EM	Elliptical modelling
EPICHEART	The influence of EPICardial adipose tissue in HEART diseases
FDA	Fisher discriminant analysis
GA	Genetic Algorithm
GAC	Geodesic active contours
GLCM	Grey level co-occurrence matrix
GWO	Grey Wolf Optimizer
HD	Hausdorff distance
HCU	Hidden Convolutional Units
HU	Hounsfield units
IHD	Ischemic heart disease
IoU	Intersection over Union
IPS	Improve the performance of the pericardium segmentation

IRP	Ideal Region Inside the Pericardium
MACE	Major Adverse Cardiac Events
MAD	Mean Absolute Distance
ML	Machine learning
MRA	Magnetic Resonance Angiography
MRG	Modified Region Growing Algorithm
MSC	Multiscale Concatenate Module
N	Number of patients
ONN	Optimal Neural Network
ORB	Oriented FAST and Rotated BRIEF
OSIC	Open Source Imaging Consortium
P	Pericardium
PS	Pericardium Segmentation
PAT	Pericardial Adipose Tissue
PCC	Pearson correlation coefficient
RANSAC	Random Sample Consensus
RBFNetwork	Radial Basis Function Network
ReLU	Rectified Linear Unit
RMSprop	Root Mean Square propagation
ROI	Regions of Interest
S	Standard
SMO	Sequential Minimal Optimization
SVM	Support Vector Machine
T	Traditional



# Chapter 1

## Introduction

### 1.1 Context

Coronary artery disease (CAD), also known as Ischemic heart disease (IHD), is the most common type of cardiovascular disease (CVD) and the leading cause of death worldwide. CAD occurs when the coronary arteries become hardened and narrowed, reducing blood flow to the heart muscle and leading to ischemia. This is caused by the buildup of plaque, made up of calcium, fatty lipids, and inflammatory cells, on the inner walls of coronary arteries. This can lead to chest pain (angina) or a myocardial infarction (heart attack) [1]. In fact, according to a study in the European Journal of Preventive Cardiology [2]:

" In 2019, there were  $\sim 56.5$  million deaths globally, with 32.9% of them (18.6 million deaths) being due to CVDs.; In 2019, IHD accounted for 197.2 million (177.7-219.5) prevalent cases, 9.1 million (8.4-9.7) deaths ..."

Many risk factors increase the probability of CAD including high cholesterol, diabetes, smoking, being overweight, or family history of heart disease. A common non-invasive technique for screening CAD is a coronary calcium scan, which uses computerized tomography (CT) to detect calcium deposits in the coronary arteries. However, CAD can be present without calcium and may not be detected by this technique, leading to misdiagnosis [3].

Studies have also suggested a link between expanded epicardial fat and cardiovascular diseases [4]. Epicardial adipose tissue (EAT) is a type of fat located within the pericardium, which is a thin sac that covers the heart. CT is currently the preferred method for quantifying EAT due to its high-spatial resolution and lower cost compared to other imaging techniques [5]. However, manually segmenting and quantifying EAT can be difficult and time-consuming, leading to the development of automatic methods using AI algorithms such as machine learning and deep learning.

## 1.2 Goals

The objectives of this thesis are:

- Develop a method to fully-automatic segment the pericardium and EAT on non-contrast CT images based in deep learning approach.
- Evaluation of the method in an external dataset provided by Centro Hospitalar Vila Nova de Gaia e Espinho (CHVNGE).

## 1.3 Document Structure

Besides the introduction, this document is structured in more seven chapters:

Chapter 2 gives an overview of CAD and the function of EAT, including crucial aspects of heart anatomy, current diagnostic strategies for CAD, and imaging methods for segmenting cardiac fats.

Chapter 3 provides a state of the art review of the current strategies to fully-automatic segment the pericardium and pericardial fats on non-contrast CT images including atlas, machine learning and deep learning based techniques. Additionally gives an overview of the main metrics to access the performance of the segmentation.

Afterwards, Chapter 4 presents a first investigation towards the fully-automatic segmentation of the pericardium on non-contrast CT using a U-Net. The effects of varying the input image size, two different HU ranges, and two distinct loss functions were assessed.

Chapter 5 presents an innovative and conventional technique to enable the model to handle certain variations observed within the CHVNGE dataset.

Chapter 6 employs three-dimensional techniques to address the concerns outlined in chapter 5, particularly the difficulties encountered in delineating the lower pericardium and defining the upper and lower borders of the heart.

Chapter 7 describes the EAT segmentation approach using the best pericardium segmentation model found. It evaluates automated EAT segmentation against intra and interreader variability, using volume and segmentation metrics.

Finally, the main conclusions, challenges and future work are discussed in Chapter 8.

## Chapter 2

# Background

### 2.1 Heart Anatomy

The heart is a vital organ that pumps blood throughout the body, supplying oxygen and nutrients to all cells. When a problem arises with the heart, it can put our lives at risk. It is essential to have a basic understanding of the anatomy of the heart, the reasons why problems may occur, and how to prevent or address them.

The wall of the heart is composed of three layers of tissue: the epicardium, the myocardium, and the endocardium. The myocardium, the thick middle layer of the heart wall, is made up of cardiac muscle cells that are responsible for the heart's ability to contract. The endocardium is the thin inner surface of the heart chambers. It is made of a connective tissue that helps the blood flow easily through the heart. The epicardium is a thin, serous membrane that covers the outer surface of the heart. Figure 2.1 depicts the layers of the heart [6].

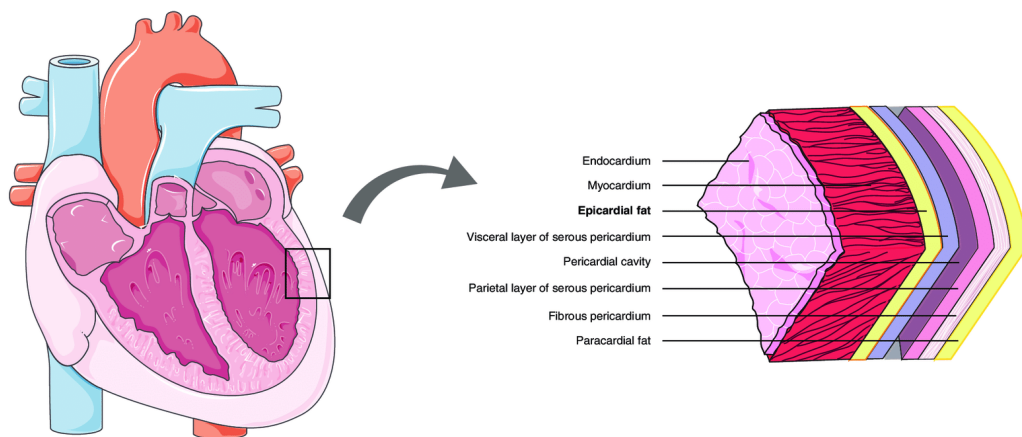


Figure 2.1: Illustration of the layers of the heart. Image from [7].

The heart itself is enclosed by the pericardium, which is a thin fibrous sac that consists of two main layers. The outer layer is called the fibrous pericardium and the inner layer is called the serous pericardium. The fibrous pericardium is a dense and inelastic structure that prevents the heart from over-expanding and holds it in place in the mediastinal space. Inside the fibrous pericardium, there is the serous pericardium. The part of the serous pericardium that is closely attached to the inner surface of the fibrous pericardium is called the parietal layer. The part that covers the inner surface of the heart is called the visceral layer or epicardium [6]. Figure 2.1 illustrates the layers mentioned. It is important to note that the terms "epicardium" and "visceral layer" refer to the same structure. However, when referring to the pericardium, the term "visceral layer" is used and when referring to the heart, the term "epicardium" is used.

The cardiac fat or adipose tissue surrounding the heart is an important topic in this work. The EAT is a type of fat that is located between the myocardium and the epicardium (as well as within the pericardium). The adipose tissue that is located outside of the pericardium is commonly referred to as paracardial or mediastinal fat (MAT). The union of these fats is commonly considered as pericardial fat (PAT). Figure 2.1 illustrates also these cardiac fats.

There is some inconsistency in the literature regarding terminology and anatomical definitions, particularly with regards to the PAT. Some researchers consider it separate from the EAT and paracardial fat, while others define it as the combination of both. In this work, we use the more commonly accepted definition of PAT as the union of EAT and paracardial fat.

## 2.2 Coronary Artery Disease

CAD, also known as IHD, is the most prevalent form of heart disease and the leading cause of death worldwide. CAD occurs when the coronary arteries become hardened and narrowed, reducing blood flow to the heart muscle and causing ischemia. This is caused by the buildup of plaque, made up of calcium, fatty lipids, and inflammatory cells, on the inner walls of the coronary arteries. This plaque buildup is also known as atherosclerosis. As a result, the heart muscle does not receive enough blood or oxygen, leading to chest pain (angina) or a heart attack (myocardial infarction) [1]. Figure 2.2 illustrates the occurrence of the CAD.

There are many risk factors that increase the likelihood of developing CAD, including high cholesterol, diabetes, smoking, being overweight, and having a family history of heart disease. In these cases, it is recommended to undergo testing for CAD [1].

The diagnosis of CAD is established through various invasive and non-invasive imaging techniques that detect decreased blood flow to the heart muscle, weakened heart muscle contractions, and other structural changes [9]. Coronary angiography (CAG) is the gold standard method for identifying CAD, as it has both diagnostic and therapeutic capabilities. In summary, during the coronary angiography exam, a catheter is inserted into the patient's blood vessels, which will then inject a special dye called iodinated contrast into the coronary arteries. This allows the coronary vessels to be visualized from different angles in real-time using a specialized X-ray technique called fluoroscopy. If it is necessary to restore blood flow to the heart, Coronary Angioplasty can

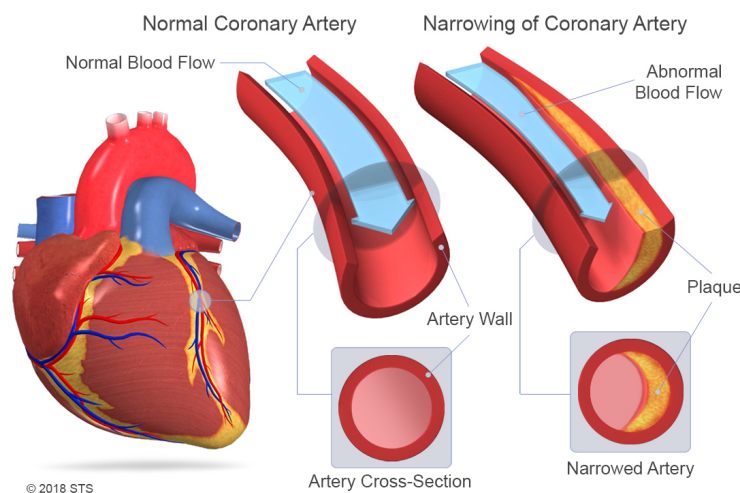


Figure 2.2: Illustration of the mechanism of CAD. Image from [8].

be performed immediately following CAG [10]. Figure 2.3 shows a CAG exam before and after a Coronary Angioplasty. However, due to its invasive nature, the potential for complications associated with the procedure, and the high cost, it is not always the best option for diagnosis.

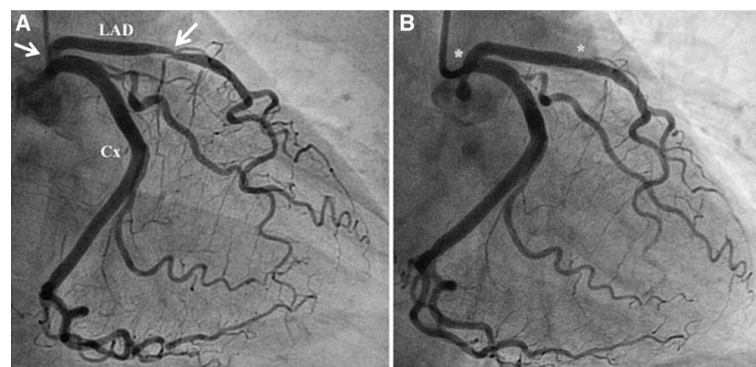


Figure 2.3: CAG before (A) and after (B) Coronary Angioplasty to restore blood flow to the heart. Image from [11].

Fortunately, there are alternative non-invasive methods available that provide diagnostic and prognostic information to help determine the level of risk, plan further testing, and guide treatment decisions. Anatomic tests for CAD include coronary arteries calcium score (CACS), coronary computed tomography angiography (CCTA), and magnetic resonance angiography (MRA), although MRA has limited clinical use.

One of the most widely used non-invasive methods for screening CAD is a calcium scan, which employs CT to detect calcium deposits in the coronary vessels. This test doesn't require the use of contrast agent and it's based on the detection of calcified plaque lesions. A higher score of coronary calcium is an indicator of a higher risk of presence of plaque and future heart attack. However, it's important to note that CAD, particularly in people under 50 years of age, can be present without calcium (non-calcified plaque) and may not be detected by this test, leading to

misdiagnosis in some cases [3]. If the CACS exam indicates a potential risk of plaque buildup, a CCTA may be recommended for further evaluation. Figure 2.4 illustrates the presence of calcium in the CACS exam.

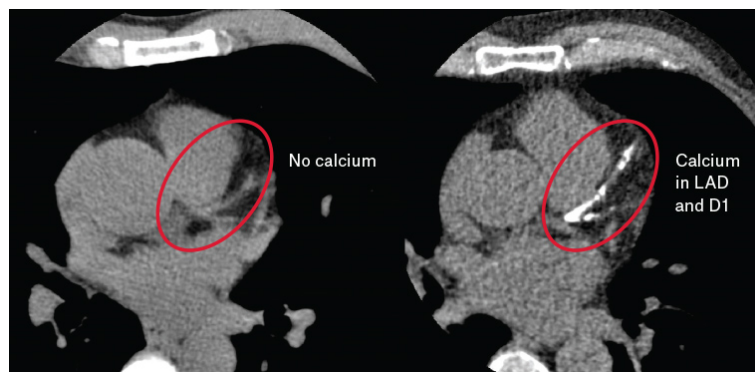


Figure 2.4: CACS exam without (left) and with (right) calcium in the arteries. Image from [12].

CCTA has been shown to be an effective means of assessing patients at intermediate risk for CAD. CCTA has the advantage of being less risky and potentially more cost-effective than CAG while providing similar diagnostic information [13]. In brief, the procedure involves injecting a contrast agent into the patient's bloodstream and then using a CT scanner to take detailed images of the coronary arteries. CCTA allows for the visualization of the coronary arteries, including any blockages, both calcified and non-calcified stenosis, whereas CACS only measures the amount of calcium present in the coronary artery walls. However, this test is not recommended for asymptomatic individuals because of the risks associated with the contrast agent used during the exam, such as allergic reactions and kidney damage, and also the high cost of the test. Figure 2.5 depicts a CCTA exam and the coronary arteries, including the severity of stenosis.

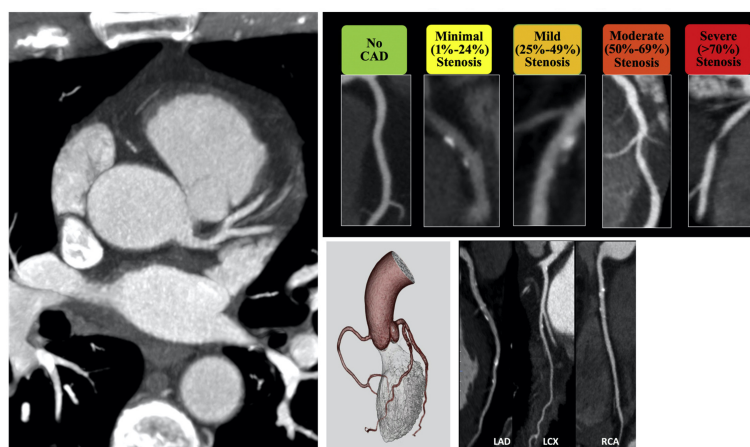


Figure 2.5: CCTA exam with the visualization of the coronary arteries and respective severity of stenosis. Image from [14].

In summary, there are two main non-invasive imaging techniques for screening CAD: CCTA and CACS. CACS is cost-effective and reliable but it can only detect calcified plaques in the

coronary arteries, which is a major limitation as non-calcified plaques are associated with higher mortality rates [15]. On the other hand, CCTA is able to identify both calcified and non-calcified plaques, as well as the severity of coronary stenosis and plaque composition and morphology, but it is more expensive and involves the use of potentially nephrotoxic contrast agents. Therefore, a new predictor of CAD is desirable to bridge the "gap" between CACS and CCTA.

### 2.2.1 The Role of EAT in CAD

EAT has been linked to various CVDs such as CAD, atrial fibrillation, and heart failure. In fact, due to its proximity to the myocardium and coronary arteries, EAT is thought to play a role in the development and progression of CVDs [4]. Early studies indicate that an expansion of this fat depot may be associated with these conditions.

EAT is mostly located in the atrioventricular and interventricular grooves, is primarily composed of adipocytes but also includes nerve cells, inflammatory cells, stromal cells, vascular cells, and immune cells.

EAT shows metabolic activity and has the same microcirculation as the myocardium, which implies the possibility of interactions between them. The EAT has several known functions, including protecting the heart, supporting the coronary arteries, acting as a barrier against pathogens, and regulating energy supply [16]. Unfortunately, EAT can become dysfunctional and contribute to the development of atherosclerosis through various possible mechanisms such as inflammation, an overactive immune response, and oxidative stress, among others. Inflammation is a major feature in patients with CAD, with high concentrations of macrophages and mast cells. EAT's close proximity to the coronary arteries allows its pro-inflammatory proteasome to surround the coronary adventitia and directly enter the coronary lumen via paracrine or vasocrine pathways [4].

EAT's role in coronary atherosclerosis can be used in clinical settings for early diagnosis and risk stratification. Additionally, recent research has shown a strong link between the amount of adipose tissue within the pericardium and the likelihood of future heart disease.

Research has shown that EAT volume and thickness are greater in patients with CAD compared to individuals without the condition. Studies such as Goeller et al. have found that EAT volume is higher in patients with coronary calcium, particularly in the early stages [17]. On the other hand, Eisenberg et al. studied the relationship between EAT and CACS in the EISNER trial and found that both EAT and CACS were independent predictors of Major Adverse Cardiac Events (MACE) in the future. Additionally, they found that EAT alone was a predictor of MACE in patients with no coronary artery calcification (CAC) [18]. Other research supports the idea that the volume of EAT can indicate the early development of atherosclerosis in asymptomatic people, regardless of their obesity [19] or diabetes status [20]. Other studies indicate that EAT can be changed and may serve as a viable therapeutic target [21].



## 2.3 Imaging Techniques for Cardiac Fat Measurement

Given its potential significance and association with CAD, measuring and determining the amount of EAT and PAT in a patient can be a useful tool for risk stratification. However, quantification follows image segmentation.

Imaging techniques play a vital role in contemporary cardiology. EAT can be evaluated by using a variety of equipment, such as CT, Cardiac Magnetic Resonance (CMR), and echocardiography [22].

### 2.3.1 Echocardiography

An echocardiogram uses sound waves to create images of the heart. Standard 2D echocardiography can be used to visualize and measure the thickness of EAT. EAT is typically identified as the area without echoes between the outer wall of the heart muscle and the visceral layer of the pericardium, but it can also appear as an area with echoes when there is inflammation or a large amount of EAT [23]. EAT thickness is typically measured as the distance perpendicular to the right ventricular free wall at end-systole (the end of a heartbeat). Figure 2.6 illustrates the location of EAT and MAT in an echocardiogram.

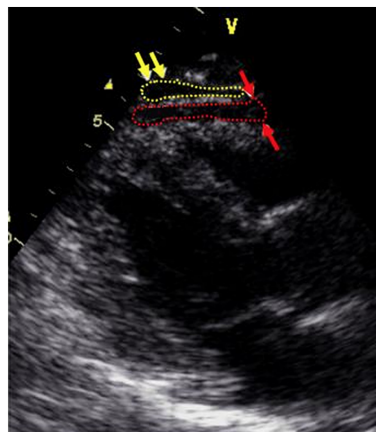


Figure 2.6: Epicardial versus paracardial fat thickness. Paracardial fat (within yellow arrows and yellow dashed shape), epicardial fat (within red arrows and red dashed shape). Image from [23].

There are several benefits to using echocardiography to measure EAT thickness, such as its affordability, availability, and repeatability. However, it also has some limitations. The image quality is inadequate for quantitative analysis (volume and area) and the accuracy of echocardiography is highly dependent on the operator's expertise [22].

### 2.3.2 Cardiac Magnetic Resonance

Cardiac magnetic resonance (CMR) employs a strong magnetic field, radio waves, and computer technology to create detailed images of the heart and surrounding structures. CMR is known for its high spatial resolution and is considered the most reliable method for imaging adipose tissue,



being highly sensitive to fat. However, this sensitivity to fat can make it difficult to visualize the myocardium [24]. Figure 2.7 illustrates EAT and MAT in a CMR.

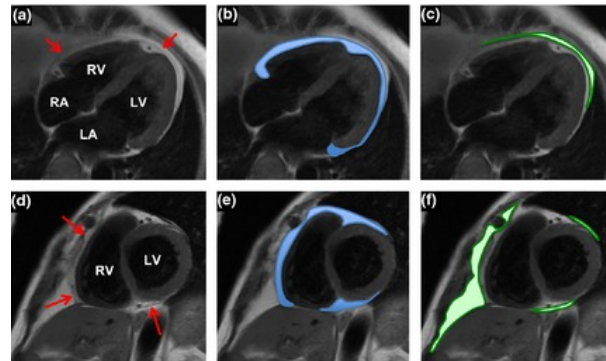


Figure 2.7: The quantification of EAT and MAT by CMR. The pericardium was identified. The EAT and MAT are identified in blue and green, respectively. Image from [25].

MRI allows for the calculation of volume, area, and thickness without the use of radiation or iodinated contrast. However, it comes with some downsides such as high cost, limited availability, long scan times, and difficulty accommodating severely obese individuals [5]. Besides that, MRI can be challenging to use due to difficulties in locating the pericardium in cardiovascular images (i.e. it is hardly visible as a very thin line, often blurred due to partial volume effects).

### 2.3.3 Computed Tomography (CT)

Computed Tomography (CT) is the most frequently used technique for assessing cardiac fat due to its ease of use, ability to create three-dimensional images, high spatial resolution, and capability to quantify X-ray attenuation. Furthermore, the prevalence of CACS scans has led to the availability of large datasets, facilitating studies of EAT using precise and efficient quantification methods [5]. As previously mentioned, there are also CCTA scans which are less commonly used but still studied. Figure 2.8 illustrates the differences in EAT quantification on a standard CT scan and a CCTA.

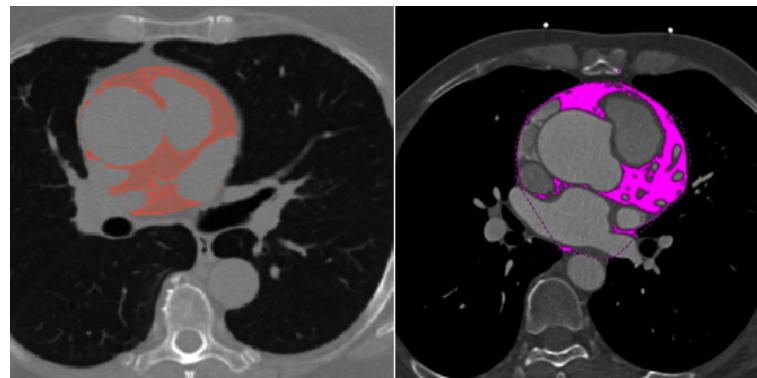


Figure 2.8: Segmented EAT in Non-Contrast (left) and Contrast-Enhanced (right) CT Images, from [18].

In CT imaging, the density of a tissue is represented using the Hounsfield scale, where the attenuation is expressed in Hounsfield Units (HU). Water has a value of zero HU, tissues denser than water have positive HU values, and tissues less dense than water have negative HU values. In general, cardiac adipose tissue attenuation ranges between  $-45$  HU and  $-195$  HU, where a lower negative means higher density [5]. It's important to note that some studies may use slightly different ranges for HU because HU values may vary depend on the manufacturer and specific system of the scanner used, as each uses a unique combination of X-ray sources [26]. Therefore, to improve visibility of cardiac fat and facilitate its segmentation, many experts use this HU range.

One major drawback of CT is the exposure to ionizing radiation, which can be a concern when serial assessments are needed. To minimize radiation exposure, some CT scans may have the heart cropped, which can make it more challenging to quantify the EAT.

## 2.4 Towards Automatic Pericardial Fat Segmentation

Despite the availability of imaging techniques for evaluating cardiac fat, their clinical application is limited due to the difficulty in precisely measuring their volume.

The manual segmentation and quantification of EAT and PAT can be a laborious and time-consuming task, and is often subject to inter-observer variability as a result of the complex distribution of these fats around the heart [27]. The process of manually segmenting EAT from CT scans can take several minutes to an hour, depending on the number of slices in a scan and the software being used [28]. Thus, computer-aided tools, which can deal with these issues, are essential to enhance the repeatability of the results and also to reduce time.

The main challenge in this process is the accurate identification of the pericardium, which is the boundary between EAT and other external fats. In CAC scans, the pericardium can be difficult to discern as it is estimated to be less than 2 mm thick. The fact that the heart is a constantly moving organ only exacerbates this problem. The thin pericardium in a CT axial slice can be seen in Figure 2.9.

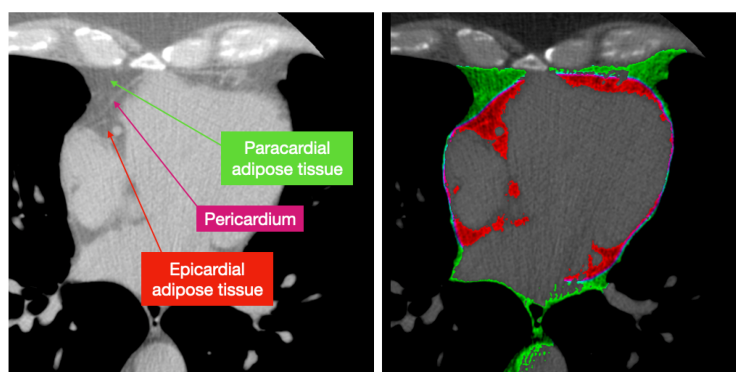


Figure 2.9: Illustration of the pericardium, EAT and paracardial fat in a CT image. Images from [29].

## Chapter 3

# Automatic Pericardial Fat Segmentation on Non-contrast CT Images

This chapter provides an overview of different methods and techniques used for fully-automatic segmentation of EAT, PAT and pericardium from non-contrast CT images. Additionally, a section is dedicated to explaining evaluation metrics used for segmentation and another section discusses the conclusions and limitations of the various approaches discussed in the chapter.

### 3.1 State of the Art

This section will discuss the techniques and methods used for segmenting EAT, PAT, and the pericardium on non-contrast CT images. It is structured into three sub-sections, each focusing on a particular type of approach: methods based on atlases, machine learning-based techniques, and deep learning-based techniques. Table 3.1 displays the studies covered in this review, along with their respective results.

#### 3.1.1 Atlas-based Techniques

The first studies that aimed to fully segment the EAT and pericardium employed atlas-based approaches. Atlas-based segmentation involves using prior knowledge and a reference image (called an atlas) in which structures have already been segmented. The goal is to find the best way to transform the atlas image to fit the test image. This is achieved using a registration algorithm. During the registration process, the input image is adjusted to align with the fixed target image. The resulting image, known as the Registered image, displays both the input and target images overlaid, with the input image now transformed as a result of the registration algorithm. Figure 3.1 illustrates this process. The atlas plays a crucial role in the segmentation process because the regions of interest (ROI) in the atlas and test images should be similar in order to establish

Authors	Structure	N	Method	Performance (DSC)
Dey et al. 2010 [30]	P	20	Atlas	0.89
Ding et al. 2015 [31]	EAT	50	Atlas + GAC	0.92
Rodrigues et al. 2016 [26]	EAT, PAT	20	ML	0.981; 0.976
Rodrigues et al. 2017 [32]	P	20	EM	-
Commandeur et al. 2018 [33]	EAT, PAT	250	DL: 2 CNN	0.823; 0.905
Commandeur et al. 2019 [34]	EAT	614	DL: CNN	0.871
Priya et al. 2019 [35]	EAT, MAT	20	ML	0.987; 0.982
Li et al. 2019 [36]	P	53	DL: U-Net	-
Kazemi et al. 2020 [37]	EAT, MAT	20	ML	0.978; 0.984
Zhang et al. 2020 [38]	P, EAT	20	DL: Dual U-Net	0.950; 0.911
Hoori et al. 2021 [39]	EAT	89	DL: CNN	0.885
Molnar et al. 2021 [40]	EAT	411	DL: 2 3D-CNN	0.90
Bencevic et al. 2021 [41]	EAT	20	DL: Semi-3D U-Net	0.857
Siriapisith et al. 2021 [42]	EAT	220	DL: 3D U-Net	0.901
Qu et al. 2022 [28]	EAT	103	DL: CNN + MRA-U-Net	0.883

Table 3.1: Summary of fully-automatic algorithms for segmentation of pericardial fats from non-contrast CT scans and respective results. DL= deep learning; EM=elliptical modelling; GAC= geodesic active contours; ML= machine learning; N= Number of patients; P= pericardium.

valid correspondences. Using a general anatomy patient atlas can improve the accuracy of the segmentation results [43].

In 2010, Dey and colleagues published one of the earliest works on the topic [30]. The authors proposed an automated algorithm for segmenting the heart and pericardium from non-contrast CT scans that uses a multiple co-registered segmented atlas. The atlas was created by manually segmenting the whole heart (cardiac region and pericardium) in 8 CT scans. The algorithm then segments new CT scans by registering them to the atlas and calculating a segmented volume mask from all co-registered atlas. The CT scans that were included in the atlas were aligned with a singular reference image through an iterative linear registration technique that involved translation, rotation, and linear scaling. Then, a non-rigid transformation was applied using deformation fields that were generated by the iterative demons algorithm. The final transformations were applied to both the images and volume masks in order to ensure that all atlas datasets and their corresponding masks were in the same spatial coordinates. Once the atlas has been created, the segmentation process can be carried out using a similar approach to the one used for atlas creation. However, the final masks for each slice will be generated by combining the masks from all the atlases using a weighted decision scheme.

Ding et al. employed a similar approach but improved the segmentation of the pericardium using a geodesic active contours deformation method [31]. Figure 3.2 demonstrates the rationale about the geodesic active contour. In previous studies, such as the one conducted by Dey et al, the pericardium is directly segmented after the target image registration and the mask are generated by combining the masks from all the atlases using a weighted decision scheme. However, Ding et al, applied this process only to initialize a global location of the pericardium, creating an atlas from 10 patient scans. Afterwards, they used a knowledge-based thin membrane detection method

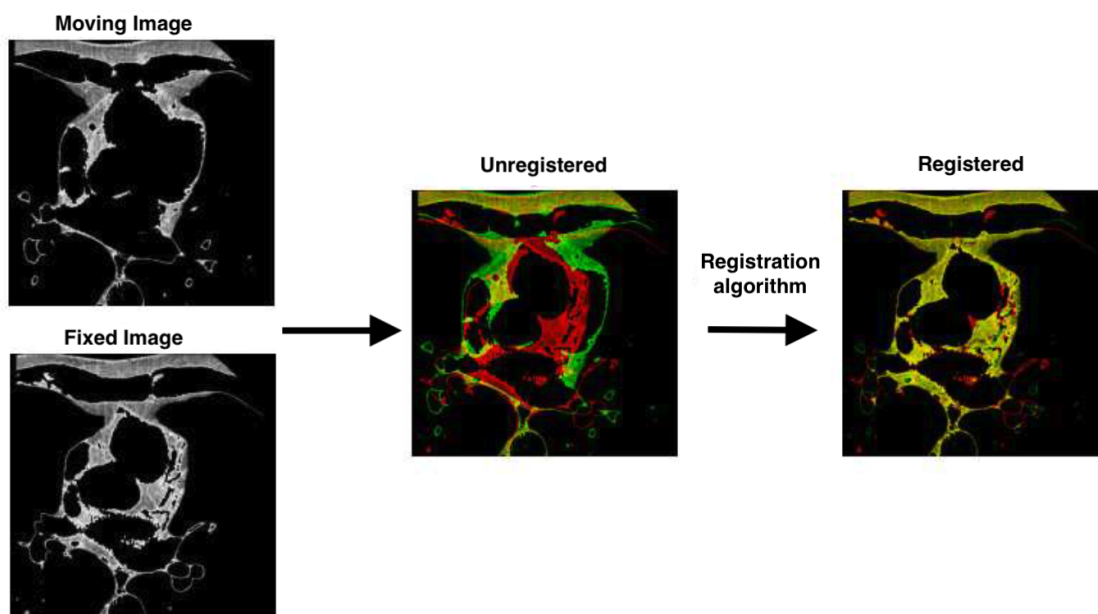


Figure 3.1: Illustration of the registration process, adapted from [29].

to identify the visible part of the pericardium, and designed a feature detector that specifically responds to the pericardium based on previous research [44]. After the pericardium has been segmented, or isolated, from the rest of the image, a threshold of -190 to -30 HU is applied to fat-containing voxels within the pericardial sac.

Despite achieving satisfactory results, atlas-based techniques have limitations that need to be taken into account, such as ambiguity in atlas construction and poor performance due to inaccuracies in registration. These methods may not be suitable for cases where there is high variability in morphology. Additionally, the use of atlases for detecting regions of interest assumes that all cases being analyzed can be mapped to a common model, which can lead to inaccuracies in fat volume quantification when the patient's morphology differs from the atlas model.

### 3.1.2 Machine Learning Techniques

Machine learning involves the use of algorithms that allow computers to learn and make predictions, and improve over time by analyzing large sets of data [45]. It is divided into three categories: supervised, unsupervised, and reinforcement learning. In supervised learning, the algorithm is given labeled data by human experts, while in unsupervised learning, the computer finds patterns in the data without any prior labels. In reinforcement learning, the computer learns through interactions with its environment and receiving feedback.

One of the first studies assessing machine learning techniques for the segmentation of the fat surrounding the heart compartments was reported by Rodrigues et al. in 2016 [26]. The method includes two main steps: intersubject registration and classification. The registration step involves rescaling each slice to a common proportion, but this process can result in part of the heart being cropped out. To overcome this issue, the registration is achieved using a combination of two

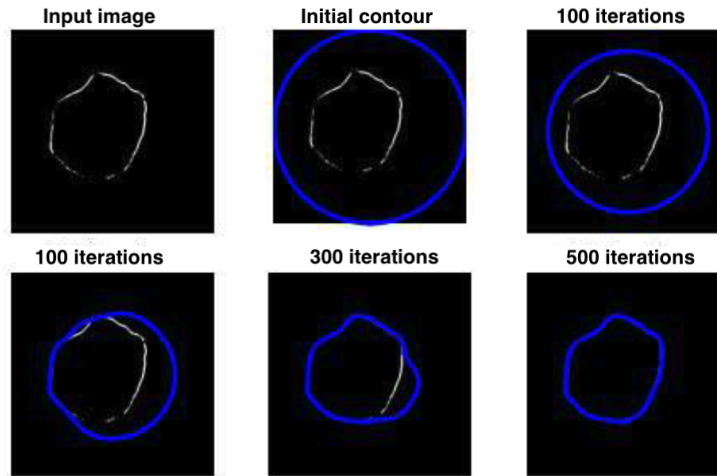


Figure 3.2: An illustration of using active contours to fill in missing sections of a pericardium line, depicted in white. The contour is displayed in blue, starting as a perfect circle around the image. The contour is gradually altered in each iteration to better conform to the image. Note that this is only the rationale about geodesic active contour and not the exact method used by Ding et al. Image adapted from [29].

techniques: a landmark approach and an atlas approach. The landmark chosen is the retrosternal area, located on the back of the sternum, which does not vary greatly among patients and is always visible within the boundaries of the CT image. Therefore, an atlas was created by manually aligning 10 randomly selected slices by these retrosternal areas, as illustrated in the Figure 3.3. For each target image, a similarity score is computed between the fixed image (target image) and the atlas (moving image). Once the parameters for the translation are determined, the entire set of images for the same patient is transformed according to the scaling and translation parameters found. After the registration, a classification algorithm was performed to achieve EAT, MAT and pericardium segmentation. In the first step, a feature extraction process was conducted and features were selected according to three main categories: pixel information (gray value, axis coordinates), related to the image or to a neighborhood window (arithmetic mean, geometric moments, etc.) and related to the already extracted data (co-occurrence matrix, run percentage, etc.). In the second step, a predictive model was trained using known classification algorithms such as Support Vector Machine (SVM), Sequential Minimal Optimization (SMO), Naive Bayes, Radial Basis Function Network (RBFNetwork), Random Trees, and others. The authors found that the Random Forest algorithm performed the best. It is worth noting that the dataset used in this study was made publicly available and can be accessed online.

In 2017, Rodrigues and colleagues introduced a new technique to decrease the significant processing time required for registration [32]. With the previous method, it takes around 1.8 hours to automatically segment the cardiac fats of a patient. To address this issue, they proposed using elliptical modeling to detect the contours of the pericardium. By identifying the pericardium in one of the images, it becomes possible to calculate the overall volume of EAT without having to segment each individual image. The fitting 5 parameters for the ellipse (orientation, center co-



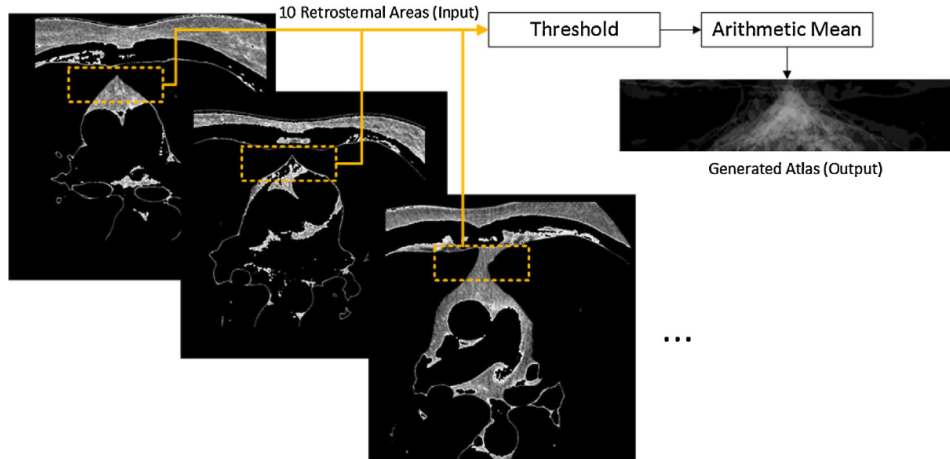


Figure 3.3: Creation of the atlas by Rodrigues et al. [26]. Image from [26].

ordinates and major/minor axis) are determined using a Genetic Algorithm (GA) [46]. GA are a set of computational optimization methods that are based on the principles of Darwin's theory of evolution and natural selection. They work by evolving a population of individuals over time, by preserving the desirable characteristics of each individual as determined by a fitness function. These characteristics can be represented by parameters that are specific to a problem being solved. Therefore, this GA was executed using a population of 20 individuals, each with 5 characteristics (ellipse parameters) in order to maximize a specified fitness function (determined by evaluation of the ground-truth images). If the GA is given adequate time to run, through multiple iterations, the outcome will be satisfactory and close to the optimal solution. The previously proposed framework [26] was improved in terms of accuracy through the reduction of false positive voxels outside the pericardium contour and false negative EAT voxels within the pericardium contour, resulting in an overall improvement in performance. Furthermore, by fitting an ellipse to represent the pericardium on a select number of slices from a CT scan, the pericardium contour can be interpreted on the remaining slices, resulting in a reduction of processing time.

In 2019, Priya and colleagues proposed a new method for segmenting EAT and MAT using a combination of a feature extraction technique based on Grey level co-occurrence matrix (GLCM) and an Optimal Neural Network (ONN) for pixel classification [35]. The resulting fat images were then processed using a Fruit Fly-based Modified Region Growing Algorithm (MRG) for segmenting cardiac fat. The proposed methodology is demonstrate in the Figure 3.4. The method begins by eliminating any irregularities and noise in the input CT DICOM images during preprocessing, and aligning the images through image registration (like Rodrigues et al. [26]). GLCM is then used to extract texture-related features [47]. A GLCM is a matrix that compares the gray levels of each pixel with the gray levels of its surrounding pixels at a specific distance and direction. From this matrix, they extracted several features such as Energy, Entropy, Contrast, Correlation, and Homogeneity. Then an ONN with a Grey Wolf Optimizer (GWO) was used to classify the pixels with only two classes (fat and non-fat). The GWO is an optimization algorithm based on the hi-

erarchy and hunting (optimization) mechanism of grey wolves in a pack, with the alpha being the most important and the omega being the least important [48]. Finally a Fruit-fly based Modified region growing algorithm is applied to segment fat accurately. A region growing algorithm (RG) is a popular segmentation algorithm particularly useful in images with low contrast and noise. The process starts by selecting a seed point in the image, and then iteratively adding neighboring pixels to the region if they meet a certain similarity criterion. This process continues until all the pixels in the region are connected and have similar properties [43]. Then, the authors replaced the pixels of the output of the ONN by the respective gray pixels, calling the result as  $I_p$ . The similarity criterion for the RG is a threshold of the distance among the pixels between the  $I_p$  and the texture image ( $I_t$ ). The Fruit Fly algorithm [49] is used to optimize the two threshold values (for EAT and MAT) in the algorithm, allowing for dynamic adjustments in the process.

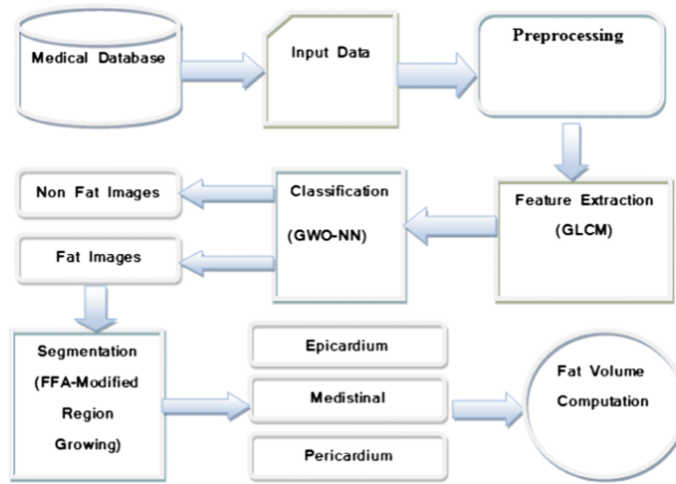


Figure 3.4: Workflow for the proposed methodology of Priya et al [35]. Image from [35].

In 2020, Kazemi et al. proposed an approach that builds upon the work of Rodrigues et al [37]. Both studies employed similar methodology and used the same dataset. Kazemi et al. claims that their approach extracts more robust features from textures, which is the main difference from the work of Rodrigues et al.. The approach included a pre-processing step where the images were processed using contrast limited adaptive histogram equalization (CLAHE) to enhance the contrast and improve the image quality for feature extraction and subsequent segmentation. CLAHE function works by dividing the image into a series of equal blocks and performing an equalization histogram separately in each segment using linear interpolation. In the feature extraction step, the authors extracted 22 features for each pixel. They used a combination of two commonly used methods for texture analysis: a Gabor Filter Bank and a gray-level co-occurrence matrix (GLCM). These methods are used to detect edges, textures, and orientations properties. From these methods, the authors extracted features such as contrast, correlation, energy, inertia, inverse difference, entropy, and homogeneity besides the features related to pixel information previously mentioned in Rodrigues et al. [26]. Finally, they applied dimension reduction using Fisher discriminant analysis (FDA). FDA is a method that increases the discriminant between members of different classes and



reduces the discriminant between members of the same class. The goal of this step is to reduce the dimensionality of the feature space in order to improve the performance of the classification algorithm. The best result was achieved for EAT segmentation with the TreeBagger classifier.

### 3.1.3 Deep Learning Techniques

Deep learning is a subset of machine learning algorithms that utilizes multiple hidden layers to learn and understand data [45]. The different models include 2D convolutional neural networks (CNNs) and 3D models, with CNNs being the most commonly used in medical imaging analysis. Deep learning differs from traditional machine learning in that it uses multiple processing layers to learn data representation with multiple levels of abstraction. It also differs from artificial neural networks in terms of the number of hidden layers, connections and its ability to learn significant abstractions of inputs.

These methods have become increasingly popular in recent years, resulting in improved performance for various tasks such as computer vision, speech recognition, and natural language processing [33]. The development of CNNs began three decades ago, inspired by the animal visual process. These networks are composed of convolution filters that act as feature extractors and are used in a wide range of applications, including medical imaging. CNNs have been used for tasks such as cancer detection, lung nodule detection, and heart segmentation in MRI. They have also been widely used in cardiac imaging for tasks such as ventricle and heart segmentation. Additionally, previous studies have described the use of CNNs for automatic coronary calcium scoring from CT datasets.

In 2018, one of the first applications of CNNs to EAT segmentation was proposed by Commandeur et al. [33]. The authors employed 2 CNNs to perform a multi-task framework. The first CNN ( $Net_1$ ) was used to accomplish three different tasks. The first task was to classify whether or not a certain input slice is within the limits of the heart by incorporating a fully connected dense layer. After, the Hidden Convolutional Units (HCU) were upsampled to perform two segmentation tasks: segmentation of pericardial fat, and another to differentiate EAT from MAT. The use of a combination of fully connected dense layers and an upsampling block to perform both classification and segmentation tasks was one of the contributions of their approach. The second CNN,  $Net_2$ , was used to detect the contour of the pericardium. The epicardial mask from  $Net_1$  was transformed into cylindrical coordinates and provided as input to  $Net_2$ . Again, by upsampling the Hidden Convolutional Units (HCU) from  $Net_2$ , the output of the  $Net_2$  was a probability map of the pericardial sac. Note that the authors considered the epicardial mask as the entirety of the interior region of the pericardium. Therefore, to segment the EAT within the "epicardial mask", a post-processing step was applied using the standard adipose tissue HU range [-190, -30] HU. The training process for each CNN involved using a cross-entropy loss function and a stochastic gradient descent algorithm to optimize the network's parameters.

In 2019, Commandeur et al. conducted one of the first multicenter studies on EAT segmentation using a large dataset from four different cohorts [34]. Instead of 2 CNNs, a single CNN network was implemented to perform multiple tasks through a computational graph. The network

first identifies the upper and lower limits of the heart, and then segments the pericardium and its interior region. To accomplish this, the CNN is provided with three consecutive axial slices: the one to be classified and segmented ( $k$ ), as well as the previous ( $k - 1$ ) and next sections ( $k + 1$ ). The architecture of the network is presented in Figure 3.5. In summary, the network is composed by six blocks, each containing two convolutional ( $3 \times 3$ ) layers. It is noteworthy that the authors also used a residual learning for each block to improve and optimize the training [50]. Similarly to the previous work [33], at the end of the five residual blocks, two dense layers were added to determine if the input section intersects the heart. For the second task, which is to segment the pericardium, five transposed convolutions were performed to upsample the image to the original size ( $512 \times 512$ ), one for each block. After concatenation, the pericardium is segmented. To improve the model, data augmentation was applied by adding random affine transforms and Gaussian noise. The Adaptive Moment Estimation Algorithm (Adam) optimizer was used with a combined loss of binary cross-entropy and DSC loss.

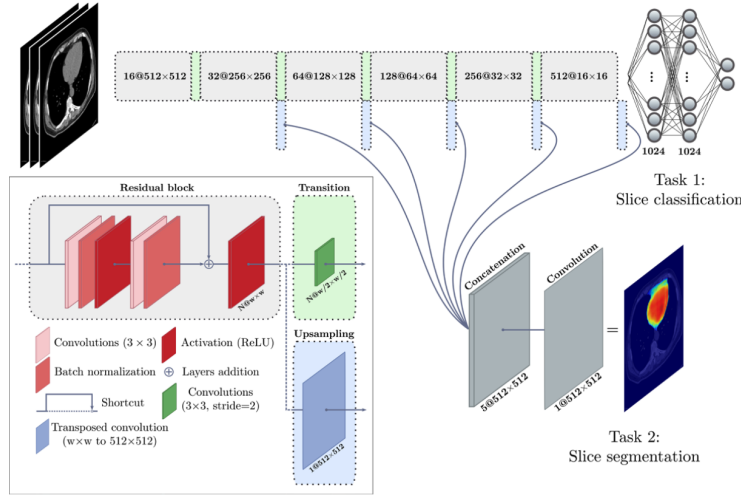


Figure 3.5: CNN architecture proposed by Commandeur et al. [34]. Image from the same source.

Li et al. proposed a method to detect the pericardium contour using a U-Net architecture [36]. The U-Net is a deep network architecture introduced by Ronneberger et al. in 2015 [51]. This architecture is based on convolutional networks for biomedical image segmentation. The U-Net framework has a unique structure that includes a U-shaped architecture and skip connections, shown in Figure 4.7. The U-shape is symmetric and divided into two sections: the contracting path on the left and the expansive path on the right. The contracting path is made up of multiple layers of  $3 \times 3$  convolutions, each followed by a rectified linear unit (ReLU) and a  $2 \times 2$  max pooling for downsampling. The expansive path uses transposed 2D convolutional layers to increase the image size back to its original size. Each step in the expansive path includes an upsampling of the feature map and concatenation with the corresponding channels from the contracting path, providing more detailed features for image segmentation.

The dataset used for training in the study of Li et al. was the same as that used by Rodrigues et al. [26], but instead of using an atlas-based approach to align and standardize all the images,

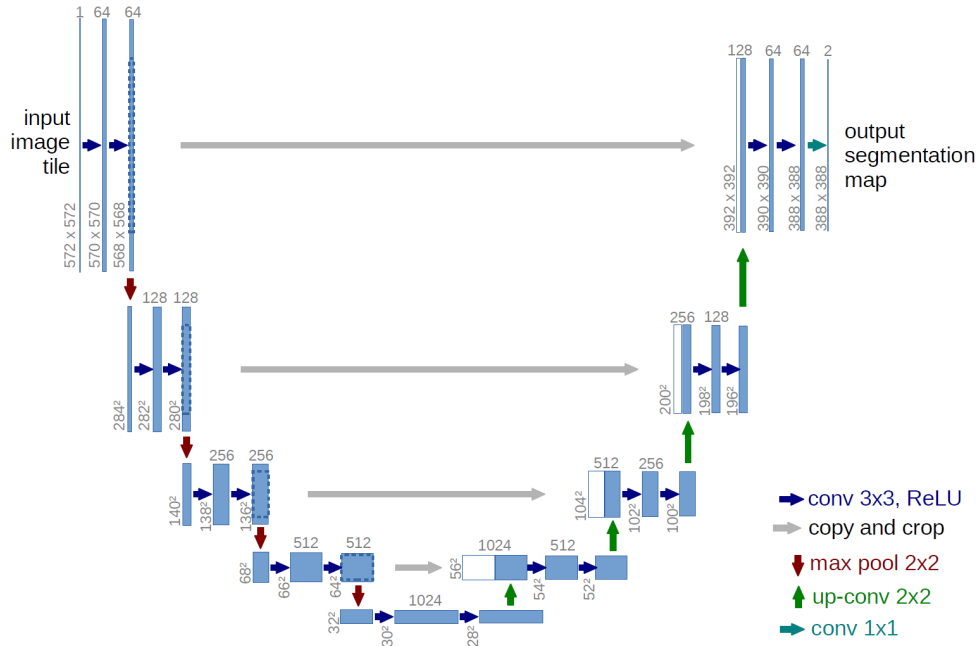


Figure 3.6: U-Net architecture presented by Ronneberger et al. [51].

they employed a simpler method. Before detecting the pericardium, Li et al. developed a ratio-based algorithm for locating the heart by consulting radiologists. The algorithm aims to identify an approximate region of the heart by determining the upper and lower boundaries of the abdomen and using a specific ratio to calculate the heart's location. Afterwards, they modified the U-Net by adding a pyramid pooling structure after three rounds of convolution. The Figure 3.7 shows the architecture of the pyramid pooling module. The pyramid pooling module enables the integration of information from different scales in the feature maps. This is done by applying pooling kernels of varying sizes on the original feature map. Then, a  $1 \times 1$  convolution and upsampling is applied to the pooled feature maps, and the resulting information is concatenated with the original feature map. The method used to find the pericardium's boundaries is to segment the EAT and the MAT separately, and then calculate the overlap area between the two fats. Two separate models were trained, one to segment the EAT, and another to segment the MAT. Focal Loss was used to handle imbalanced data during the training phase, and data augmentation techniques such as cropping, rotation, and scaling were applied. The Adam optimization algorithm was employed. The model was trained using the Cardiac Fat Dataset [26] and tested on an external dataset of chest CT images from 33 patients.

In 2020, Zhang et al. claimed to be the first to use a dual U-Net approach to segment the EAT [38]. The first U-Net is used to segment the pericardium and remove interference from surrounding organs during the segmentation of EAT. However, the pericardium segmented by the first network can have noise, irregularities or holes. To address this, the authors applied a morphological layer to obtain the ideal region inside the pericardium (IRP). The morphological layer consists of three main steps: firstly, using a method called morphological reconstruction

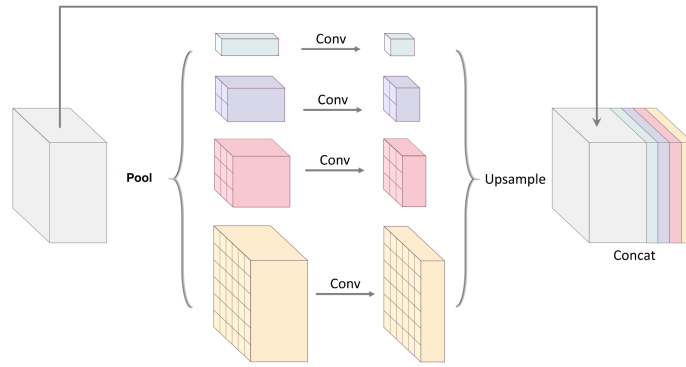


Figure 3.7: Architecture of a pyramid pool module from [52].

to fill the holes in the segmented IRP, secondly, using an open operation to remove small noisy elements outside the pericardium and lastly, using a close operation to smooth the boundary and internal noise of the segmented IRP. To analyze and segment the EAT from the interior region of the IRP, a second network based on a U-Net architecture was employed. However, this U-Net has only three encoders and three decoders of the same level. Both networks were trained and tested using the publicly available data of Rodrigues et al. [26] with the Adam optimizer. The authors note that one limitation of their approach is that the use of two U-Nets and a morphological layer together can be computationally expensive and requires more time.

In 2021, Bencevic et al. introduced a new technique for segmenting EAT called Semi-3D Neural Network, which utilizes a single 2D U-Net architecture, but with an added depth channel [41]. This approach is similar to that proposed by Zhang et al, but instead of using two U-Nets, only one U-Net is employed and the input consists of two channels: the slice image and the depth information, which is normalized between 0 and 1. The method involves first segmenting the interior region of the pericardium and then applying a threshold to segment the EAT. They used the same dataset as Rodrigues et al. which includes CT images with a range of  $[-200, -30]$  HU for adipose tissue. The images were manually segmented to create a closed pericardium region as the ground truth, using the contour of the labeled data from the original dataset. The images were also resized from  $512 \times 512$  to  $128 \times 128$ . Data augmentation techniques such as horizontal flipping, scaling, and translation were used during training, which was done using Dice loss and Adam optimizer.

In the study by Siriapisith et al., a 3D deep learning approach was presented for EAT segmentation using a modified U-Net architecture [42]. This approach includes the addition of an attention gate (AG) module and a deep supervision (DSV) module at the end of the network, as can be seen in Figure 3.8. The AG focuses on the relevant regions of the image by suppressing irrelevant regions [53]. This is achieved by connecting each layer of the encoding with the decoding path, allowing the feature map from the encoding path to be aligned with the feature map from the decoding path. This alignment allows, after an element-wise-sum, the focus on important features, as the relevant weights are summed together. The DSV module was added at the final stage of the network in order to generate multiple segmentation maps at different resolution levels, which

were then combined together, increasing the number of learned features. This approach differs from the standard approach, where supervision is only provided at the output layer. The losses from this secondary segmentation map are weighted and added to the final loss function, which effectively improves the performance of the model. The model was trained using mean squared error as the loss function and Root Mean Square propagation (RMSprop) as the optimizer. The authors concluded that the incorporation of AG and DSV modules improved the performance of the model.

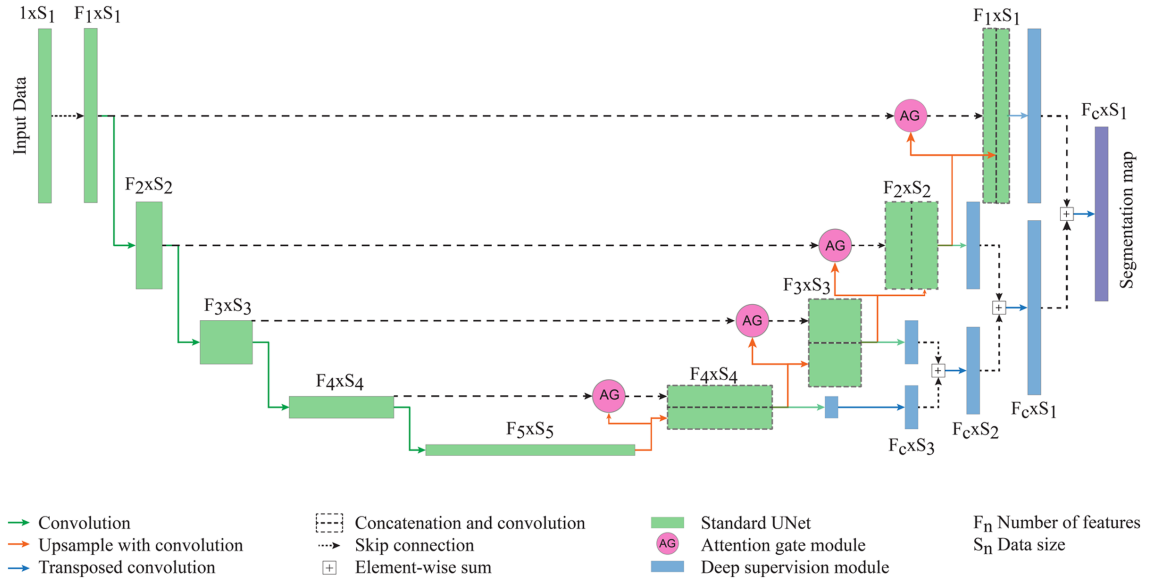


Figure 3.8: U-Net architecture proposed by Siriapisith et al. [42] with the AG and DSV modules implemented in the standard U-Net.

Hoori et al. developed an alternative approach to the work of Commandeur et al. due to errors they found associated with the top and bottom slices [39]. The authors of the study segmented the interior region of the pericardial sac using a CNN and then applied the standard thresholding  $[-190, -30]$  HU to determine the EAT. A diagram of their methodology is presented in Figure 3.9. Their approach involves feeding the CNN with the slice to be segmented ( $k$ ) and the two next slices ( $k + 1$  and  $k + 2$ ). Additionally, they added a pre-processing step to improve the contrast of the images and help the network to focus on features more related to the pericardium by applying a HU-attention-window with a window/level of 350/40 HU. Furthermore, they divided the heart slices into two halves, with the lower half being arranged in a bottom-to-middle sequence and the upper half in a top-to-middle sequence. This method ensures that the curvature of the sac is increasing and that similar images are presented to the network during training and testing. The authors applied transfer learning on the DeepLab-v3-plus network, pre-trained on the ImageNet dataset, which uses Resnet-18 as its backbone. This deep network is specifically designed for semantic segmentation tasks and features important architectural components such as the backbone network, Atrous convolution, Atrous Spatial Pyramid Pooling (ASPP) network, and a decoder section. The model was trained with the Dice loss and also applied data augmentation

techniques, with Gaussian blurring being the most effective. Hoori et al. demonstrated that all these modifications have a positive impact.

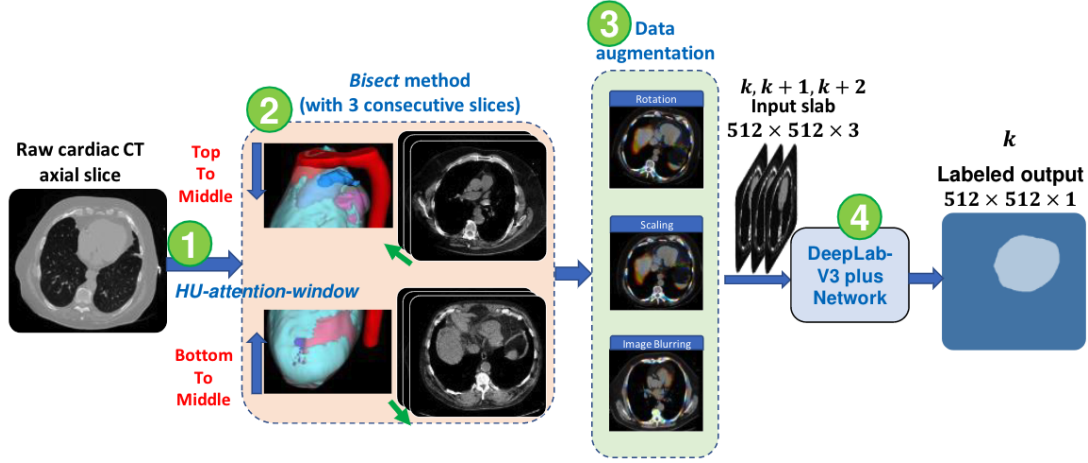
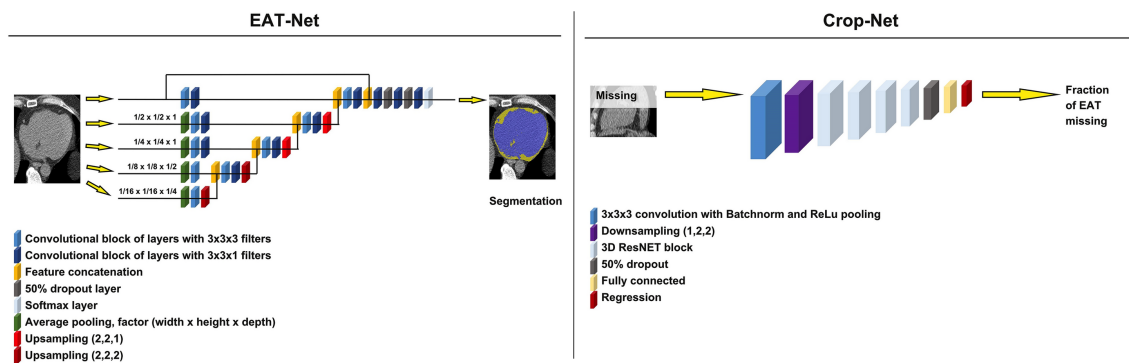


Figure 3.9: The full structure of the automated EAT segmentation training process and the preprocessing steps performed by Hoori et al [39]. Image from [39].

Molnar et al. proposed two 3D CNN models which work in series to segment and estimate the volume of the EAT [40]. The first one named "EAT-Net" is designed to segment the EAT voxels within the pericardium and can be used to calculate the EAT volume and EAT attenuation. The second model "Crop-Net" inspired by ResNet18 is aimed to estimate any missing EAT volume (EATV) when the heart is not fully captured in the image set. Figure 3.10 shows the architecture of the 2 CNNs. They first trained the EAT-Net for only two classes (pericardium and non-pericardium). Then, a new class was introduced: non-adipose tissue, which aims to identify areas that do not belong to the EAT within the pericardium. The authors utilized data augmentation techniques such as rotation and scaling throughout the process, and employed a categorical cross-entropy loss function and Adam optimizer. In the final step of prediction for the output of EAT-Net, a threshold of  $[-190, -30]$  HU was applied to segment only the EAT and discard any pixels classified as non-adipose tissue. The Crop-Net was trained using 866 CT scans that had complete representation and accurate segmentation of the pericardium. Random slices were cropped from these scans and the Crop-Net was trained to predict the missing fraction of EATV. The authors claim that this model is robust and can be applied to larger populations with incomplete and noisy images.

Most recently, Qu et al. were the first to propose a combination between a traditional 2D CNN (T-CNN) and a 2D multiscale residual channel attention U-Net (MRA-U-Net) to segment the EAT [28]. The T-CNN is used to identify the upper and lower limits of the EAT to reduce the oversegmentation. Afterwards, the MRA-U-Net is employed to segment the EAT on the selected slices. Before segmentation, the authors preprocessed the images by setting a threshold of HU to  $[-210, -10]$  and normalizing the pixels using max-min normalization, resulting in a range of  $[0,1]$ . The T-CNN is responsible for selecting the slices, which are then readjusted to a HU range of  $[-100, 125]$  and passed to the T-CNN for classification, determining whether or not the





slice belongs to the heart. The slices belonging to the heart were fed to the MRA-U-Net. The MRA-U-Net architecture is illustrated in Figure 3.11. The MRA-U-Net have a residual channel attention mechanism [54]. This module analyse the information between feature maps and through convolutional layers and global average pooling, the weights are readjusted. By multiplying the feature maps with the newly adjusted weights, the network is able to selectively pay attention to essential features while disregarding unimportant ones. Additionally to address limitations of U-Net, such as the disconnection of low-level and high-level features, and its inability to analyze information across different scales, the authors introduced a multiscale concatenate module (MSC) that includes atrous spatial pyramid pooling (ASPP) [55]. ASPP utilizes atrous convolution to extract features at multiple scales without compromising image information. The authors trained the T-CNN and MRA-U-Net using binary cross-entropy loss and Dice loss respectively, with Adam optimizer. They also claim that their model reduces the time and effort required for EAT segmentation compared to 3D and cascaded CNNs, taking only 2.2 seconds.

Figure 3.11: MRA-U-Net architecture proposed by Qu et al. [28].

## 3.2 Evaluation Metrics

In this section, the most common evaluation metrics used in semantic segmentation will be addressed.

### 3.2.1 Accuracy

The first well-known metric used in the study is Accuracy, which measures the percentage of correctly classified pixels in the image. This metric is easy to understand and is represented mathematically as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (3.1)$$

with TP, TN, FP and FN being the true and false positives and negatives, respectively. However, this metric may not be the most appropriate for evaluating segmentation, particularly for imbalanced classes.

### 3.2.2 Precision

Precision is a metric which measures the accuracy on predicting a specific class by quantify how many predicted pixels are actually true. This metric is mathematically defined as:

$$Precision = \frac{TP}{TP + FP}, \quad (3.2)$$

with TP, TN, FP and FN being the true and false positives and negatives, respectively. The Precision may be useful to evaluate if a model can actually be reliable when its prediction is positive for a certain class.

### 3.2.3 Recall

Recall, also known as sensitivity or true positive rate, quantifies the model's ability to correctly identify positive instances out of the total number of the actual positive instances present. Recall is represented mathematically as:

$$Recall = \frac{TP}{TP + FN}, \quad (3.3)$$

with TP, TN, FP and FN being the true and false positives and negatives, respectively. A high recall suggests that the model is accurate to predict a certain class.

### 3.2.4 Dice Similarity Coefficient (DSC)

The most popular and used metric to evaluate the performance of a segmentation is the Dice similarity coefficient (DSC) [56], defined as:



$$DSC = \frac{2TP}{2TP + FN + FP} \quad (3.4)$$

Simply put, the DSC is two times the overlapping area divided by the total number of pixels in both images. The value of DSC ranges from 0, indicating no spatial overlap between the two segmentations, to 1, indicating complete overlap.

### 3.2.5 Hausdorff Distance (HD)

Hausdorff distance (HD) measures the maximum segmentation error in terms of distance between two sets of points [57]. The formula for HD is:

$$HD(X, Y) = \max \left\{ \max_{x \in X} \min_{y \in Y} d(x, y), \max_{y \in Y} \min_{x \in X} d(x, y) \right\}, \quad (3.5)$$

where  $X$  and  $Y$  are the boundaries of predicted segmentation and the ground truth, and  $d(x, y)$  is the distance function between two surface points  $x$  and  $y$ . The HD value ranges from 0 to infinity, with 0 indicating a perfect match. This metric can be particularly useful when evaluating the performance of the segmentation in combination with other metrics such as IOU or DSC.

### 3.2.6 Mean Absolute Distance (MAD)

MAD, or Mean Absolute Distance, is a measure of segmentation error that is similar to HD. Instead of using the maximum error as HD does, MAD calculates the mean of the absolute errors. The formula for MAD is:

$$MAD(X, Y) = \frac{1}{2} \left( \frac{1}{N_X} \sum_{x \in X} \min_{y \in Y} d(x, y) + \frac{1}{N_Y} \sum_{y \in Y} \min_{x \in X} d(x, y) \right), \quad (3.6)$$

where definitions of  $X$  and  $Y$  remain the same as in (3.5),  $N_x$  and  $N_y$  represent the total number of surface points in  $X$  and  $Y$ , respectively.

### 3.2.7 Pearson Correlation Coefficient (PCC)

The Pearson correlation method is a widely used method for determining the relationship between numerical variables [58]. It assigns a value between -1 and 1, with 0 indicating no correlation, 1 indicating a strong positive correlation, and -1 indicating a strong negative correlation. A positive correlation means that as one variable increases, the other also increases, while a negative correlation means that as one variable increases, the other decreases. This metric is often used to evaluate the correlation between the volume segmented by an expert and the predicted one in medical imaging segmentation tasks. This can be determined using the following formula:

$$PCC = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.7)$$

where  $x_i$  and  $y_i$  represent the estimated and actual values of sample  $i$  in the dataset, respectively,  $\bar{x}$  and  $\bar{y}$  are the averages of the estimated and actual values, and  $n$  is the total number of samples.

### 3.3 Conclusion

This chapter conducts a thorough examination of fully-automatic methods for segmenting cardiac fat, specifically EAT, on non-contrast CT images.

It begins by discussing atlas-based algorithms which are more older articles and traditional approaches. Although atlas-based methods offer good segmentation results and take into account the spatial relationships of different structures, their main limitation is that their performance is heavily dependent on the selection of the atlas. The atlas chosen must be a highly representative CT scan, with minimal noise and an accurate manual segmentation by an expert. However, due to the diversity of human anatomy, this can be a challenge. To address this issue, many approaches have employed the use of multi-atlas, which is a combination of multiple atlases. This way, the segmentation of an input image can be achieved by combining the masks of all the atlases through a weighted decision scheme. However, the problem still persists and using more atlases results in an increase in computational demands during the registration process.

After, machine learning techniques were addressed. The selection of features is a vital step in creating a robust and generalizable machine learning model, but it can be challenging and requires a deep understanding of the structure being segmented. The features chosen should be informative, relevant, and non-redundant in order to effectively capture important patterns in the data. Dimensionality reduction techniques may be employed to eliminate irrelevant features and improve model performance. Studies have shown that texture-related features can play an important role in the segmentation of EAT, and these features can be extracted using techniques such as Gabor filter banks or using a GLCM matrix. However, even with the use of texture-related features, the segmentation results remain similar to those of other studies that do not focus on these features. Also, it is important to note that most studies in this field have been conducted on small datasets of 20 patients or less, which limits their generalizability. Furthermore, it is often difficult to compare different studies as the patients used in testing are not always disclosed.

Recently, there has been an increase in the use of deep learning techniques in EAT segmentation research. As these methods continue to advance, it is expected that this trend will continue. A significant challenge for models in EAT segmentation is accurately identifying the upper and lower boundaries of the heart, avoiding oversegmentation. To address this issue, many approaches use a specific task to accomplish this. Additionally, accurately identifying the pericardium is a difficult task, even for experts, as the thin layer separating the interior and exterior regions of the pericardium can be difficult to distinguish. To tackle this challenge, some approaches start by segmenting the pericardium and its interior first, and then focus on segmenting the EAT. However, many modifications of the traditional CNN or U-Net can be employed to help the model to focus on most relevant features for the segmentation, and an overall improved performance can be verified.

Overall, there is still much research and work to be done in the area of automatic EAT segmentation. The results of different studies cannot be directly compared as the amount of data used varies greatly. Furthermore, most studies use data from a single cohort with similar properties, not providing good generalization for the model. As a result, direct comparisons between studies can be unfair. Another important factor to consider is the time required for the model to segment a CT scan. For example, the first approach by Rodrigues et al. [26] takes 1.8 hours to segment a CT scan with high accuracy, while the earliest work by Qu et al. [28] only takes 2.2 seconds. It is important to strike a balance between segmentation performance and the time taken, considering the computational resources and the clinical scenarios. Additionally, while 3D networks can provide valuable contextual information and enhance the model's performance, the computational effort required to train and run them is demanding and poses a significant limitation. An issue that affects the reliability of automatic methods is the possibility of them producing segmentations of structures that are not anatomically accurate, such as not being able to properly segment the pericardium in the presence of small calcifications in the patient, which is somewhat incomprehensible.



## Chapter 4

# A First Approach for Pericardial Segmentation

This chapter presents a first approach towards the automatic segmentation of the pericardium in CT imaging. A 2D U-Net is trained for this purpose using publicly available data from two datasets and the performance is then validated in an external private dataset. Throughout this chapter three topics were examined to assess the optimal model: 1) the impact of the images input size; 2) the impact of the loss function; and 3) the impact of changing the HU ranges.

### 4.1 Methods

#### 4.1.1 Datasets

In this work two publicly available datasets were employed - the Cardiac Fat dataset from Rio de Janeiro and the Open Source Imaging Consortium (OSIC) from a Kaggle challenge - and a private dataset acquired at the CHVNGE.

##### Cardiac Fat Dataset

The Cardiac Fat database [60] was acquired in Rio de Janeiro and released publicly by Rodrigues et al. [26]. The dataset includes 20 CT scans with 878 slices belonging to 20 patients as DICOM images ( $512 \times 512$  pixels) with an average pixel spacing of 0.33 mm. The CT scans were obtained with two different scanners (Philips and Siemens) with a slice thickness of 3 mm. The average age of the patients is 55.4. The original ground truth was obtained via manual segmentation by a physician and a computer scientist who labeled the EAT, MAT, and finally, the pericardium layer. It is relevant to point out that CTs were manually centred and cropped to the pericardium and

---

Part of this work has been published at the 11th EAI International Conference on Wireless Mobile Communication and Healthcare (EAI MobiHealth 2022) [59].

thresholded to the adipose tissue range of  $[-200, -30]$  HU. Manual annotations of different structures are made available for these images. For this reason it is not possible to make a direct match of the labels with the original DICOM images. An example of a CT slice and the corresponding manual segmentations is shown in Figure 4.1.

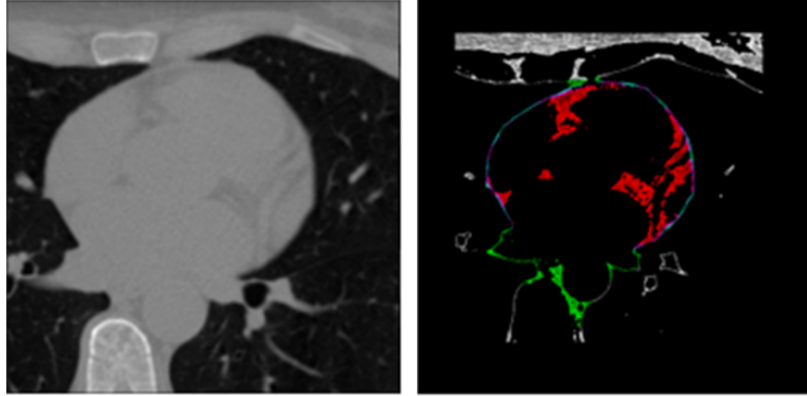


Figure 4.1: Cardiac Fat dataset. Left: DICOM images; right: DICOM in fat HU range and manual segmentations of cardiac fat. Red indicates EAT, green indicates MAT.

### OSIC Dataset

The OSIC dataset was created from the OSIC Pulmonary Fibrosis Competition hosted on Kaggle, available to the general public. The aim of the competition was to investigate the progression of Pulmonary Fibrosis in a patient by analyzing CT images of the lungs. A total of 176 CT scans were provided for this purpose. Fortunately, an experienced radiologist from the hospital center in Bad Berka, Germany, took the initiative to segment the hearts and other anatomical structures of 85 patients, and these segmented scans were utilized in this project [61]. Furthermore, the OSIC dataset consists of 85 CT scans with 12,133 slices whose scans were conducted using six distinct scanners (GE Medical Systems, Hitachi Medical Corporation, PACSMatt, Philips, Siemens, Toshiba), and the scans have been obtained with 15 different slice thicknesses ranging from 0.5 mm to 7 mm. The DICOM images within the dataset exhibit four different sizes (512, 632, 768, 888), with the majority of them (70 CT scans) having a size of  $512 \times 512$  pixels and an average pixel spacing of 0.67 mm. Figure 4.2 shows two CT scans and the respective manual segmentations.

### CHVNGE Dataset

The CHVNGE dataset is a subset of 190 patients randomly selected from the EPICHEART (The influence of EPICardial adipose tissue in HEART diseases) Study (ClinicalTrials.gov: NCT03280433), collected at the CHVNGE in Vila Nova de Gaia, Portugal. The EPICHEART

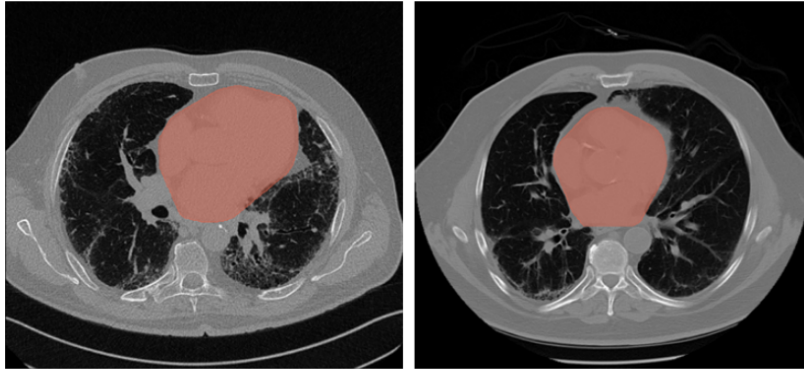


Figure 4.2: OSIC Dataset. DICOM images with manual pericardium segmentation in red.

study is a translational study designed to investigate the association between EAT and CAD, left atrial remodelling and postoperative atrial fibrillation, and frailty syndrome in patients with symptomatic severe aortic stenosis referred to aortic valve replacement. This is a prospective cohort including 574 patients who underwent pre-operative assessment, intra-operative samples collections, and follow-up in hospital and at 6 months after surgery. All data was anonymized prior to analysis for the purposes of this study. The dataset includes 190 CT scans with 8661 slices as DICOM images ( $512 \times 512$  pixels) and an average pixel spacing of 0.45 mm. All CT scans were acquired on a Siemens Somatom Sensation 64 with an ECG-triggered scanning protocol (tube voltage 120 kV, tube current 190 mA, gantry rotation 330 ms, collimation  $24 \times 1.2$  mm, pitch 0.2) and a slice thickness of 3 mm. The pericardial segmentation was obtained via manual segmentation by a medicine student. Figure 4.3 illustrates two instances of CT slices along with their corresponding manual pericardial segmentations.

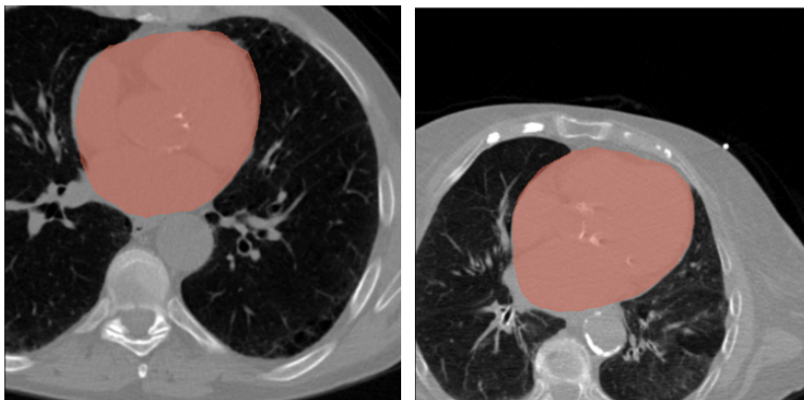


Figure 4.3: CHVNGE Dataset. DICOM images with manual pericardium segmentation in red.

### 4.1.2 Image Registration

Given the mismatch between the labels and original DICOMs due to the manual cropping and centering of CT slices for labelling of the Cardiac Fat dataset, an image registration methodology was used to align the labels to the DICOM images. In this way, full DICOM images can be used to train a deep learning segmentation model without the need for centering and cropping on external datasets.

First, the DICOM images were converted to the same range as labeled images of  $[-200, -30]$  HU to facilitate the recognition of key points between the two images. Image registration was then performed between pairs of DICOM and labeled CTs using the ORB (Oriented FAST and Rotated BRIEF) algorithm [62]. This process consists first of finding the key points using an algorithm called FAST, which mainly uses an intensity threshold between the central pixel and a circular ring around the center. Then, ORB uses BRIEF (Binary Robust Independent Elementary Features) descriptors to describe each keypoint. BRIEF is a bit string description constructed from a set of binary intensity tests between  $n$  pairs  $(x,y)$  of pixels which are located inside the patch (a key point is in the center of the patch). For further details, the reader is referred to the original publications of the ORB algorithm [62]. Next, a brute-force matcher was used to match the key points of the two images, selecting the best results while removing the noisy ones (Figure 4.4). The default parameters and the hamming distance were used for the ORB and the matcher, respectively. Finally, using the Random Sample Consensus (RANSAC) algorithm it is possible to highlight the inliers (correct matches) and find the transformation matrix [63].

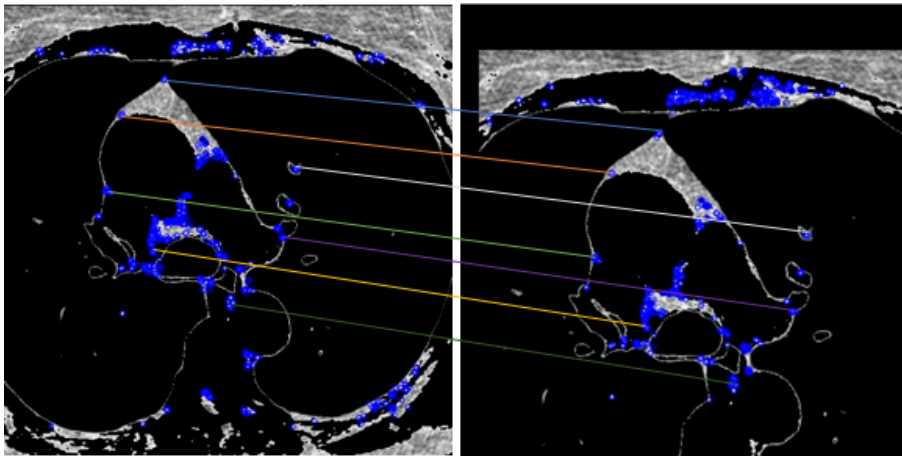


Figure 4.4: Example of the matching of key points between the two images using BRFMatcher.

It should be noted that the images are misaligned in exactly the same way within each patient, therefore only one manually chosen transformation matrix is applied per patient. With this, there will be 20 transformation matrices corresponding to the 20 patients in the database.



### 4.1.3 Pericardial Segmentation

In this section, the data preprocessing is initially presented, including the strategy adopted to obtain the pericardial reference segmentations for the Cardiac Fat dataset. Finally, the models developed will be described, as well as the approach employed to address the issue of unbalanced datasets.

#### Data Preprocessing

While the Cardiac Fat dataset has pericardial contours available, label inconsistencies were found in some images. Therefore, to train and test the segmentation model, only the EAT label was used to obtain an approximation of the pericardial mask using a Convex Hull [64]. This algorithm was applied to all binary masks of the EAT and an example is shown in Figure 4.5.

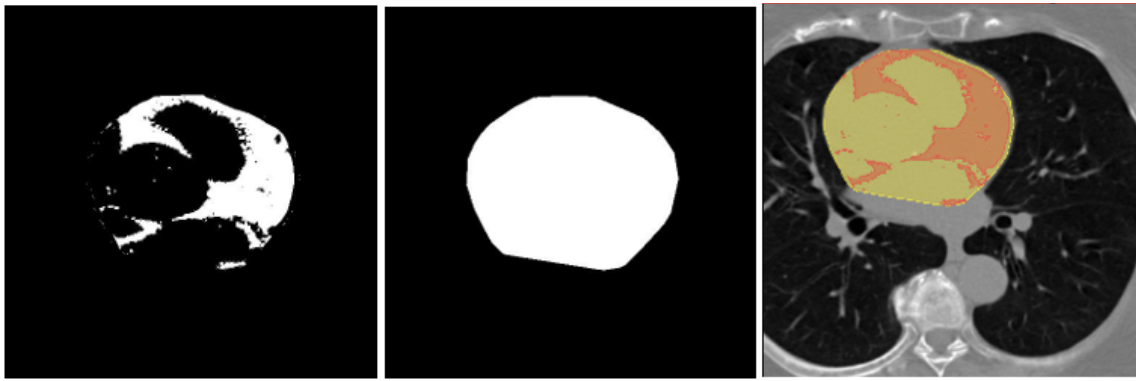


Figure 4.5: Convex hull pericardium labelling example. Left: EAT mask; middle: generated pericardium label using convex hull; right: EAT (red) and Pericardial (yellow) labels registered and overlapped with DICOM images.

Before training the network, the CT slices were first clipped to  $[-1000, 1000]$  HU using 16 bit images. Additionally, the CT scans were also clipped to the  $[-125, 225]$  HU in order to enhance the pericardium layer and assess whether this thresholding could enhance the model's performance. Finally the images were normalized to a range between 0 and 1. Figure 4.6 illustrates the difference between the two ranges HU for the same slice.

For the training, validation and testing of the model, each one of the Cardiac Fat and OSIC datasets were randomly divided as follows: 60% of the CT scans for training, 20% of the CT scans for validation, and the remaining 20% for testing. A patient-wise division was applied to avoid bias given that different CT slices of the same patient are highly correlated.

The CHVNGE data was used for external testing of the automatic segmentation network and were thus not used for training.

#### Model Architecture and Training Setup

Pericardial segmentation was then performed using a U-Net architecture [51], as show in Figure 4.7. The U-Net is a popular framework for deep learning models, it often obtains excellent performance in image segmentation, especially in the area of medical image processing. This model

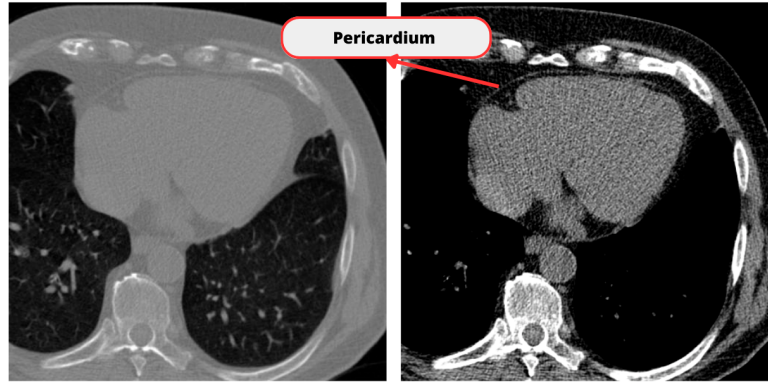


Figure 4.6: CT image shown in ranges  $[-1000, 1000]$  HU (left) and  $[-125, 225]$  HU (right) showing pericardium layer enhancement.

was trained from scratch with a total number of 31,031,685 parameters.

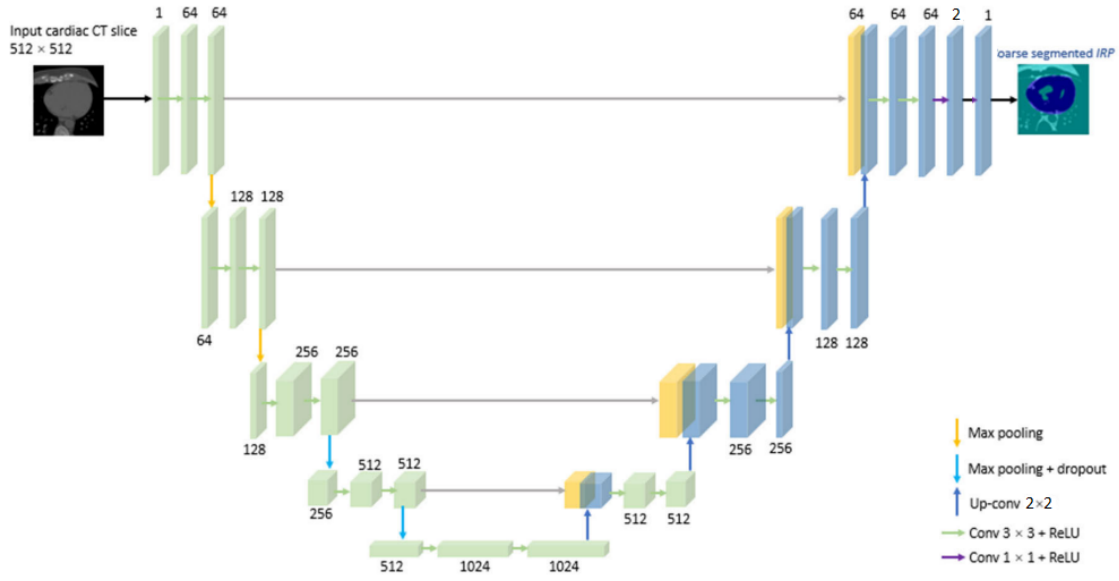


Figure 4.7: U-Net architecture used for pericardial segmentation (adapted from [38]).

Two widely utilized loss function were employed in this work: the Dice loss function, which is simply  $1 - DSC$  (eq. 3.4), and the binary cross-entropy (BCE) loss, defined as:

$$BCE = -\frac{1}{N} \sum_{i=1}^N y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i)) \quad (4.1)$$

where  $y$  is the label and  $p(y)$  is the predicted probability of the pixel being from the label for all  $N$  pixels.

An early stopping callback method was also used to stop training when the validation loss did not improve for 20 consecutive epochs. Adam was used as the optimizer. The deep learning

models were trained using Tensorflow and an NVIDIA GTX 1080 GPU.

Note that for each model, the weights were initially optimized for the slices containing the pericardium and after a new learning process was carried out, encompassing all slices, including those without the pericardium. This strategy was adopted because there are many slices without pericardium for each patient (upper and bottom ones). By employing this strategy, the model's focus was directed towards accurately segmenting the pericardium, ensuring an overall strong performance.

For training and validation, a custom image generator was utilized. This generator employed an alternating approach for each consecutive step in the epoch, switching between images from the two datasets: Cardiac Fat and OSIC. The alternating behavior is depicted in Figure 4.8. This strategy was implemented to address the challenge of unbalanced datasets, where the Cardiac Fat dataset consisted of 20 patients and the OSIC dataset included 85 patients. The approach of alternating between the OSIC and Cardiac Fat datasets enhances the variability of the data during training and avoids overfitting to the majority dataset.

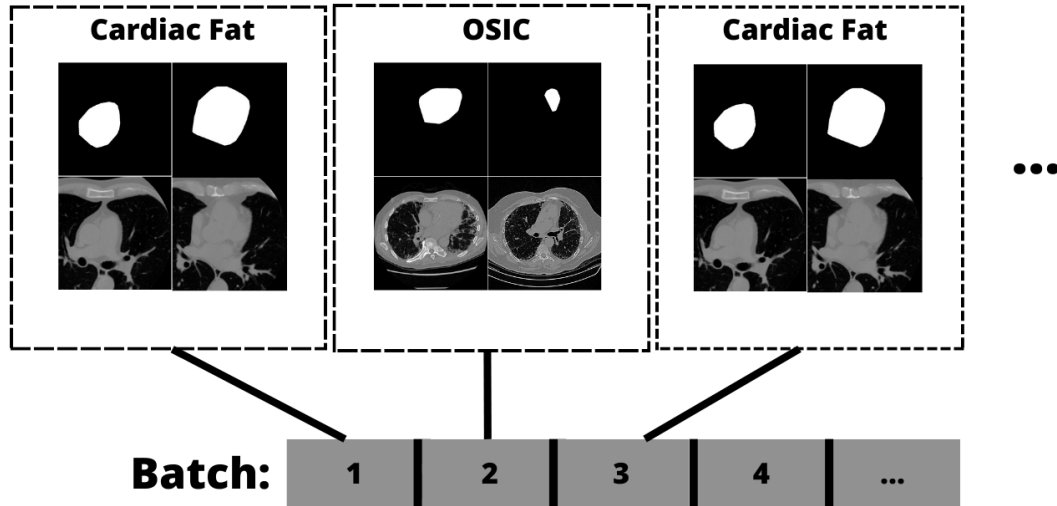


Figure 4.8: Illustration of the method for alternating between datasets for each batch during training.

## 4.2 Experiments

In this chapter, three experiments were conducted with the aim of identifying the optimal model that accurately segments the pericardium.

First, two different input sizes were tested to evaluate the importance of contextual information and network receptive field:  $512 \times 512$  and  $256 \times 256$ . Note that for these experiments, the range of  $[-1000, 1000]$  HU was considered for the CT scans and the Dice loss as the loss function. A batch size of 4 and learning rate of  $1e-5$  was used to train the models with  $512 \times 512$  images. For

the models trained with  $256 \times 256$  images, a batch size of 12 and a learning rate of  $1e-4$  was used. These hyper-parameters were selected based on empirical experiments, taking into account the memory constraints of the workstation used.

After choosing the best image input size, a comparison between the Dice loss function and BCE was made in order to determine which of the two popular losses is better for segmentation tasks. Besides that, two different HU ranges were used:  $[-1000, 1000]$  HU, which includes almost all the values in a DICOM format, and  $[-125, 225]$  HU, which enhance the pericardium layer. The  $[-125, 225]$  HU range is applied on literature aiming to help the model in focusing on the pericardium [40]. The objective of these experiments was to determine the optimal combination of loss function and HU range in order to achieve an accurate segmentation of the pericardium.

## 4.3 Results and Discussion

### 4.3.1 Image Registration

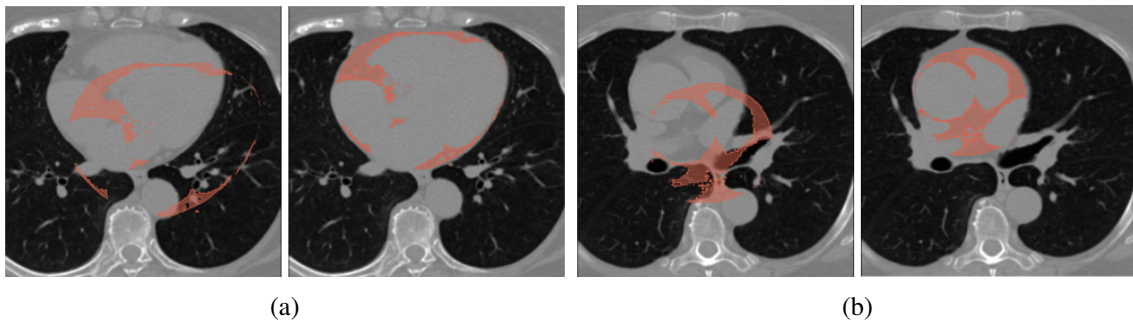


Figure 4.9: Two examples ((a) and (b)) before (left) and after (right) image registration.

The result of the image registration for two patients is shown in Figure 4.9. It can be seen that before the application of image registration for this patient, the labels were out of alignment with the DICOM image. After registration, a correct alignment of the mask with the DICOM image is verified as seen in Figure 4.9. The registration results were evaluated qualitatively and a successful alignment was verified for all patients.

### 4.3.2 Pericardial Segmentation

The discussion will be divided into two parts: firstly, the performance on the public datasets (where the models were trained) will be addressed and then, the performance on the CHVNGE dataset will be evaluated.

Note that, when a specific value of a metric is mentioned, it will be presented as the mean value with the standard deviation (mean  $\pm$  standard deviation).

### Impact of the Input Sizes

In this subsection, the impact of the images input sizes on the performance of the model will be addressed. The DSC, HD and MAD means values for the models studied, are presented in the Figure 4.10 with the respective boxplot.

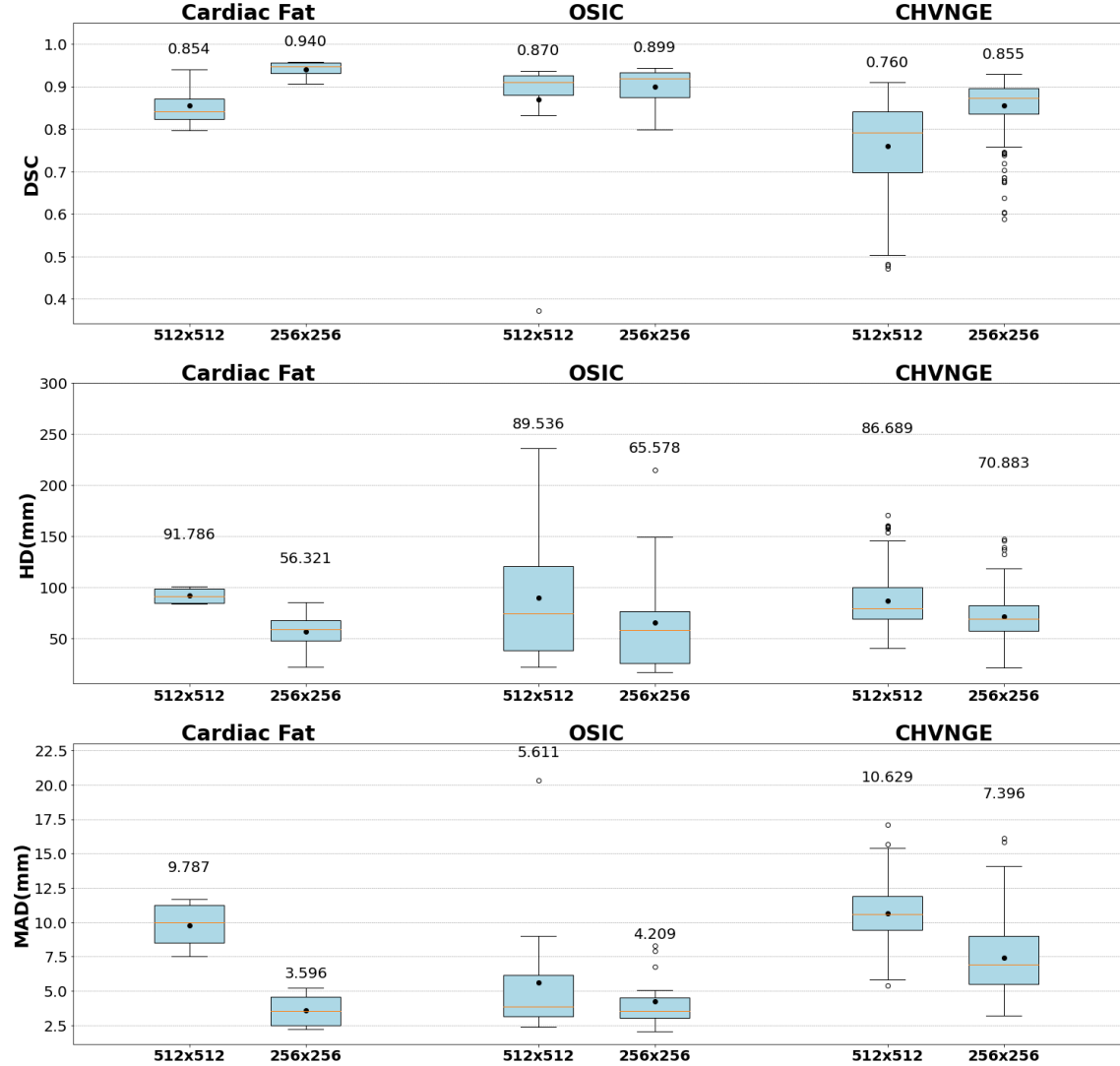


Figure 4.10: Results for the three datasets for each image size, where the three metrics are displayed using boxplots. The mean is depicted as a point with the corresponding value written above the maximum value of each boxplot.

By analysing the boxplots presented in the Figure 4.10, it can clearly be seen that the 256x256 image size achieved a higher performance for all the three metrics regarding the Cardiac Fat dataset. In fact, the DSC reaches  $0.940 \pm 0.02$  for 256x256 images and  $0.854 \pm 0.053$  for 512x512 images. The same trend is observed for HD and MAD where the superiority of the 256x256 image size is evident. A similar outcome can be extracted for the OSIC Dataset but with less difference between the two images sizes model's performance. Indeed, a DSC value of 0.899

$\pm 0.044$  and  $0.870 \pm 0.136$  was obtained for the  $256 \times 256$  and  $512 \times 512$  image size, respectively. As expected, it was possible to notice that, according to the DSC, better results were obtained for the Cardiac Fat Dataset, when analysing the  $256 \times 256$ . This can be attributed to the smaller number of patients in the Cardiac Fat test set (4 patients) compared to the OSIC test set (17 patients). This rationale is valid for the subsequent analysis. However the values of the HD and MAD are similar between the two datasets with a slightly better performance for the Cardiac Fat, exhibiting more consistent results with less outliers.

Three examples of automatic segmentations predicted by both models are presented in Figure 4.11. It can be seen that the images predicted by the  $256 \times 256$  model present better results, being more similar to the corresponding label. In the case of the  $512 \times 512$  images, there are some flaws in the classification of pixels and even inside the pericardium, some pixels are not attributed to it. The bottom row of Figure 4.11 highlights the difficulty in segmenting the lower portion of the pericardium due to the presence of other organs in the CT slice.

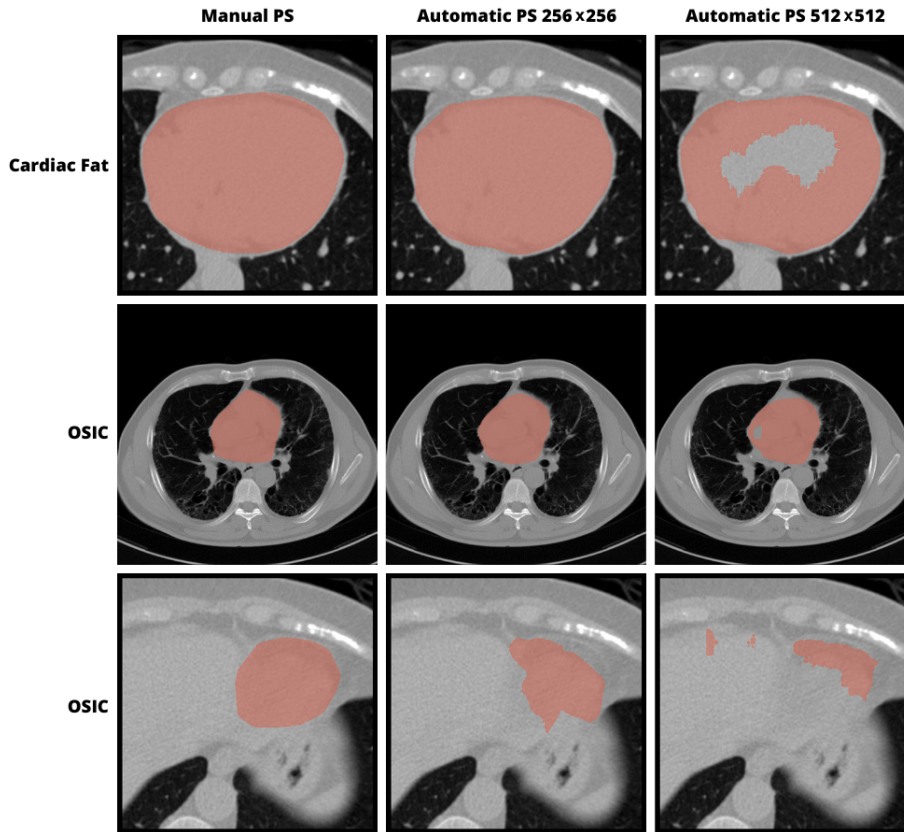


Figure 4.11: Examples of automatic pericardium segmentation (PS) for the two input sizes in the public datasets.

The observed worse performance for the  $512 \times 512$  images is likely due to the fact that the larger image size results in a decrease in the size of the receptive field thus giving less contextual information to the network, which can be important for segmentation. However, the use of different hyperparameters for the two models (batch size and learning rate) can also play a role.



Regarding now the CHVNGE dataset, It is possible to note that, again, images with  $256 \times 256$  showed better results with a DSC of  $0.855 \pm 0.063$  for images with  $256 \times 256$  pixels and  $0.760 \pm 0.101$  for  $512 \times 512$  images. Similar findings can be observed for HD and MAD, which values are worse for the  $512 \times 512$  images.

Two examples of automatic segmentations by both models are presented in Figure 4.12. On the top row, it can be observed that the results are better using  $256 \times 256$  images and that the prediction is more similar to the corresponding label. However, examining the bottom row of the Figure 4.12, it can be seen that the significant calcification in this patient led to a failure of the segmentation network with a large portion of the center of the pericardium excluded from the segmentation. This example shows the need to train the model with a larger amount of data and especially with as much variability as possible with the aim of making the model more generalizable.

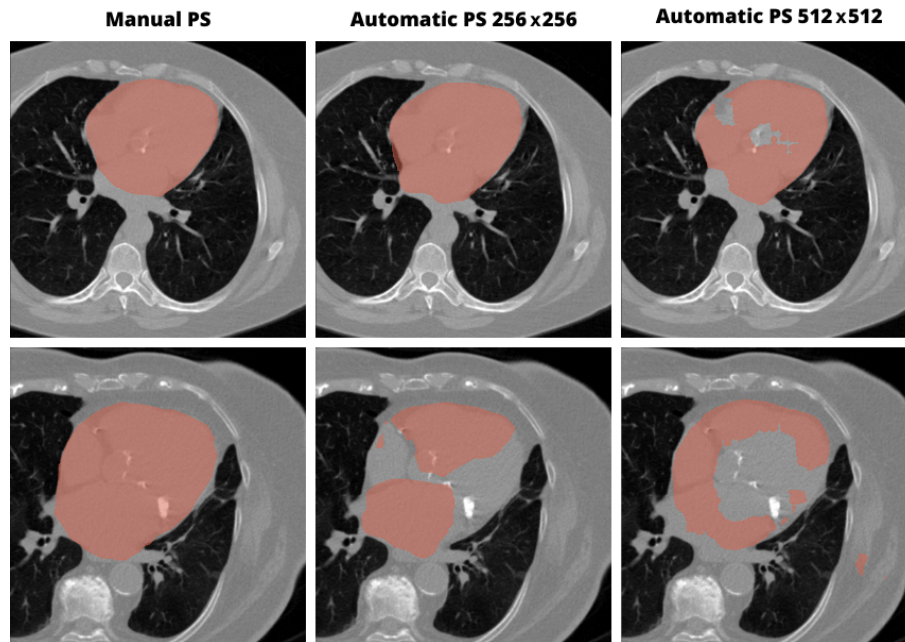


Figure 4.12: Examples of automatic pericardium segmentation (PS) for the two input sizes in the CHVNGE datasets.

Still it is noteworthy that the results were moderately worse compared to those obtained in the Cardiac Fat and OSIC dataset in all three metrics. This decrease in performance was however expected since the model was trained on the public datasets. In addition, the data from the hospital has significantly more patients with 190 patients tested against 4 and 17 for the Cardiac Fat and OSIC dataset, respectively, increasing the variability and consequently the difficulty for the model to predict correctly.

In summary, after analyzing the results, it was found that the  $256 \times 256$  image size consistently yielded better performance in all metrics. Therefore, our study will exclusively focus on using this image size for the CT scans in future experiments.

### Impact of the Loss Function and HU Ranges

This subsection will address the impact of the loss function on the performance of the models. Note that all the models until here were trained with the Dice loss. Therefore, the same models were trained with the BCE loss and the impact will be evaluated. Figure 4.13 displays the metrics means values for the models studied, along with their corresponding boxplot.

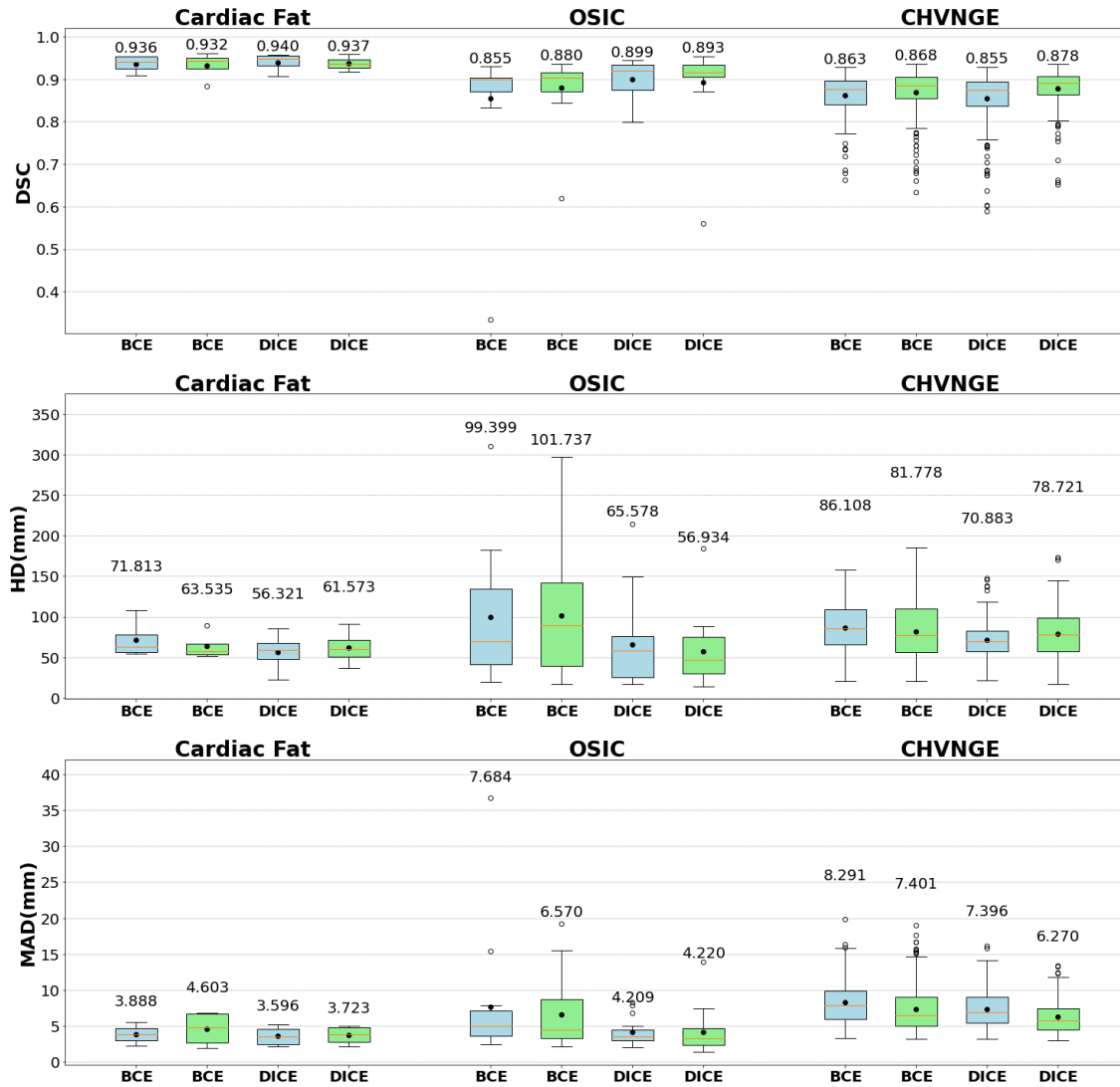


Figure 4.13: Results of the loss function and HU range for the three datasets, where the three metrics are displayed using boxplots. Note: blue and green color for the boxplots represents the [-1000, 1000] HU and [-125, 225] HU range, respectively.

The results observed in the boxplots of the Figures 4.13 demonstrate a decrease in performance when using the BCE loss function, compared to the Dice loss for both HU ranges. In fact, it is possible to see that, while the DSC for the Cardiac Fat remains quite similar but a slightly worse for the BCE models, the MAD and HD are significantly higher for the BCE loss, specially in the [-1000, 1000] HU range. The same outcome can be drawn for the OSIC dataset. Indeed, the DSC



experiences a significant decrease from  $0.899 \pm 0.044$  to  $0.855 \pm 0.142$  when training the model with the BCE loss and  $[-1000, 1000]$  HU range. The  $[-125, 225]$  HU range shows the same trend but with a limited effect and upon analysis, it was determined that the deviation in performance can be attributed to the presence of an outlier that the model struggled to accurately segment. The only exception to this trend is observed in the CHVNGE dataset where the BCE model performs slightly better in terms of DSC with  $0.863 \pm 0.048$  for the model trained with BCE loss and  $0.855 \pm 0.063$  for the one trained with the Dice loss, regarding the  $[-1000, 1000]$  HU range. However, the situation is different within the range of  $-125$  to  $225$  HU, as the DSC remains superior for the Dice loss function, exhibiting a  $+0.01$  difference compared to the model trained with the BCE loss. The same can be observed for the other metrics, where the HD and MAD are once again worse in the BCE model.

In summary it was decided to choose the Dice as the loss function in the next experiments, as this one outperforms the BCE loss in all metrics and for all datasets with the exception mentioned before. Therefore, a analysis of the impact of two different HU ranges on the model's performance is presented, with the Dice as loss function.

Regarding the influence of the used HU range on the public datasets, the results demonstrate that the  $[-1000, 1000]$  HU range outperforms the  $[-125, 225]$  HU range in terms of almost all of the three metrics for both datasets. Although the DSC results for the Cardiac Fat dataset are similar for both ranges, the HD and MAD indicate significantly better results for the  $[-1000, 1000]$  HU range. This outcome can also be applied to the OSIC Dataset with one exception. Indeed, the HD is slightly better in terms of mean value for the  $[-125, 225]$  HU range. Nonetheless, the significant standard deviation observed for the  $[-125, 225]$  HU range, which can be attributed to the presence of two outliers as depicted in Figure 4.13, indicates that the  $[-1000, 1000]$  HU range may yield a more consistent performance in this scenario.

The Figure 4.14 illustrates two examples of automatic pericardial segmentation for the Cardiac Fat and OSIC datasets. It can be observed that the images with the range  $[-1000, 1000]$  HU exhibit a closer resemblance to the manual segmentation for both datasets. The example for the Cardiac Fat is demonstrative of the difference of the HD values between the two ranges, as the line delineating the  $[-125, 225]$  HU segmentation is further away from the line representing the  $[-1000, 1000]$  HU segmentation in comparison to the manual segmentation. This larger distance results in a less favorable HD value. The same trend can be observed for the OSIC dataset.

These results can be explained due to the fact that the  $[-1000, 1000]$  HU images contain more information than the  $[-125, 225]$  HU ones, as a larger range of HU values are retained. Therefore, the model can likely extract more relevant features for the pericardium when more comprehensive information during training is given, resulting in a better performance.

Paying attention now to the results for the CHVNGE dataset in Figure 4.13, the results are not consistent to those observed on the public datasets. Contrarily of the insights given from the public datasets, the results demonstrate that the  $[-125, 225]$  HU range achieved better performance than  $[-1000, 1000]$  HU range in terms of DSC with a difference around  $+0.02$  between the two HU ranges. The MAD is also worse for the  $[-1000, 1000]$  HU range. However, better results according

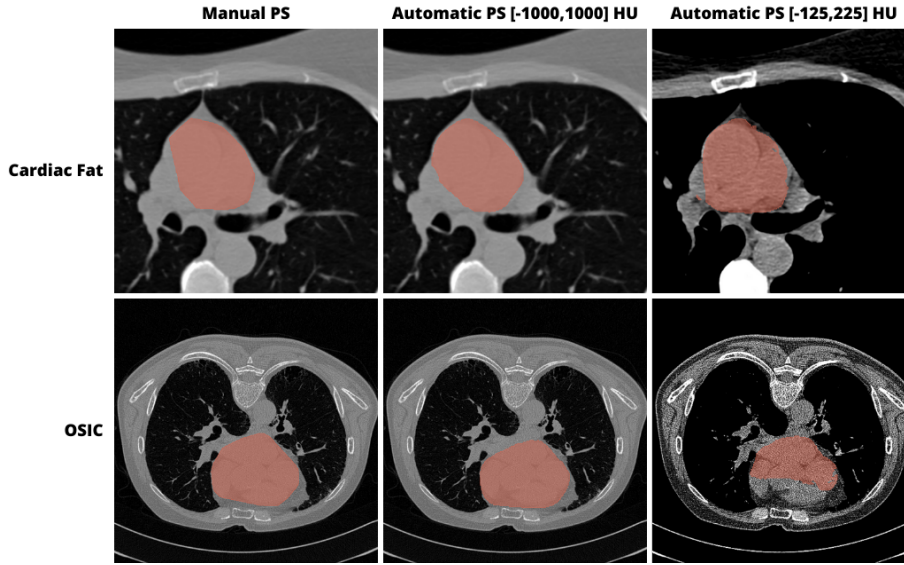


Figure 4.14: Examples of automatic pericardium segmentation (PS) for the two different HU ranges in the public datasets.

to HD are obtained for the  $[-1000, 1000]$  HU range, outperforming the  $[-125, 225]$  HU range by around -8 mm.

The Figure 4.15 shows two pericardium segmentations examples for the CHVNGE dataset. For the first patient it is evident that the presence of calcifications in the CT slice affects more the model trained with the  $[-1000, 1000]$  HU images than the model trained with  $[-125, 225]$  HU images. In fact, the segmentations obtained with the  $[-1000, 1000]$  HU are poor with numerous disconnected components surrounding the calcifications. However, this effect seems to be mitigated when using the  $[-125, 225]$  HU images and the automatic segmentations are similar to the manual ones. The second patient exemplifies a case where the  $[-1000, 1000]$  HU model performs better in terms of DSC and, mainly, HD metrics, as the contour that delineates the  $[-125, 225]$  HU segmentations is farther from the manual segmentations.

These results can be explained by the fact that calcifications are represented by pixels with high HU values. When the HU range is clipped to  $[-125, 225]$ , all values above 225 are disregarded, which includes the values associated with calcifications. As a result the model segments more accurately the pericardium without considering the calcifications. However this clipping is not always advantageous as it can be seen by the second example of the Figure 4.15 or by the results of the public datasets and the examples provided in the Figure 4.14. This happens because, although the clipping removes the negative influence of the calcifications on the model, it also eliminates important information related to the pericardium and other organs, as explained before. This lack of comprehensive information can lead to more difficulties in accurately segmenting the pericardium. It should be noted that the public datasets used to train the models do not contain many calcifications, and the results for the  $[-1000, 1000]$  HU range are better.

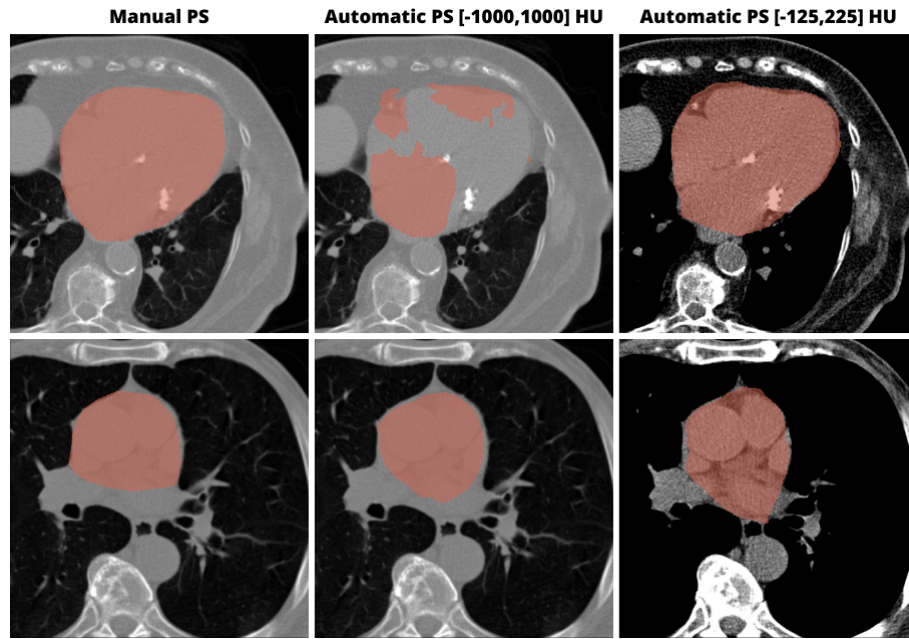


Figure 4.15: Examples of automatic pericardium segmentation (PS) for two patients from CHVNGE dataset, considering the two different HU ranges.

## 4.4 Final Outcomes

This chapter represents a initial investigation conducted to determine the best approach to achieve an accurate pericardial segmentation. The models were trained with two publicly available datasets (Cardiac Fat and OSIC) with a further performance validation on an external private dataset (CHVNGE). The impact of the input size of the images, loss function and the HU range were accessed.

In summary, it was determinate that, an input size of  $256 \times 256$ , the  $[-1000, 1000]$  HU range, and using the Dice loss function can be the best approach for the next experiments. Despite the improved results observed for the CHVNGE dataset within the  $[-125, 225]$  HU range, it has been decided to disregard this range and continue exploring the  $[-1000, 1000]$  HU range. It has been recognized that the benefits of the  $[-125, 225]$  HU range are primarily associated with the presence of calcifications in CHVNGE patients. However, the upcoming chapter will address the challenge of effectively dealing with these calcifications while still benefiting from the comprehensive information provided by the  $[-1000, 1000]$  HU range. Beside that it was observed that the models can often fail, particularly for the lower pericardium and the use of networks to provide further context during segmentation would thus be an extremely important strategy.



## Chapter 5

# Data Augmentation and Artificial Calcifications

In the previous chapter, a baseline was established to determine the optimal model for pericardial segmentation. Overall, it was determined that the best input size is  $256 \times 256$ , and the Dice loss function is better than the BCE loss function.

This chapter introduces a new approach aimed at addressing the challenges identified in Chapter 4, particularly concerning the model trained on images within the  $[-1000, 1000]$  HU range. First, data augmentation techniques will be employed to add some variability in the existing data. Next, artificial calcifications will be added in the training data to help the  $[-1000, 1000]$  HU model dealing with actual calcifications found in some patients.

### 5.1 Methods

This section aims to provide an overview of the data augmentation techniques utilized in our work, as well as introduce the innovative idea of creating artificial calcifications.

#### 5.1.1 Data Augmentation

Data augmentation techniques are heavily used in deep learning as it allows to increase the variability of the data without the need for additional data acquisition. This is achieved by applying some transformations in the images while preserving their inherent characteristics and the corresponding labels. In our case the following transformations were employed in the training images: rotations within a range of 5 degrees, horizontal and vertical flip, horizontal and vertical shifts within a range of 10% of the total length of the image, zoom within a range of 20% (more or less the size of the image). The models subjected to data augmentation will be labeled with "DA" preceding their names.

The Figure 5.1 illustrates some examples of these transformations for one slice.

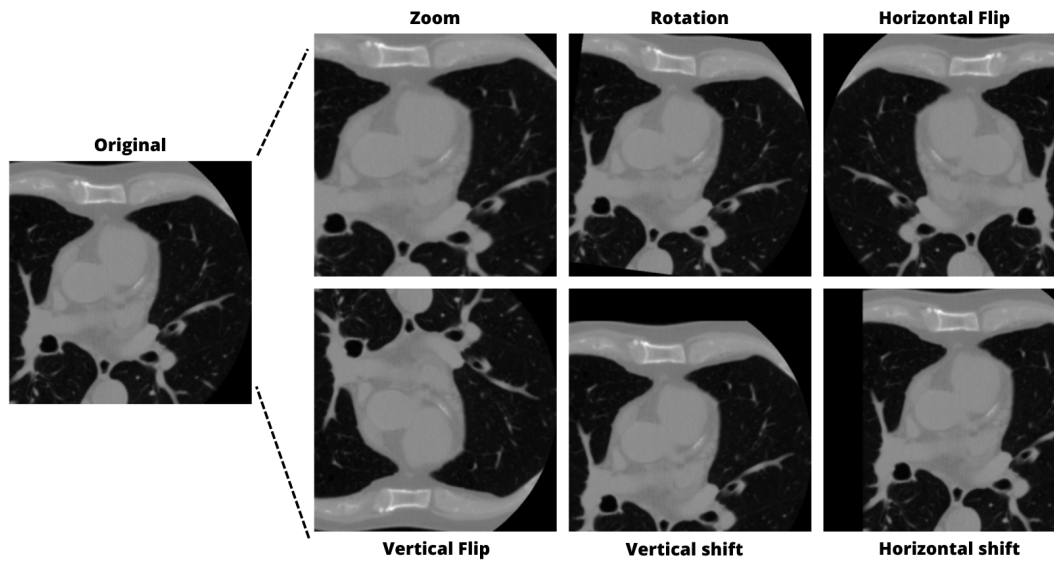


Figure 5.1: Example of the applied data augmentation transformations on an illustrative CT slice.

### 5.1.2 Artificial Calcifications

As discussed in the previous Chapter, the presence of calcifications, mainly on the CHVNGE dataset, negatively affected the segmentation performance. Although the use of the range  $[-125, 225]$  HU mitigate these effects, its usage is not always suitable and performs worse in other scenarios, as discussed in Chapter 4. The challenge is to find a solution that can handle the presence of calcifications while still utilizing the comprehensive information provided by the  $[-1000, 1000]$  HU range. Figure 5.2 illustrates calcifications within the CHVNGE dataset.



Figure 5.2: Example of a few calcifications in the CHVNGE dataset.

Therefore, an innovative solution was to artificially create these calcifications and add them to the training data.

After analysing the pixel values of the calcifications it was evident that they represent high HU values, close to the limit of our window of HU (1000). Besides that, it was decided to replicate these calcifications with a Gaussian distribution. What is proposed on this work is to create these high intensity values with a 2D Gaussian shape and add to the CT slice.

The 2D Gaussian distribution can be defined by the following expression:

$$G(x,y) = A \exp(-(a(x-x_o)^2 + 2b(x-x_o)(y-y_o) + c(y-y_o)^2)) \quad (5.1)$$

with,

$$a = \frac{\cos^2(\theta)}{2\sigma_x^2} + \frac{\sin^2(\theta)}{2\sigma_y^2}$$

$$b = -\frac{\sin(2\theta)}{4\sigma_x^2} + \frac{\sin(2\theta)}{4\sigma_y^2}$$

$$c = \frac{\sin^2(\theta)}{2\sigma_x^2} + \frac{\cos^2(\theta)}{2\sigma_y^2}$$

Here  $\sigma_x$  and  $\sigma_y$  are the standard deviation in the  $x$ -direction and  $y$ -direction,  $\theta$  represents the angle of rotation,  $(x_o, y_o)$  represents the coordinates of the Gaussian center's and  $A$  the amplitude or intensity of the gaussian.

By manipulating these parameters of the gaussian, it was possible to create different calcifications with different orientations, sizes and shapes as can be seen by the Figure 5.3. The models that were trained using artificial calcifications will be denoted by the prefix "AC" in their names.

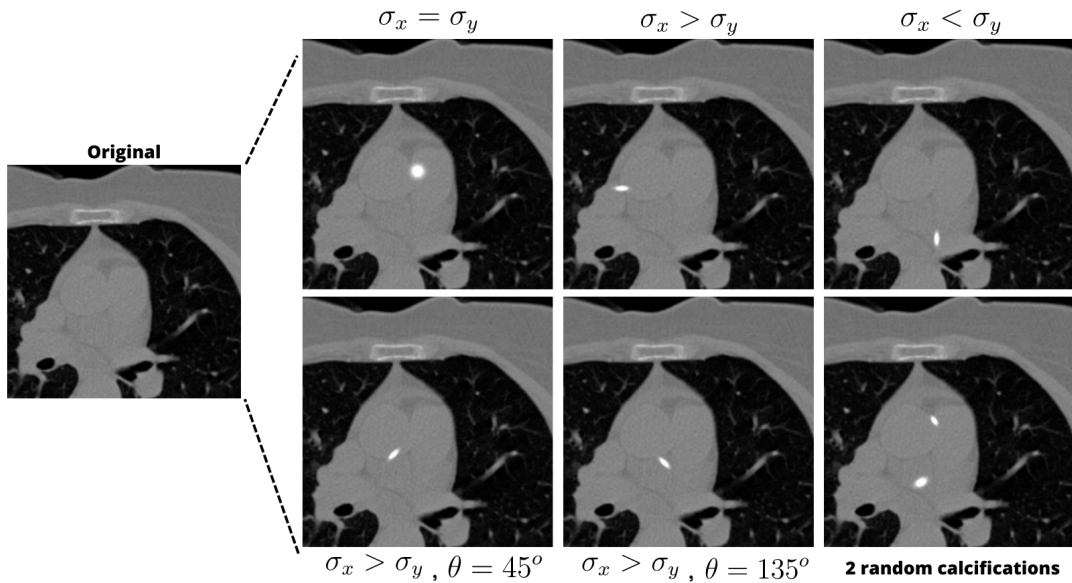


Figure 5.3: Example of artificial calcifications and the effect of manipulating parameters of the Gaussian.



### 5.1.3 Training Setup

The training of the model was conducted similarly as described in the Chapter 4, with  $256 \times 256$  image size, the Dice Loss function and for  $[-1000, 1000]$  HU images. However, during the training of the models, both data augmentation techniques and artificial calcifications were implemented.

The calcifications were added in the image generator with a  $1/3$  probability of occurrence. Besides that, it was noticed that in the CHVNGE dataset, calcifications were rarely observed in the lower slices, where the pericardium size is small. To replicate this pattern in the training data, the calcifications would appear only when the ratio between the numbers of pericardium pixels and the total number of pixels was above 0.1.

## 5.2 Experiments

In this chapter, two main experiments were performed to evaluate the impact of each one of the new implementations to the model trained with  $[-1000, 1000]$  HU images.

The impact of augmentations is initially examined by comparing the outcomes of models with and without the utilization of data augmentation techniques described before.

Finally, the effect of adding artificial calcifications is evaluated by comparing the results with and without the implementation of this new technique in combination with the data augmentation.

Again, all the three metrics (DSC, HD, MAD) were used to evaluate the performance's model. Throughout these experiments, numerous examples were presented to analyze and understand the impact of these new approaches.

## 5.3 Results and Discussion

Figure 5.4 displays these metrics for each one of the datasets. Across all the discussion, whenever a specific value of a metric is mentioned, it will be presented as the mean value accompanied by the standard deviation (mean  $\pm$  standard deviation).

Regarding the public datasets, and analysing the boxplots for the *DA* and *Standard* models presented in the Figure 5.4, it is possible to see that for the Cardiac Fat dataset, no significant differences was observed for the DSC but more consistent results with less variability of the values was obtained. However the HD shows a high increase for the model with augmentations achieving a mean of  $86.553 \pm 20.161$  mm against  $56.321 \pm 22.605$  mm without augmentations. The same can be concluded for the MAD. Regarding now the OSIC dataset, all the three metrics were superior when augmentations were used. Notably, the DSC exhibited the most significant improvement, increasing from  $0.899 \pm 0.044$  to  $0.918 \pm 0.023$  when utilizing the *DA* model.

Figure 5.5 shows an example of the impact of the data augmentation on the two public datasets. The top row is illustrative of the difference between the HD when data augmentation is employed or not. It is possible to note that the line that delineate the segmentation performed by the *Standard* model is closer to the manual one, when comparing with the *DA* model. The bottom row shows



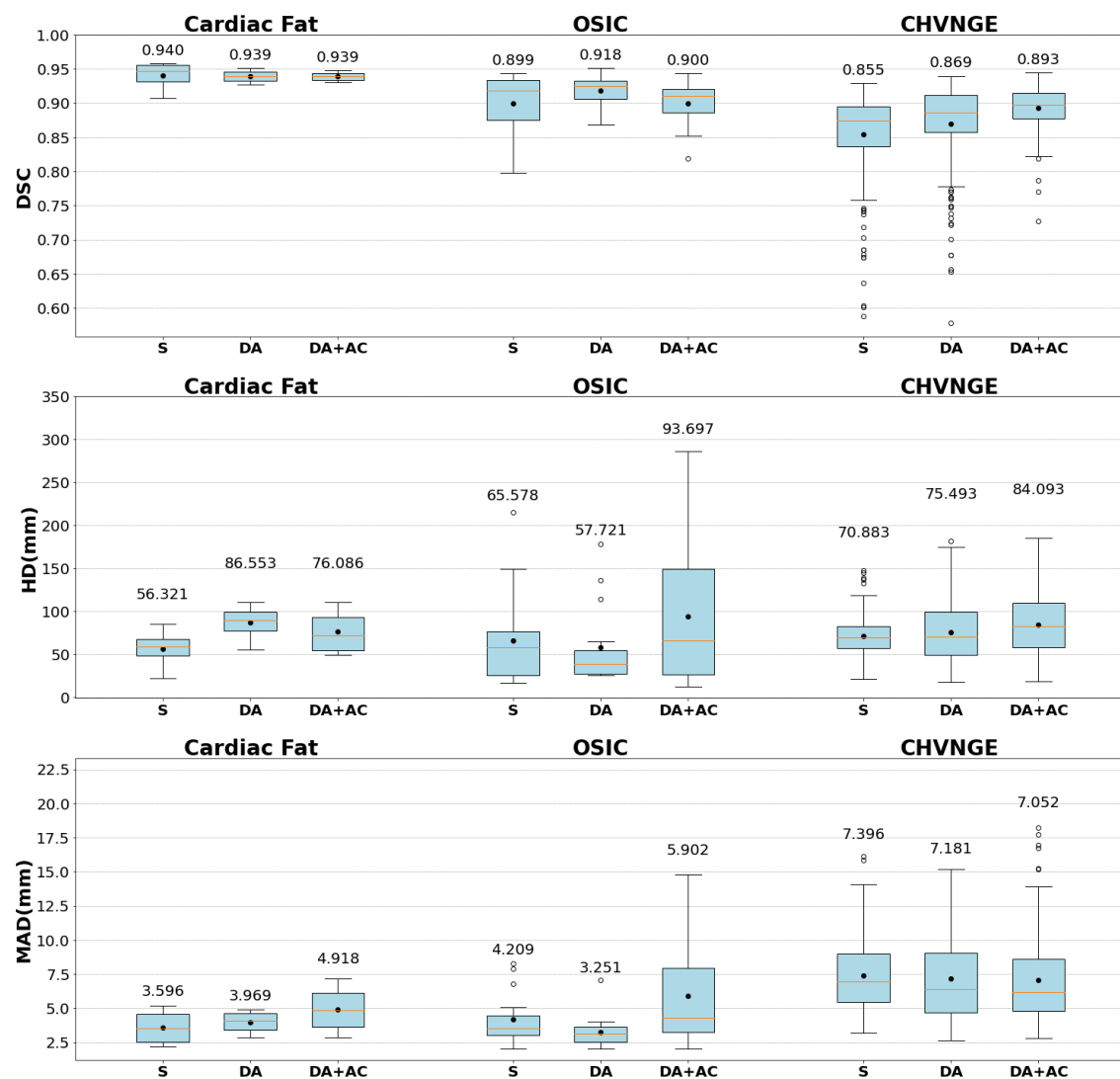


Figure 5.4: The results of each model for the three datasets, where the three metrics are displayed using boxplots, with the mean depicted as a point with the corresponding value written on the upper side. Note that S= Standard model; DA= DA model; DA + AC= DA+AC model

an example of the OSIC dataset. Through the analysis of this example, it becomes evident that the utilization of augmentations has proven to be highly beneficial. In fact, maybe the vertical flip or zoom played a significant role in this specific example.

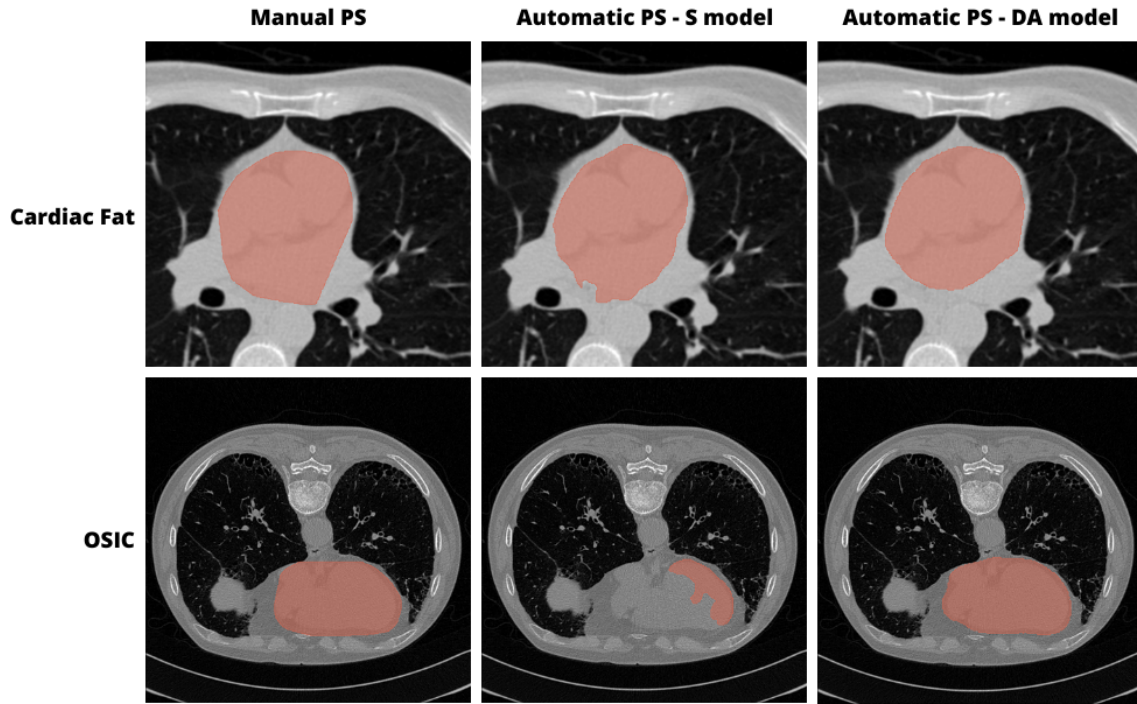


Figure 5.5: Examples of the effect of data augmentation on the automatic pericardium segmentation for the public datasets.

In general, the application of data augmentation yielded positive results in terms of the DSC for the public datasets, with a more pronounced impact observed in the OSIC dataset. However, the implementation of data augmentation exhibits a negative impact for the other metrics in the Cardiac Fat dataset.

Regarding now the private dataset, a significant improvement was observed when the model was trained with augmentations, specially for the DSC. Indeed, when evaluating the DSC, the *DA* model achieved a higher value of  $0.869 \pm 0.061$ , compared to the *Standard* model which had a value of  $0.855 \pm 0.063$ . The same outcome can be drawn for MAD but not for the HD. In fact, for HD, the values were worse for the *DA* model, but for MAD an improvement was observed when augmentations were employed.

Two patients are presented in the Figure 5.6. In the top row it is clear that the augmentations employed to the model were important to improve the accuracy of the automatic segmentation. In fact, for this specific case the transformation of the vertical shift in the training data, certainly had a important role for this accurate prediction of the model. The second patient serves as another example highlighting the impact of augmentations, as the difference between the model with augmentations and the one without is clearly evident.

In summary, the application of data augmentation techniques to the models had a positive

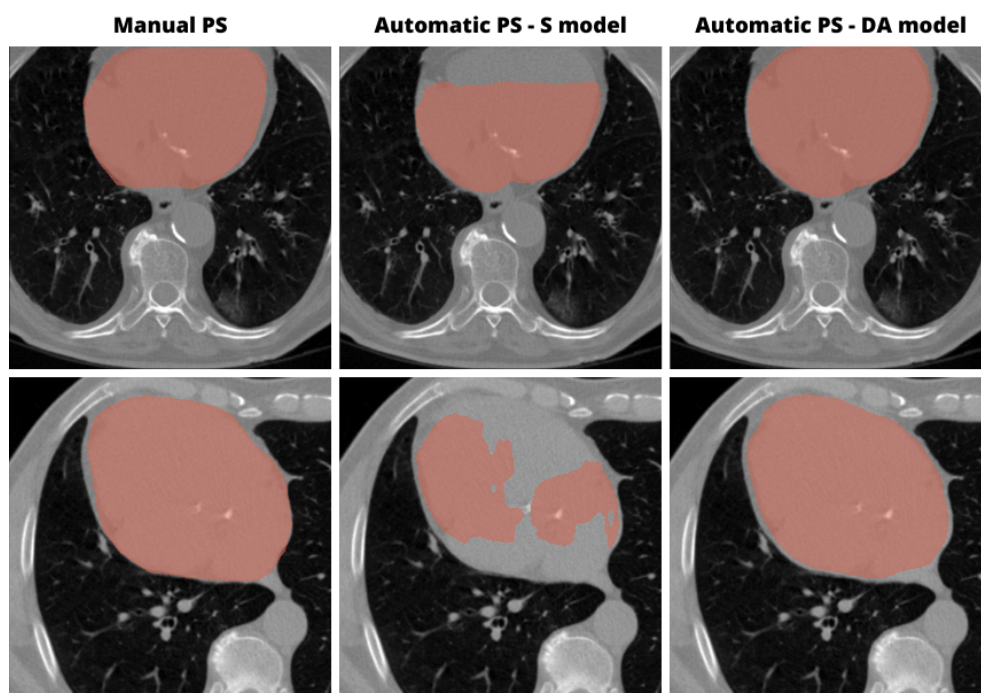


Figure 5.6: Examples of the effect of data augmentation on the automatic pericardium segmentation for the CHVNGE dataset.

impact on both the public and private datasets, particularly in terms of the DSC. Another inference made is that data augmentation has a negative effect on the HD for almost all datasets. This can be explained to HD high sensitivity towards spatial misalignment and distortions between the predicted labels and ground truth. Since data augmentation often involves scaling, rotations, and distortions, these modifications can introduce additional misalignments and lead to poorer results in this specific metric.

Regarding the *DA + AC* results, it is possible to see that similar results were obtained for the public datasets. In fact, for the Cardiac Fat dataset, the DSC remains practically the same, maintaining the value of  $0.939 \pm 0.006$  compared to the two other models, albeit a more stable performance is observed across the patients. However, in terms of the HD, an evident improvement is observed compared to the *DA* model, even though a decrease is observed compared to the *Standard* model, reaching the value of  $76.086 \pm 24.726$  mm. Nevertheless, when observing the MAD this behaviour is not verified, and a higher MAD is obtained with a value of  $4.918 \pm 1.660$  mm. For the OSIC dataset, the combination of both techniques resulted in a general worse performance. In fact, the DSC value is comparable to the *Standard* model but worse than the *DA* model. However the HD and MAD achieved significant worse results, with values of  $93.697 \pm 80.415$  mm and  $5.902 \pm 3.714$  mm, respectively. These values indicate poorer performance compared to the other two models under consideration mainly due to an outlier that can be observed in the Figure 5.4.

Figure 5.7 presents two patients from the public datasets, demonstrating the outputs generated by the three different models: first the *Standard* model, next the *DA* model, and finally the *DA +*

AC model. The first patient from the Cardiac Fat exemplifies the advantage of combining both techniques, as the automatic pericardium segmentation is notably improved compared to the other two models. For the second patient from the OSIC dataset, the output of the two last models appears to be similar and better than the first one. However, when both techniques were used together, a slightly worse performance was observed for pericardium segmentation compared to the *DA* model. The last row of the figure exhibits lower CT slices from the same patient. Indeed, the main reason for the decrease in performance from the *DA + AC* model is that it struggles in accurately segmenting the pericardium in the lower slices, as can be observed by the third example. These observation explains the decrease in performance for the DSC and mainly, for the HD compared do the *DA* model.

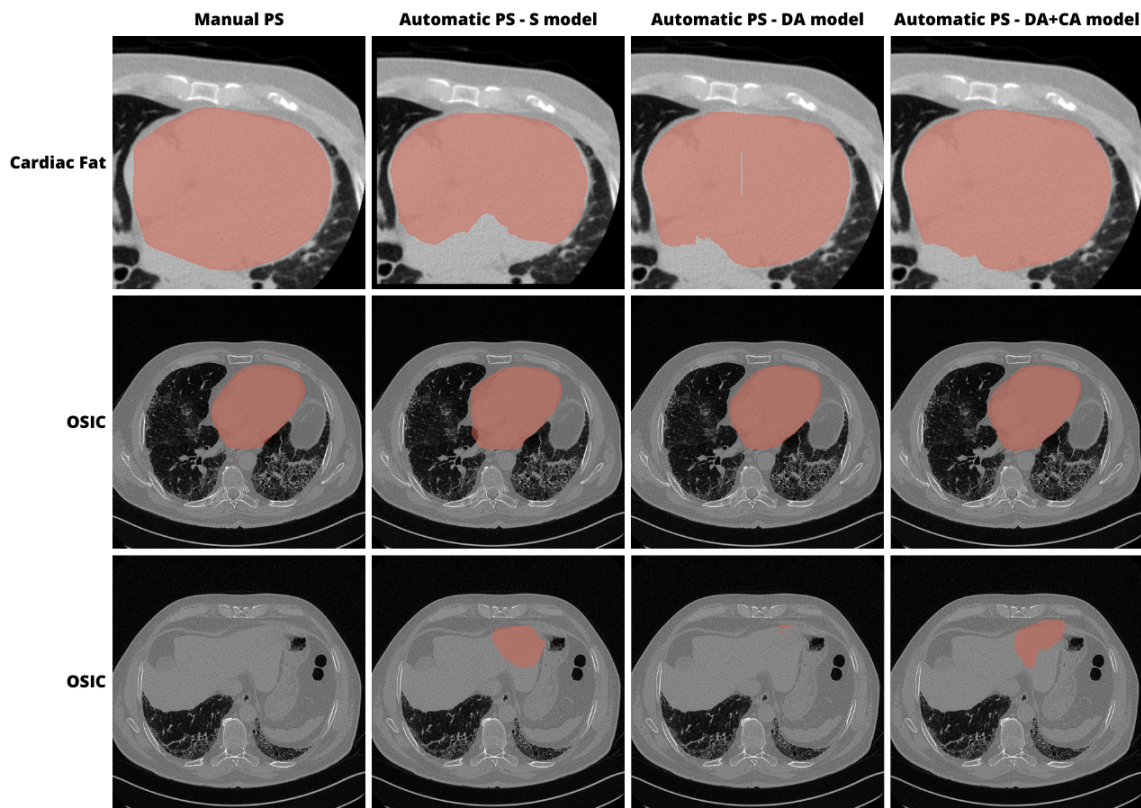


Figure 5.7: Examples of the effect of each one of the techniques and their combination, on two patients from the public datasets.

In summary, the training of the model with the combination of the two techniques, provides a balanced approach for the Cardiac Fat dataset where the DSC remains almost the same but the HD improves relatively to the *DA* model and decreases comparing to the *Standard* model. For the OSIC dataset the results were worse for almost all of the metrics. However, as can be seen by the example given by the Figure 5.7, these results can be explained by the difficulties of the model to segment the pericardium in lower slices where other organs are present. It is important to highlight that the model's performance when the pericardium is present is quite accurate. In fact, in certain cases, the model even achieves a more precise delineation of the pericardium compared



to the other two models. However, a slight decrease in performance was expected considering that the public datasets do not contain actual calcifications. Therefore, their inclusion to the training data can be viewed as the implementation of some noise for these specific datasets.

Regarding now the results for the private dataset presented in the Figure 5.4, it is possible to see that the combination of both techniques provides an improvement of performance. Indeed, when considering the DSC, the combined approach yielded the best result thus far with a value of  $0.893 \pm 0.031$ , surpassing the performance achieved by the two other models. Besides that the boxplot shows a more precise and consistent performance across all patients, indicated by a greater concentration of values towards the upper side of the graph and a decrease in the number of outliers. The HD exhibited a worse value than both the other models with a value of  $84.093 \pm 35.431$  mm. However the MAD showed improvement compared to both models, yielding a value of  $7.052 \pm 3.184$  mm, although presenting many outliers.

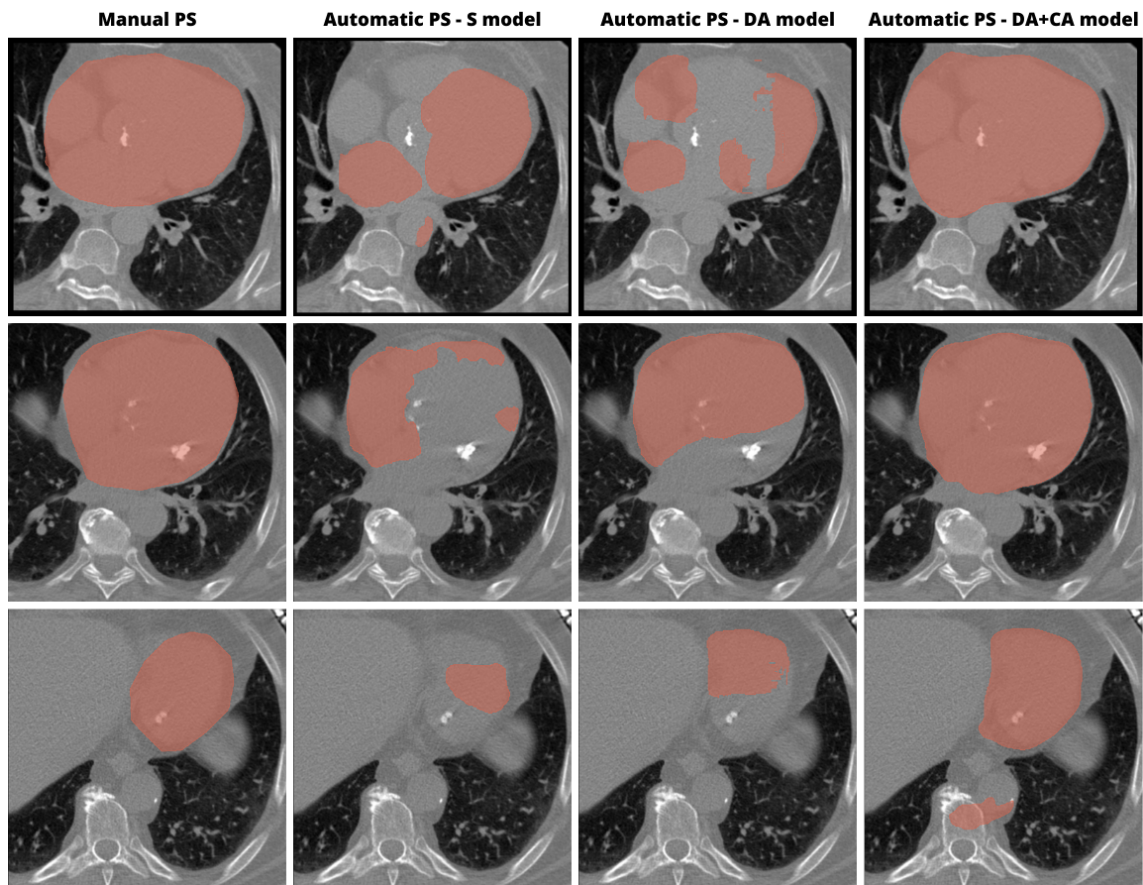


Figure 5.8: Examples of the effect of the combination of the artificial calcifications with data augmentation on two patients from CHVNGE dataset.

Three examples are demonstrated in the Figure 5.8 where the impact of the implementation of both techniques together was beneficial. In fact for the first patient it can be seen that the output from the *DA* model fails when calcifications are present in the CT, resulting in scattered components around the calcification. However, when the artificial calcifications are incorporated

into the model training, a noticeable change can be observed. Indeed, the *DA + AC* model now recognizes the inner region, including the calcification, as the pericardium, and a more accurate automatic segmentation is achieved. The second patient exhibits a similar scenario, where the *DA* model fails to identify the calcification as part of the inner region of the pericardium. However, when both techniques are combined in the training, this issue is reduced, resulting in a more accurate segmentation of the pericardium. The third example is a case which is yet common in the lower slices of the CT scans where the presence of neighboring organs can cause confusion for the model. This results in the segmentation of portions of other organs being mistakenly identified as the pericardium. Even the *DA + AC* model struggles to accurately delineate the pericardium in this case. The model, mistakenly identifies a portion of the descending aorta as a separate pericardium, indicating an error in its segmentation process.

## 5.4 Final Outcomes

In this chapter two techniques were employed to address the problems observed in chapter 4, mainly the incapacity to the model to deal with calcifications in patients from CHVNGE.

Therefore, artificial calcifications were created with various shapes and sizes using a 2D Gaussian distribution and added to the training data. Furthermore, data augmentation was incorporated in these studies.

In conclusion, the combination of both techniques was considered to be the most suitable approach for the automatic pericardium segmentation, as a more balanced performance was achieved for the three test sets. Besides that, it was achieved the best value for the metric of DSC on the external and more diversified dataset. However, there is still room for improvement as the model struggles to accurately segment the lower slices of the pericardium, where other organs may be present. The presence of disconnected components in the pericardium segmentation in these slices, may require additional post-processing techniques or the integration of more contextual information into the network. These measures could potentially address these issues and improve the values of the HD and MAD in particular. Therefore the next chapter will introduce a new approach aimed to overcome these issues.

## Chapter 6

# Three Dimensional Approaches

In the previous chapter, the model trained with the combination of artificial calcifications and data augmentation was considered the most suitable for an accurate pericardial segmentation, having general good performance according to the three metrics DSC, HD and MAD studied. However it was visible that the model fails sometimes in the lower slices where other organs are present. Additionally, when the patient's anatomy is unusual, the model may struggle to accurately identify the pericardium and may segment other structures besides the pericardium.

Therefore, this chapter aims to introduce a new approach to give more contextual information to the model. Besides that, pre and post-processing steps are proposed to deal with some inconsistencies found in the automatic segmentation of the pericardium.

### 6.1 Methods

#### 6.1.1 2.5D U-Net

In this chapter it is proposed to change the input of the 2D U-Net in order to give more contextual information to the network. To accomplish this, the U-Net is provided with three consecutive axial slices: the one to be classified and segmented ( $k$ ), as well as the previous ( $k - 1$ ) and next sections ( $k + 1$ ). Figure 6.1 illustrates this approach.

This model was named as a 2.5D U-Net rather than the 2D U-Net, because although the architecture and operations such as convolutions and padding are performed in 2D, the model utilizes 3D information by considering the upper and lower slices of each slice.

The training of 2.5D model was conducted similarly as the 2D described in the Chapter 4, with  $256 \times 256$  image size, the Dice Loss function,  $[-1000, 1000]$  HU range, and implementing both techniques discussed in the Chapter 5. However, it was decided to decrease the learning rate from  $1e-4$  to  $1e-5$  when the loss didn't improve for 10 consecutive epochs to avoid the increase of the loss observed in the ending epochs.

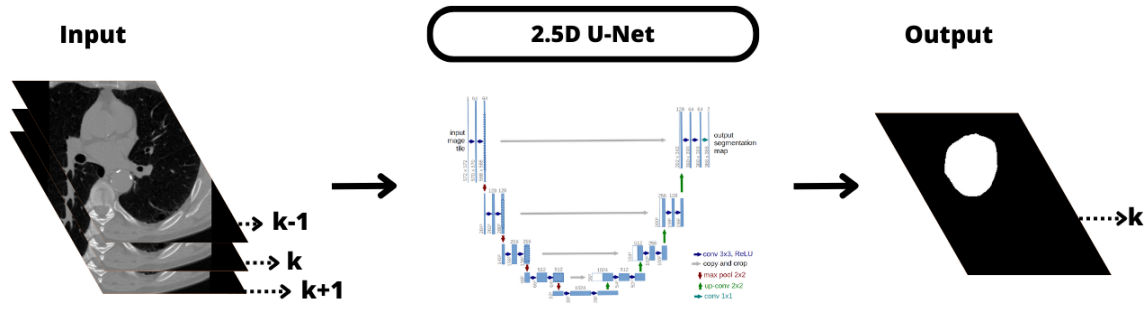


Figure 6.1: Example of the input and output of the 2.5D model.

### 6.1.2 Preprocessing

As mentioned before, one of the main issues of the 2D model was often the incapacity to recognize the limits of the heart on the CT scan, leading to an over segmentation of the pericardium. Therefore two different preprocessing steps were proposed to help the model to focus only on the slices containing the pericardium.

#### Slice Classification Network

The first strategy adopted was to classify each of the CT slices as containing or not the pericardium by using a simple 2D CNN [28], as the one presented in Figure 6.2.

The CNN purpose is for a classification task and instead of employing the Dice loss, the BCE loss function was utilized. The remaining steps of the process were similar to those conducted previously. However, of course, the weights were not optimized for the pericardium as described in the Chapter 4. The main issue was the unbalanced classes where (mainly due to the OSIC dataset) slices without the pericardium were more represented than the slices with pericardium. Therefore, an adjustment of the class weights was performed according to the ratio between the classes to address this issue.

After training the model, to limit the boundaries of the heart, the first and last slice containing the pericardium, were selected and all the slices in between were input to a pericardium segmentation model such as the 2D or 2.5D models proposed previously. This process is illustrated in the Figure 6.2. Consequently, throughout this document, this approach is denoted as the "slc+2/2.5D model."

#### 3D U-Net

Another solution tested was to use a 3D U-Net to segment the pericardium and help to set the boundaries. However, due to its high computational requirements, the CT scan had to be resized to a lower resolution ( $64 \times 64 \times 64$ ) to be entirely input into the model. The main objective of this 3D U-Net was not to provide an accurate 3D pericardial segmentation but rather to leverage the rich contextual information it offers to accurately identify the upper and lower boundaries of the pericardium. Similarly to the previous approach, the pericardium is then accurately segmented



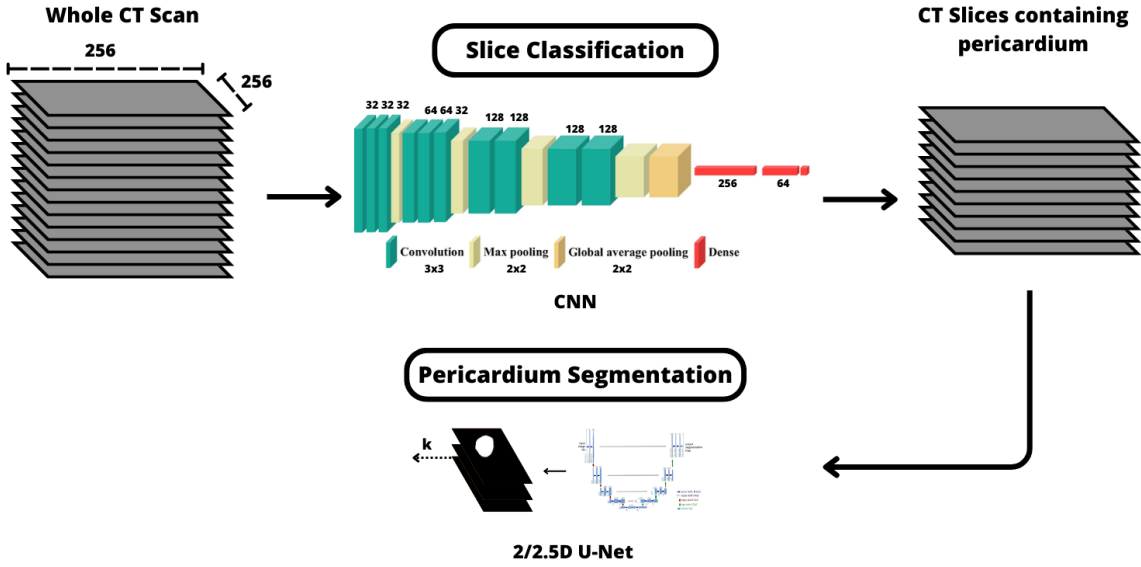


Figure 6.2: Illustration of the strategy of slice classification with the 2/2.5D model for pericardium segmentation.

using the 2/2.5D U-Net. This process is illustrated in the Figure 6.3. As a result, throughout this document, this approach is referred to as the "3D+2/2.5D model." The training process for the 3D model followed a similar approach to the 2.5D model, with a few differences. The BCE loss function was employed, and the input size was set as a  $64 \times 64 \times 64$  cube. Additionally, due to the demanding computational requirements of the task, a batch size of 1 was utilized, processing one CT scan at a time during training.

Occasionally, reducing the resolution of the CT scan posed challenges for certain patients. While the coronal and sagittal axes could be easily resized to 64, the axial axis encountered a few difficulties. In cases where patients had fewer than 64 slices, the empty spaces in the upper and lower limits were filled with black slices until the total count reached 64. On the other hand, some patients had their hearts cropped since the pericardium extended beyond the limited 64 slices. To address this issue, a down-sampling process was employed to ensure a complete representation of the pericardium.

### Preprocessing Evaluation

In order to evaluate the performance of the preprocessing step, an specific personalized metric named Mean Absolute Error Distance (MAED) was created. The MAED aims to quantify the error of the model in the task to correctly identify the upper and lower limits of the heart. Therefore, the MAED involves the calculation of two distances for each patient. The first distance measured the gap (in *mm*) between the first slice containing the pericardium in the manually segmented data and the corresponding slice in the predicted results. The second distance captured the difference between the last slices containing the pericardium in both the manual segmentation and the predicted outputs. Finally a mean of these two distances was performed.

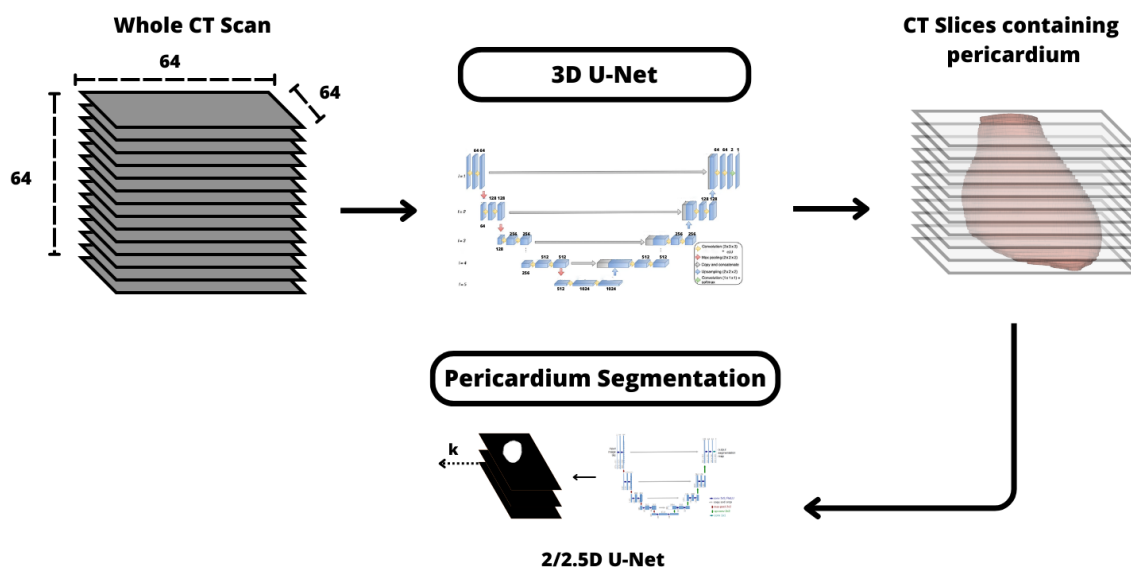


Figure 6.3: Illustration of the strategy of 3D segmentation for preprocessing with the 2/2.5D model for pericardium segmentation.

In the validation set, the MAED was utilized for threshold tuning in the  $slc + 2/2.5D$  model. Specifically, the threshold responsible for determining whether a slice belonged to the pericardium or not was optimized. The objective was to enhance the MAED and, as a result, reduce errors in accurately identifying the boundaries of the heart.

### 6.1.3 Post-processing

An issue identified in the previous chapters was the presence of inconsistencies in automatic pericardium segmentations. The pericardium should be represented as a single, solid, and connected component. However, the automatic segmentations often exhibit small disconnected components, particularly in the lower slices of the pericardium. The example in Figure 6.4 illustrates a 3D representation of a patient's heart, comparing the manual segmentation with the automatic segmentation obtained from the chosen model. The difference between the two segmentations is evident, with the automatic segmentation displaying numerous disconnected components that need to be removed for accurate representation of the pericardium. While most of these inconsistencies are noticeable in the lower pericardium, some are also present in the upper slices, as depicted in Figure 6.4. Therefore, the preprocessing step can help to remove these inconsistencies in regions where the heart is absent (upper and lower slices) but not in the regions containing the heart. Thus, a post-processing step may be necessary to handle the latter case.

To tackle these inconsistencies, three steps were implemented. Firstly, only the biggest 3D connected component was considered, while the disconnected components were discarded. Additionally, to maintain continuity in the segmentations, a 3D approach was implemented, which involves considering the each three consecutive slices, which we refer to as "Axial Fill". In this

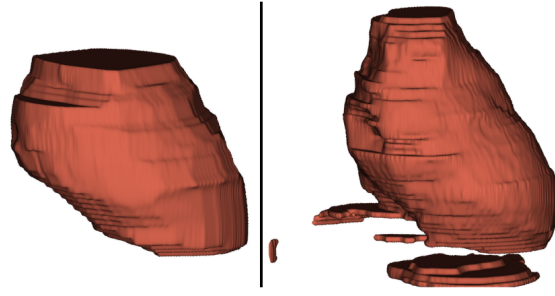


Figure 6.4: Example of a segmentation in 3D of a heart. Left: manual segmentation. Right: automatic segmentation from the 2D model.

step, if a pixel was segmented in the upper and lower slices, it was also included in the intermediate slice to ensure continuity. Finally, a 2D convex hull was applied to each slice, filling certain holes and smoothing the segmentation. All these steps are presented in the Figure 6.5. Models that incorporate post-processing techniques are denoted with a "+pp" at the end of their names.

## 6.2 Experiments

In this chapter, three main experiments were done to evaluate the effect of giving more contextual information to the network for different purposes (direct segmentation or preprocessing step).

First, a direct comparison of the performance of the pericardium segmentation task was conducted between the previous 2D model (*DA + AC*) and the 2.5D model presented in this chapter. Next, the two different preprocessing approaches to limit the upper and lower limits of the heart were discussed and evaluated as well as the final predicted segmentation by using each one of them. Finally, the effect of the post-processing techniques were studied.

Once again, all the three metrics (DSC, HD, MAD) were used to evaluate the performance's model. However, during the preprocessing evaluation, MAED was also used to assess the performance of these approaches.

## 6.3 Results and Discussion

### 6.3.1 2D Model vs 2.5D Model

Figure 6.6 allows us to compare the results for the 2D and 2.5D model across all three datasets.

Regarding the public datasets it is notorious that the 2.5D model outperformed the 2D model in all three metrics. In fact, for both datasets, the DSC achieved the best results until now with  $0.955 \pm 0.009$  and  $0.930 \pm 0.016$  for Cardiac Fat and OSIC, respectively. Besides that, the MAD achieved also the best results among all the models with  $2.591 \pm 0.803$  mm for the Cardiac Fat dataset. Paying more attention to the boxplots and more precisely to the HD for the OSIC dataset, one exception is observed for one outlier patient which results were better for the 2D model, but even with that the mean HD value was better for the 2.5D model.

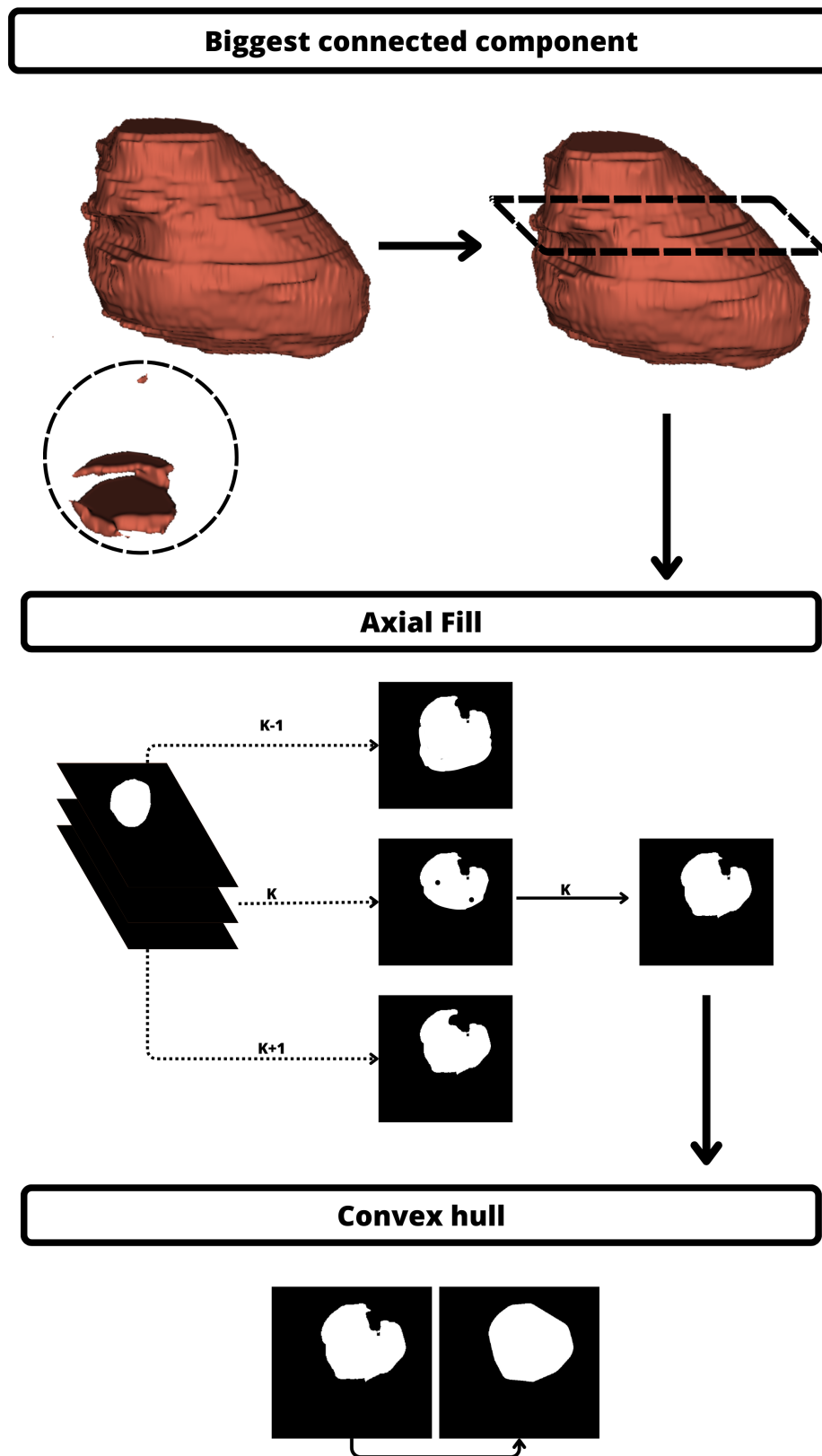


Figure 6.5: Illustration of the post-processing steps including: the selection of the biggest component component, application of the "Axial Fill" in consecutive slices and application of the Convex hull for each slice.

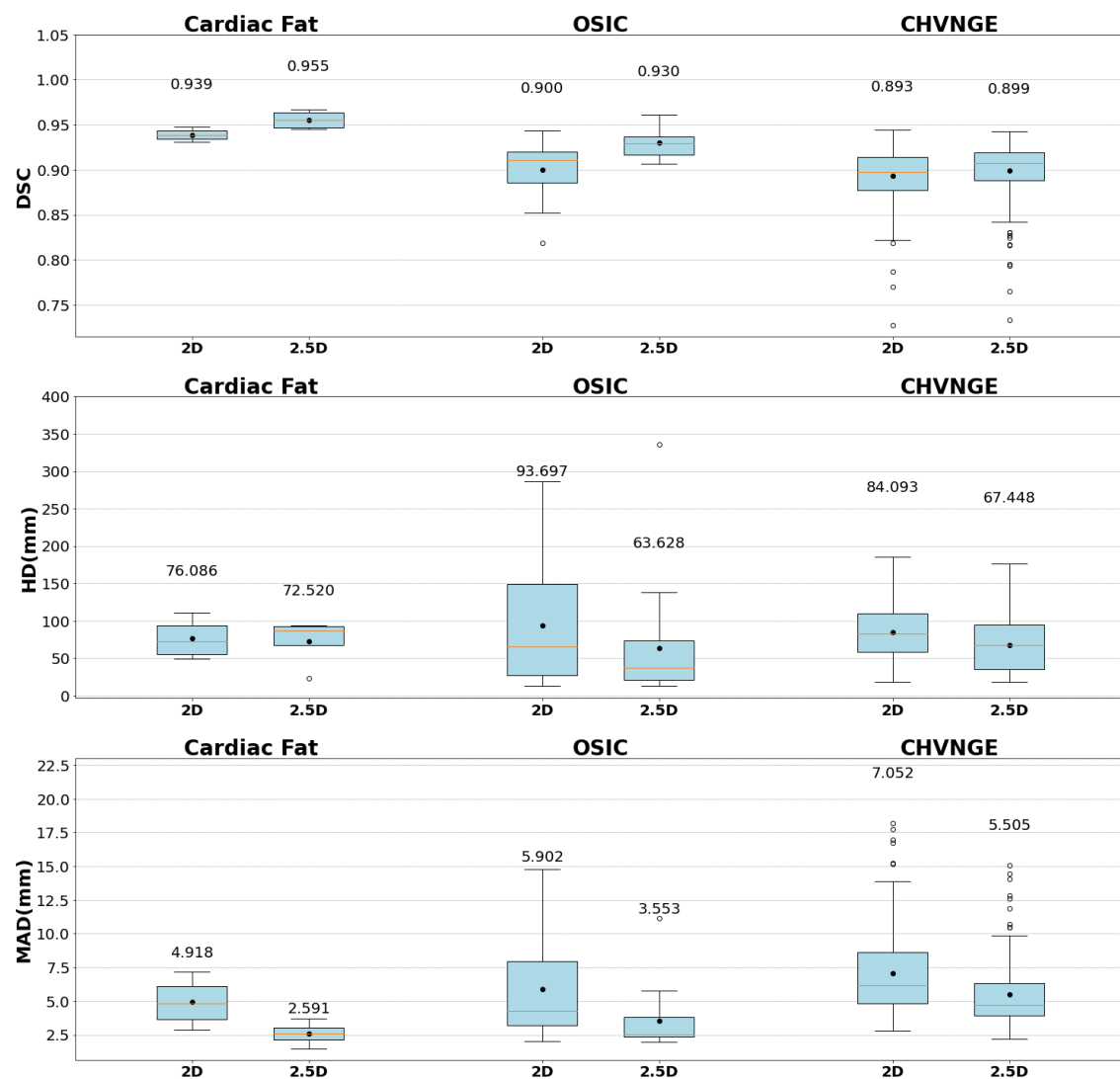


Figure 6.6: The results of the 2D and 2.5D for the three datasets, where the three metrics are displayed using boxplots, with the mean depicted as a point with the corresponding value written on the upper side.

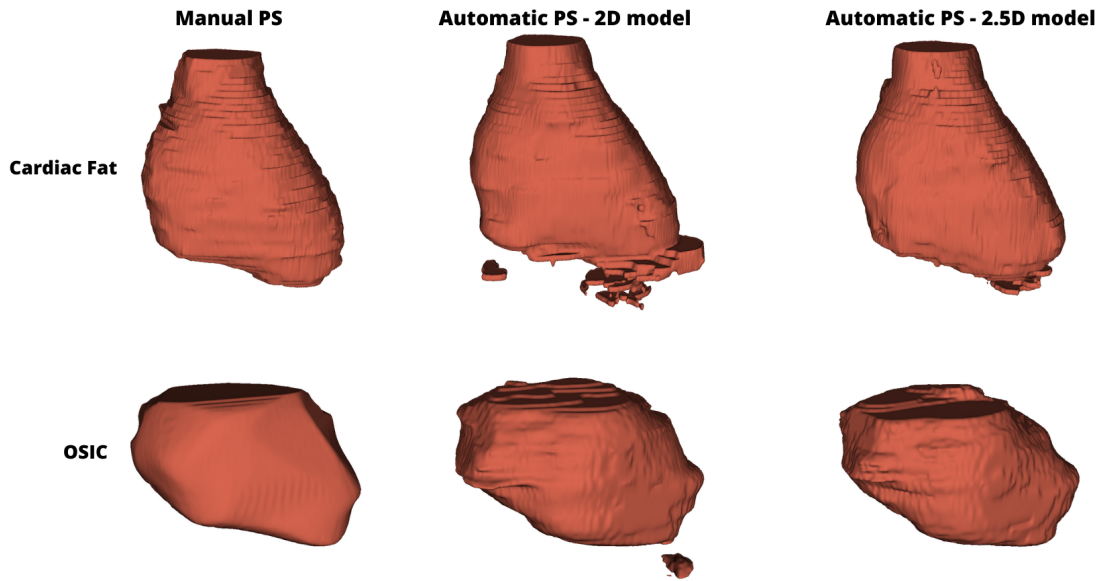


Figure 6.7: Examples of automatic pericardium segmentations using the 2D and 2.5D model for the public datasets.

In Figure 6.7, two instances in 3D are depicted to demonstrate the advantages of employing the 2.5D model. In fact for the Cardiac Fat dataset it is evident that the segmentation performed by the 2.5D model is better, because it exhibits fewer inconsistencies and eliminates certain disconnected components in the lower pericardium, when compared to the 2D model. Similarly, the second example, concerning the OSIC dataset, follows the same pattern, with the key distinction being that the 2.5D model effectively avoids segmenting disconnected components.

Paying now attention to the CHVNGE dataset, once again, the 2.5D model performed better than the 2D model, achieving the best values for all three metrics compared to previous models. Specifically, the DSC reached a value of  $0.899 \pm 0.031$ , the HD measured  $67.448 \pm 33.41$  mm, and the MAD was calculated as  $5.505 \pm 2.33$  mm.

Figure 6.8 present two examples to illustrate the superior performance of the 2.5D model. Indeed, the first example demonstrates that the segmentation provided by the 2D model shows some disconnected components below the pericardium which are not present in the 2.5D model. Besides that, both automated segmentation methods, when compared to the manual segmentation, demonstrate an oversegmentation of the pericardium, particularly in the upper slices. The second patient is a clear example of the 2.5D model's superiority. In fact, the 2D model fails to accurately identify the pericardium, especially in the lower pericardium region, certainly identifying other organs as the pericardium. Otherwise, while the 2.5D model may have some flaws and disconnected components, it still exhibits significantly better performance than the 2D model in correctly identifying the pericardium.

In summary, the 2.5D model outperformed the 2D model in all the parameters for both the public and private datasets. However, certain issues were observed with the 2.5D model, including

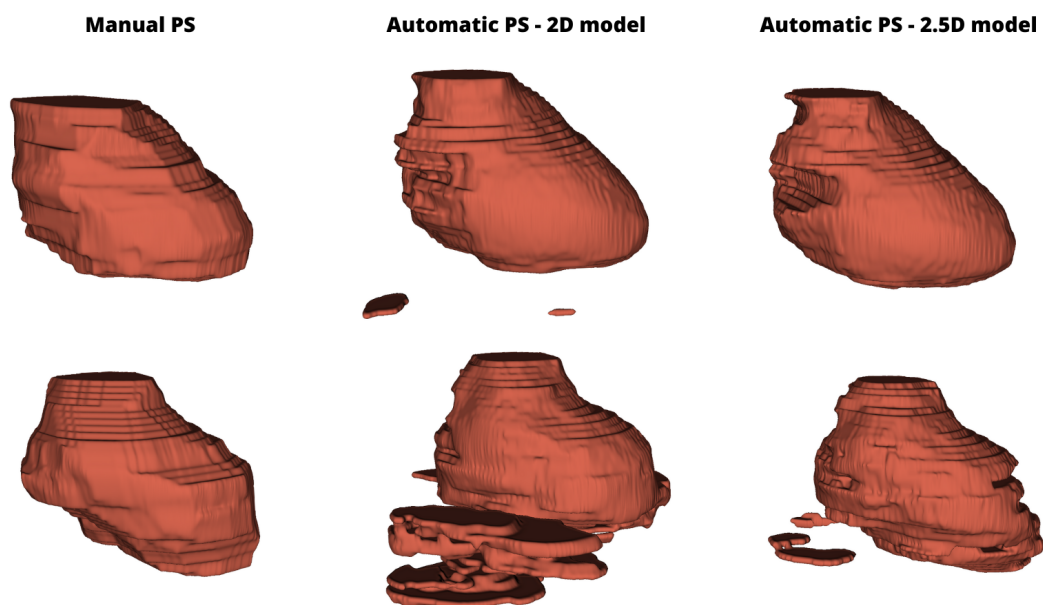


Figure 6.8: Examples of automatic pericardium segmentations using the 2D and 2.5D model for the CHVNGE dataset.

oversegmentation in the upper slices and occasional presence of disconnected components in the lower pericardium.

### 6.3.2 Preprocessing

As introduced before, the MAED was also used in the process of evaluation and discussion to determine the impact of the preprocessing step in limiting the boundaries of the heart. The results are present in the Table 6.1.

Table 6.1: Results of MAED for all the three test sets. Results shown in mean  $\pm$  standard deviation mm.

Dataset	2.5D	Slc	3D
Cardiac Fat	$8.625 \pm 8.172$	$1.875 \pm 0.651$	$1.500 \pm 1.836$
OSIC	$21.941 \pm 45.735$	$11.458 \pm 7.703$	$12.875 \pm 9.460$
CHVNGE	$10.674 \pm 7.473$	$12.234 \pm 7.122$	$8.337 \pm 5.472$

Regarding the public datasets, the lower values of the MAED metric revealed that the preprocessing methods played a vital role in accurately defining the boundaries of the heart in both datasets. Indeed, for the Cardiac Fat dataset an improvement of near 7 mm was observed when using both preprocessing approaches compared to the 2.5D model. Although both approaches yielded similar results, the *slc* + 2.5D model exhibited a lower standard deviation. In the OSIC dataset, the same trend was evident, with the preprocessing steps leading to an improvement of around 10 mm, particularly when utilizing the *slc* + 2.5D model.

During the analysis of the MAED in the CHVNGE dataset, it becomes evident that, unlike the public datasets, the results obtained from all approaches are more similar, and the preprocessing steps do not seem to offer significant advantages. However, the  $3D + 2.5D$  model exhibits an improvement of approximately 2 mm in MAED when compared to the raw  $2.5D$  model. Conversely, the  $slc + 2.5D$  model shows a decrease in performance of around 2 mm.

Figure 6.9 displays the segmentation results for the  $2.5D$  model, the slice classification approach and the  $3D$  model approach across all three datasets.

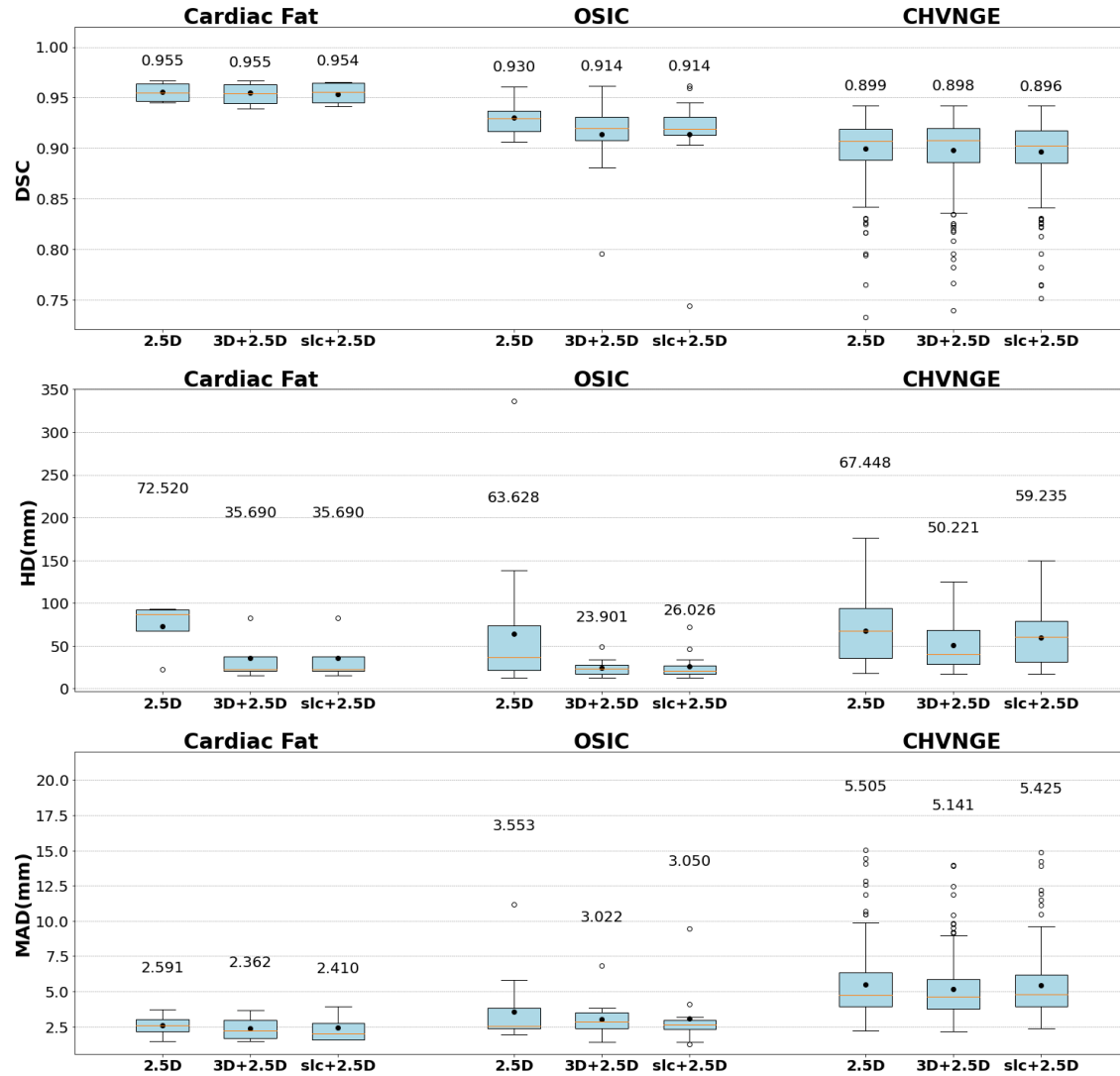


Figure 6.9: The results of the  $2.5D$  model along with preprocessing approaches for the three datasets, where the three metrics are displayed using boxplots, with the mean depicted as a point with the corresponding value written on the upper side.

Regarding the segmentation results on the public datasets, the preprocessing approaches mainly improved the performance of the HD and MAD, as expected. In fact, for the Cardiac Fat dataset, the HD improved significantly, decreasing from  $72.520 \pm 29.069$  mm to  $35.690 \pm 27.047$  mm when using both preprocessing approaches. These results are in agreement with the



MAED, where both preprocessing steps significantly reduce the MAED when compared to the 2.5D model. Regarding the OSIC dataset, a decrease in performance can be observed for the DSC when the preprocessing is applied. Nevertheless, the HD and MAD show remarkable improvement, particularly with the utilization of the  $3D + 2.5D$  model, achieving values of  $23.901 \pm 9.073$  mm and  $3.022 \pm 1.226$  mm respectively.

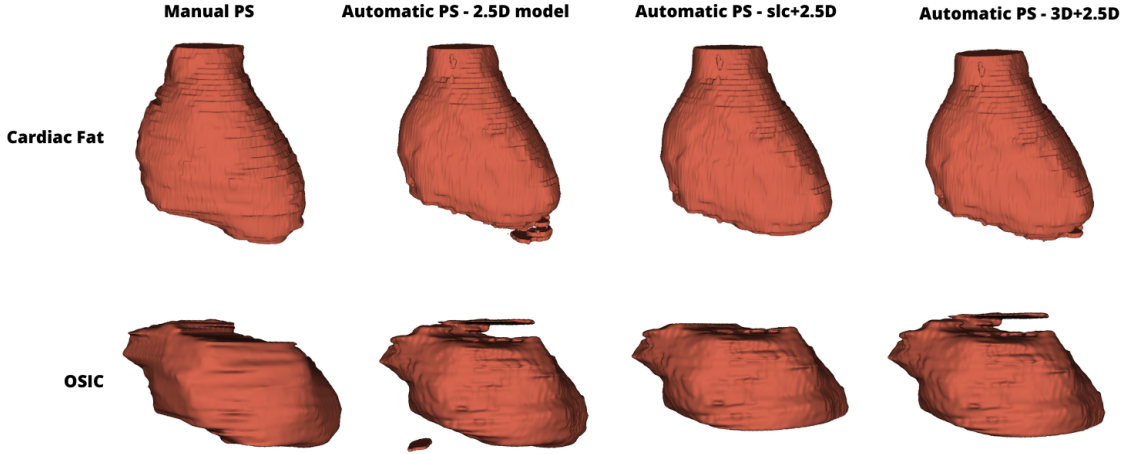


Figure 6.10: Examples of the effect of both preprocessing approaches on the automated pericardium segmentation for the public datasets.

Figure 6.10 shows the effect of each one of the preprocessing approaches for two patients from the public datasets. Regarding the example for the Cardiac Fat dataset, it can be observed that both preprocessing approaches performed well in identifying the lower pericardium in this case, avoiding the segmentation of some disconnected components carried out by the original 2.5D model. In the second example of the OSIC dataset, it is evident that both preprocessing models excessively cropped the pericardium region. This could explain the lower DSC observed for the  $slc + 2.5D$  and  $3D + 2.5D$  models. However, by cropping the pericardium region, the disconnected components in the lower pericardium are also removed, potentially resulting in lower HD and MAD values for both preprocessing models.

Turning now to the CHVNGE dataset, a slightly decrease in performance can be seen when attending to the DSC, mainly for the  $slc + 2.5D$  model. However, both HD and MAD showed a great improvement when the preprocessing approaches were employed. In fact, the HD and MAD reached a value of  $50.221 \pm 26.704$  mm and  $5.141 \pm 2.016$  mm, respectively for the  $3D + 2.5D$  model, the best results until this moment. This outcome is in agreement with the MAED results, as lower values are obtained when the  $3D + 2.5D$  model was employed. However, if we focus on the MAED for the  $slc + 2.5D$  model an apparently contradictory information is given, as the value for the 2.5D model is better and the HD and MAD is worse for this model.

By paying attention to the first example of the Figure 6.11 it can be seen that both the 2.5D model and the  $3D + 2.5D$  model achieved similar results in segmenting the pericardium. However, the preprocessing step fails slightly in correctly identifying the upper limit of the pericardium, resulting in an oversegmentation. Nonetheless, overall, both models exhibit satisfactory segmenta-

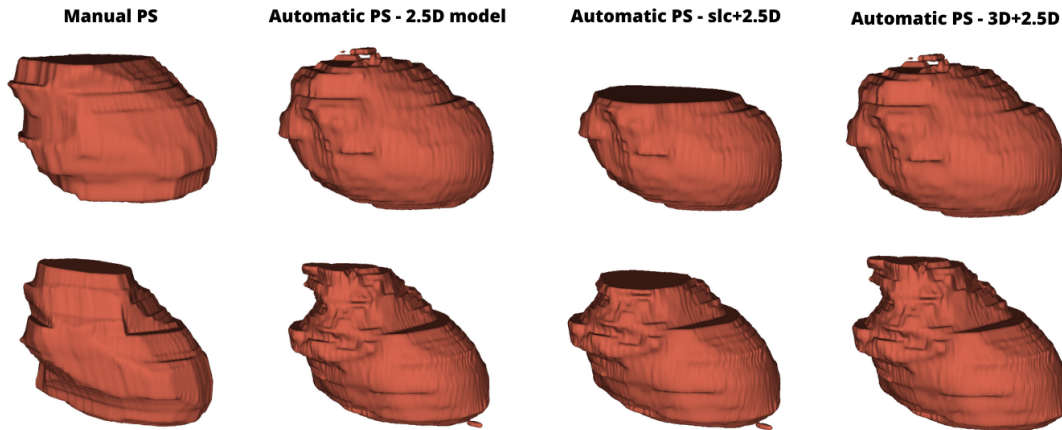


Figure 6.11: Examples of the effect of both preprocessing approaches on the automated pericardium segmentation for the CHVNGE dataset.

tion performance. On the other hand, the automatic segmentation of the  $slc + 2.5D$  model demonstrates the opposite issue as it fails to recognize the pericardium in the upper slices leading to an undersegmentation. This can explain the worse results in MAED compared to the 2.5D model as the distance between the first slice containing the pericardium of the ground truth and the predicted one would be higher. Moreover, it is reasonable to expect that the HD metric showed better results for the  $slc + 2.5D$  model compared to the 2.5D model, as segmenting less pericardium in the upper slices reduces the distance between the automatic segmentation and the ground truth, since the other models tend to oversegment the pericardium. The second example shows that the automatic segmentation carried out by the  $3D + 2.5D$  model seems to be the best one as the preprocessing employed prevents the segmentation of some disconnected components in the lower pericardium presented in the automatic segmentations produced by the other models. Once again, the  $slc + 2.5D$  model exhibits insufficient segmentation of the pericardium in the upper slices.

In summary the preprocessing approaches appear to help to recognize the limits of the heart, particularly when utilizing the 3D segmentation approach prior to segmenting with the 2.5D model. This was evidenced through better values of MAD and HD for both the public and private datasets. However, the DSC exhibited a slight decrease in performance across all datasets, especially when employing the slice classification approach. Specifically, the  $slc + 2.5D$  model failed to classify certain slices as pericardium in the upper sections, resulting in a lesser extent of segmented pericardium than desired. After the exploration of the examples of the Figures the main advantage of the  $3D + 2.5D$  model compared to the 2.5D model is that the first one is able to avoid the segmentation of some disconnected components of the lower slices.

### 6.3.3 Post-processing

After exploring both preprocessing approaches, it was decided to select the  $3D + 2.5D$  model as it yielded better results for all the metrics, resulting in the exclusion of the slice classification approach from further consideration. However the 2.5D model shows better results for the DSC

and it was seen that the main advantage of the 3D segmentation preprocessing approach was its ability to prevent the segmentation of disconnected components in the lower slices. Therefore, it was decided to apply the post-processing for both automatic segmentations: the original 2.5D model and the preprocessed one (3D + 2.5D model). Figure 6.12 displays the results for the 2.5D model, the 3D + 2.5D model before and after post-processing across all three datasets.

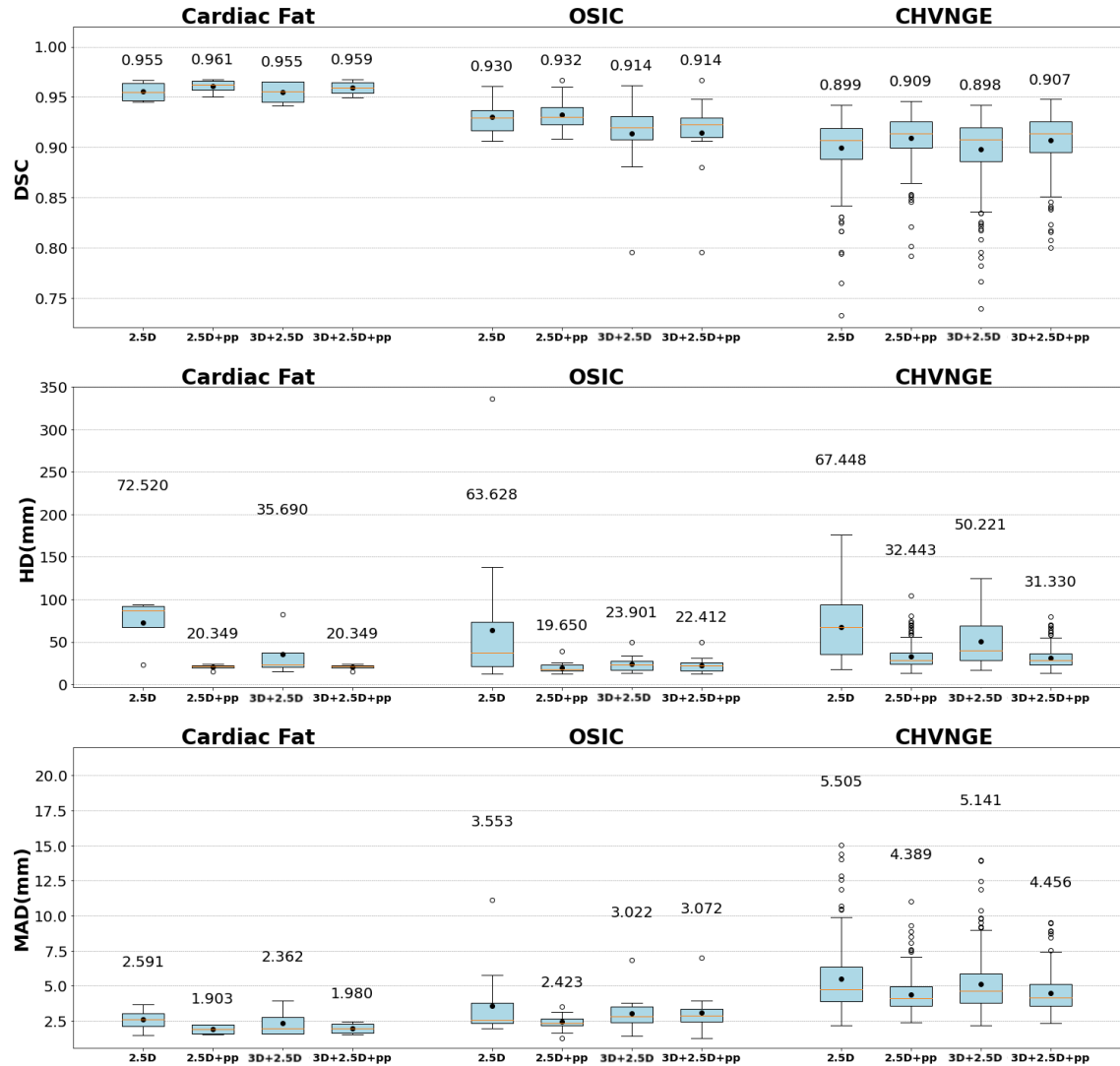


Figure 6.12: The outcomes of the 2.5D model and 3D + 2.5D model, combined with post-processing, are presented for the three datasets. Boxplots are utilized to visualize the three metrics, with the mean represented as a point accompanied by the corresponding value indicated on the upper side.

Regarding the public datasets, the application of post-processing directly to the 2.5D model has proven to be incredibly effective, resulting in the best values obtained thus far for all three metrics across both datasets. In fact, for the Cardiac Fat dataset, the DSC, HD and MAD reached the values of  $0.961 \pm 0.007$ ,  $20.349 \pm 3.277$  mm and  $1.903 \pm 0.345$  mm, respectively. These results indicate that this approach slightly outperformed the 3D + 2.5D model with post-processing. The same

trend is observed for the OSIC dataset, where this approach demonstrated the best performance achieved thus far across all metrics.

Figure 6.13 illustrates the effectiveness of employing the post-processing step for the public datasets. In the first example from the Cardiac Fat dataset, it is evident that the post-processing step successfully eliminates the disconnected components present in the lower slices of the automatic segmentation performed by the original 2.5D model. Besides, it seems that the post-processing step appears to provide a smoother surface for the segmentation, which is expected due to the application of the Convex Hull to each slice. In this case the segmentations provided by the models with and without preprocessing, after the application of the post-processing step looks quite the same. The same outcome can be extracted for the second example from the OSIC dataset.

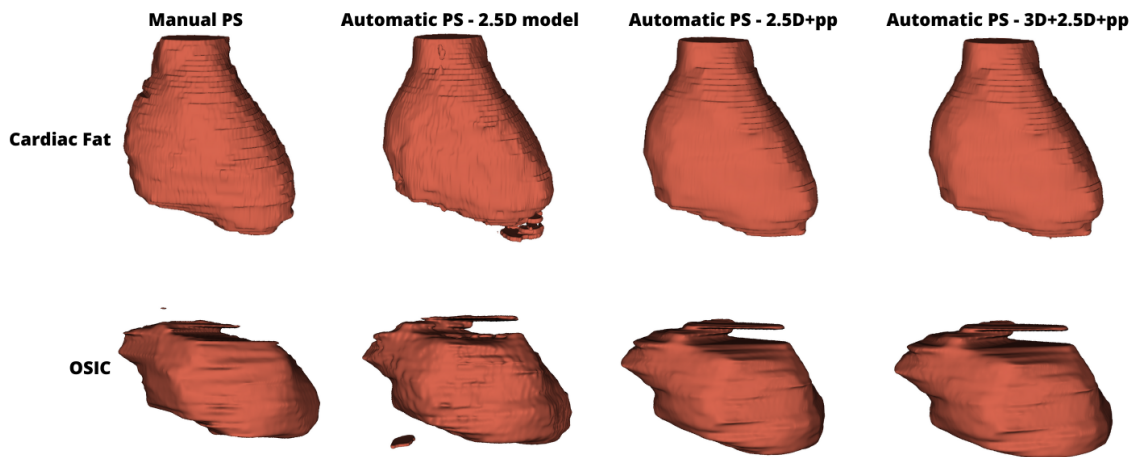


Figure 6.13: Examples of the impact of the post-processing approach on the automatic pericardium segmentation for the public datasets.

After analysing the results for the CHVNGE dataset, the same conclusion can be drawn. Indeed, the post-processing step improves the results for all the three metrics, achieving the best performance thus far. Directly applying the post-processing to the segmentations generated by the 2.5D model yields better results compared to applying the post-processing after the segmentations produced by the  $3D + 2.5D$  model. The only exception was observed in the HD, where the  $3D + 2.5D + pp$  model demonstrated a slightly better outcome than the  $2.5D + pp$  model. In addition, it appears that the  $3D + 2.5D + pp$  model exhibits slightly reduced variability and noticeably better results for a particular patient. However, the mean values are slightly worse, and the computational requirements are higher. Consequently, the  $2.5D + pp$  model yielded generally better results achieving for the DSC, HD, and MAD, the values of  $0.909 \pm 0.024$ ,  $32.443 \pm 13.729$  mm, and  $4.389 \pm 1.291$  mm, respectively. This approach was considered the most effective method for achieving accurate pericardial segmentation thus far.

Figure 6.14 helps to understand the beneficial impact of the post processing when applied over the automatic segmentations provided by the two models. Indeed, the first patient serves as a excellent example of the effectiveness of our work and approaches thus far. This particular patient is the same individual featured in Figure 6.8, where the application of the 2D model

resulted in highly inaccurate segmentation, characterized by numerous inconsistencies, particularly in the lower pericardium region. However, with the introduction of the 2.5D model and subsequent implementation of post-processing, as demonstrated in Figure 6.14, a substantial improvement is observed. The post-processing step proves effective in removing the disconnected components present in the 2.5D model's segmentation and contributes to a smoother surface, by filling holes. The segmentations produced by the  $3D + 2.5D + pp$  model closely resemble those of the  $2.5D + pp$  model. The second patient is another example of the evident beneficial impact of the post-processing step. The third example shows a particular case where preprocessing step was highly beneficial. Indeed, the  $3D + 2.5D + pp$  model didn't identify certain lower slices as the pericardium. This outcome prevented the missegmentation of those slices, which improved the segmentation performance compared to the 2.5D model.

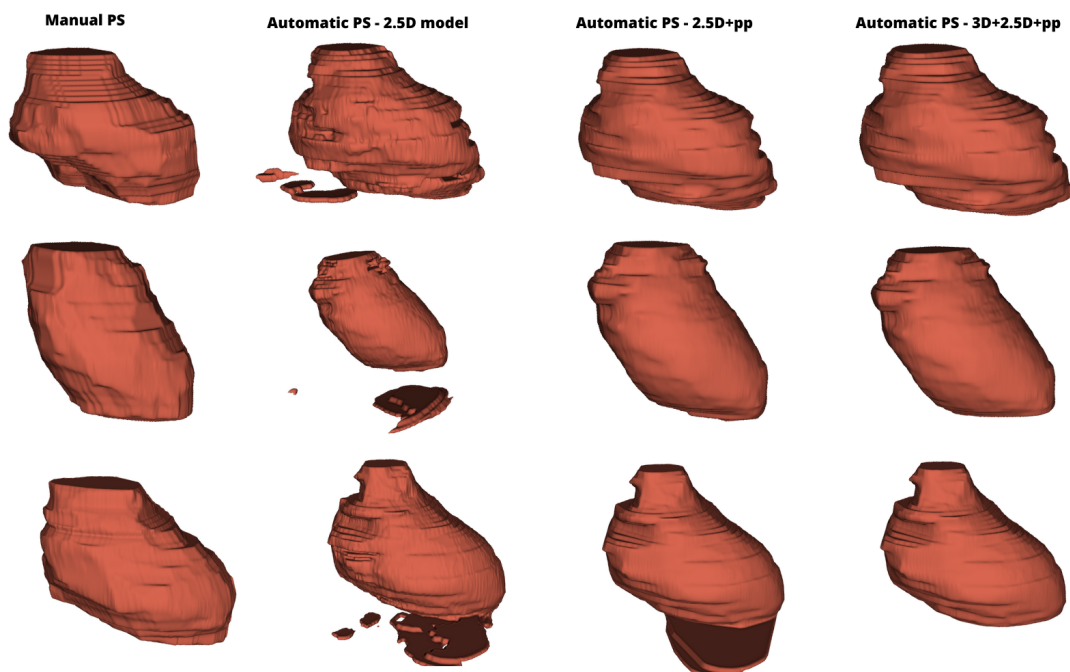


Figure 6.14: Examples of the impact of the post-processing approach on the automatic pericardium segmentation for the private dataset.

To summarize, the post-processing step has demonstrated remarkable effectiveness, achieving the best performance thus far across all metrics for both public and private datasets. The examples depicted in the figures illustrate the benefits of post-processing, including the removal of disconnected components in the lower pericardium and the filling of holes, resulting in a smoother surface. Comparing the performance of the  $3D + 2.5D + pp$  model to the  $2.5D + pp$  model, they exhibited very similar results, with a slight advantage in favor of the  $2.5D + pp$  model. Additionally, the  $3D + 2.5D + pp$  model demands more computational resources as it involves the utilization of two models (3D and 2.5D), making the  $2.5D + pp$  model the more suitable choice for achieving accurate pericardial segmentation.

## 6.4 Final Outcomes

This chapter utilized three-dimensional methods to tackle the issues identified in chapter 5, specifically the challenges faced by the 2D model in accurately segmenting the lower pericardium and determining the upper and lower boundaries of the heart. The incorporation of additional contextual information offered by three-dimensional approaches proved advantageous in assisting the model with this task.

In overall, the 2.5D model outperformed the 2D model in all three datasets for all metrics and demonstrated superior capability in accurately identifying the pericardium in the lower slices. The problem of disconnected components and artifacts, especially in the lower slices, was successfully addressed using an efficient post-processing step. The application of a preprocessing step to identify the heart's boundaries did not prove to be particularly beneficial. Although it accurately detected the lower limit of the heart, the main segmentation issues in this region were the disconnected components. Consequently, directly applying the post-processing to the segmentations of the 2.5D model was less computational demanding and yielded the most favorable results.

It is important to note that the model still struggles in segmenting patients with significant anatomical variability, such as a disproportionate size between the heart and lungs or between other organs. However, as this variability was not present in the training datasets, the only way to optimize performance in such cases is by providing the model with more diverse data to learn from. It is worth noting that the reported results were achieved using the CHVNGE dataset exclusively for testing, which comprised a substantial number of 190 patients.



## Chapter 7

# EAT Segmentation and Quantification

Recapping to the introduction in the Chapter 1 and according to many different studies, the EAT is an adipose tissue that can be related to CAD and can assist in risk stratification and the prediction of MACE. In fact, the volume and thickness of the EAT can be associated with Coronary Calcium (widely employed CAD indicator) but recent studies showed that it can be an independent predictor of CAD, thus the interest on its segmentation.

However, the EAT refers to the fat located within the enclosed area of the pericardium. Therefore, it is extremely important to accurately segment this structure prior to proceeding with EAT segmentation. The preceding chapters have focused on this particular objective, and after conducting numerous experiments, a definitive strategy has been proposed for segmenting the pericardium from CT images.

With the pericardium segmentation completed, this chapter concentrates on the subsequent task of segmenting and quantifying the EAT, which aligns with the objective of our study. The obtained results will be assessed through the analysis of segmentation and quantification metrics. Furthermore, the inter and intra-variability of the EAT and pericardium segmentation between two readers will also be examined. In this evaluation only the external private CHVNGE dataset will be employed.

## 7.1 Methods

In this section, the method to segment and quantify the EAT from the pericardium will be addressed, as well as the evaluation metrics used for both segmentation and quantification tasks.

### 7.1.1 EAT Segmentation

Throughout the previous chapters many strategies were adopted to segment the pericardium. In the last chapter it was concluded that the 2.5D model along with the post-processing step yielded the

best results thus far. As a result, this approach was selected as the optimal method for accurately and automatically segmenting the pericardium in our work.

The automatic segmentation of the EAT is performed similarly to the manual one. After identifying the pericardium, the adipose tissue range  $[-150, -50]$  HU that highlight the adipose tissue is applied, and all the fat inside is considered as the EAT. Figure 7.1 illustrates the process of the segmentation of EAT.

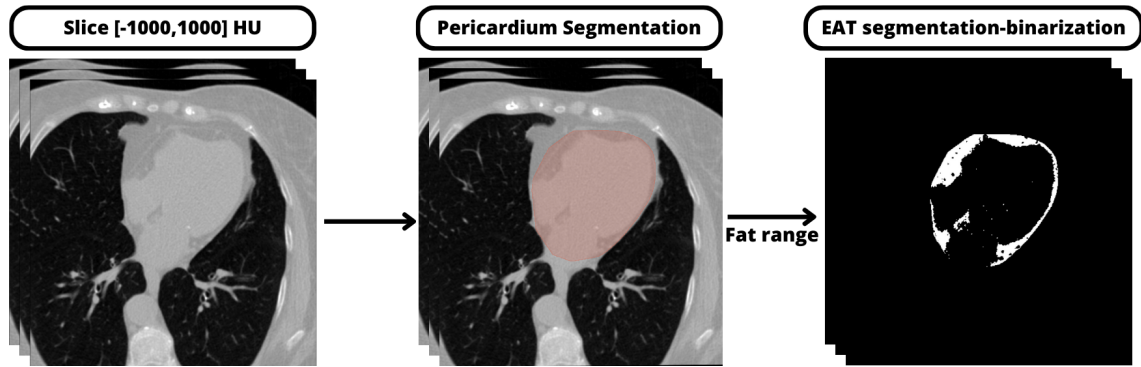


Figure 7.1: Process of segmenting the EAT from CT scans.

An important point in this chapter is the availability of more pericardial segmentations. In addition to the 190 CT scans that have been manually segmented by the Medicine student up to this point, a subset of 20 of these cases underwent a secondary manual segmentation by the same student. Similarly, these 20 CT scans were subjected to manual segmentation by a second specialist. This allows the examination of the intra and interreader variability among these patients. The goal is that the automatic method shows a lower or comparable variability, as identifying the pericardium poses a difficult task even for specialists. To simplify the identification of each person responsible for the manual segmentations, we will refer to the medicine student who provided the manual pericardium used as the reference until this chapter as "Reader 1," and the new specialist as "Reader 2."

### 7.1.2 Evaluation Metrics

Two types of metrics are utilized to evaluate and analyze the obtained results.

The first category of metrics are utilized to assess the agreement between two calculated volumes of EAT obtained from two distinct approaches. The volume of EAT is the most common way to quantify this fat and can be used as an indicator for risk stratification. By knowing the spatial dimensions of the voxels it is trivial to calculate the segmentation volume. The PCC (eq. 3.7) has been selected for its ability to visualize and compare the performance and relationship between the two volumes. The closer to 1 is the PCC, the better and closer is the relation between the two volumes from the approaches considered. Another strategy commonly used to evaluate the agreement between two variables is the Bland-Altman plot. In this plot, the difference between the volumes obtained from the two approaches is plotted against a reference volume. This graph can help to evaluate the trend and bias between the two volumes derived from the two approaches.



Although the prevalent and common approach in clinical contexts involves quantifying fat and establishing a direct CAD connection, the evaluation of segmentation remains significantly valuable. This is particularly significant due to the emergence of a promising technique called radiomics, which has gained traction as an effective means to extract a wide range of features. These features hold the potential to notably enhance prediction and classification results. Therefore it is important to accurately segment the EAT to extract these valuable features.

Consequently, the second type of metrics are particularly suitable for segmentation tasks, where the focus lies in accurately classifying individual pixels. Among these metrics, the chosen ones include the DSC (eq. 3.4), precision (eq. 3.2), and recall (eq. 3.3).

Throughout this work, the DSC has been the predominant metric used for segmentation tasks, helping in the evaluation and selection of the best model for pericardium segmentation. However, it is important to note that the pericardium and its inner region encompass a much larger structure compared to each individual voxel representing the EAT. Moreover, the incorporation or exclusion of certain fats that are not truly EAT can have an impact on the quality of the features extracted by radiomics. As a result, conducting a more comprehensive evaluation by analyzing the precision and recall specifically for the EAT would be beneficial, allowing a deep understanding of the performance of our approach.

## 7.2 Results and Discussion

The results for evaluation of the intrareader, interreader and automatic performance are depicted in the Figures 7.2, 7.3 and 7.4, respectively.

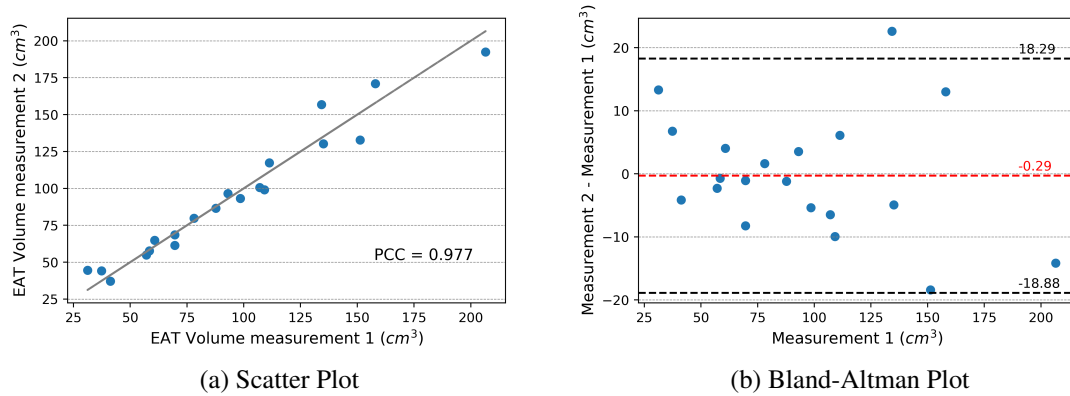


Figure 7.2: Volume quantifications from two measurements from Reader 1 to access intravariability. The gray line depicted in the Scatter plot illustrates the ideal scenario where the two measurements align perfectly, resulting in a PCC of 1.

Upon analyzing the intravariability from the Figure 7.2 is possible to see that, even between measures from the same Reader, certain variability is observed. The shows an obviously strong correlation between the two measures with a PCC of 0.977. However, the Bland-Altman is very useful to notice that the agreement between the two measures is far from perfect. Although a low

bias of 0.29 is observed, the 95% range of agreement is considerably large ranging from -18.88 to 18.29, indicating an absolute range of  $37.17 \text{ cm}^3$ . This implies that delineating the pericardium in a CT image is very challenging, even when performed by the same Reader, leading to significant uncertainty when establishing the boundaries of both the pericardium and, consequently, the EAT.

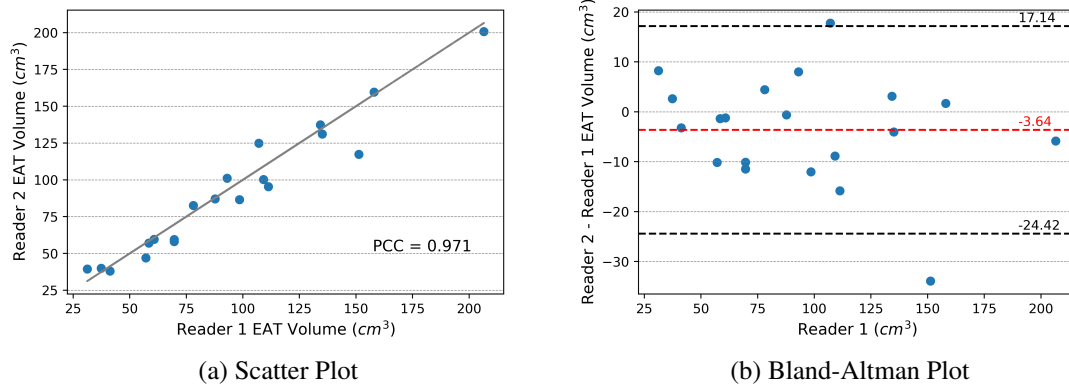


Figure 7.3: Volume quantifications from the Reader 2 versus the manual quantifications from the Reader 1 to assess intervariability. The gray line depicted in the Scatter plot illustrates the ideal scenario where the two measurements align perfectly, resulting in a PCC of 1.

Turning now to the intervariability, the depicted in Figure 7.3 reveals a strong correlation between the two volumes measured by the manual readers. In fact, the PCC achieved a value of 0.971 and the data points are close to the ideal gray line, representing the ideal behavior of total agreement with a PCC of 1. However, the Bland-Altman analysis reveals that the agreement between the two readers is not ideal. Indeed, there is a considerable bias of  $3.64 \text{ cm}^3$  observed, and the 95% range of agreement spans from -24.42 to 17.14, indicating an absolute range of  $41.56 \text{ cm}^3$ . Nonetheless, when compared to intravariability, a minimal difference in the absolute range (approximately  $4 \text{ cm}^3$ ) is detected, potentially implying a strong concordance between the readers.

After analysing the intra and interreader variability among 20 patients, the results prove that delimiting the pericardium is an extremely challenging task and leads to uncertainty even for the same reader across time. Consequently, and as expected, the disagreement is larger between different readers, given that they can be influenced by each expert's distinct background.

Now, examining our automated approach, the scatter plot in Figure 7.4 highlights a satisfactory correlation (PCC = 0.924) between Reader 1's EAT quantification and the automatic method. Notably, the data points are positioned closely to the gray line. The Bland-Altman graph shows us a presence of a small bias of  $0.98 \text{ cm}^3$  between the two calculated volumes. The 95% limits of agreement ranged from  $-29.11 \text{ cm}^3$  to  $31.07 \text{ cm}^3$  with an absolute range of  $60.18 \text{ cm}^3$ . These results indicate a reasonable agreement between the automatic approach and the quantification of the Reader 1. However, it remains a little distant from the intra and interreader variability ( $37.17$  and  $41.56 \text{ cm}^3$ ). It was an expected outcome considering the larger patient pool in this

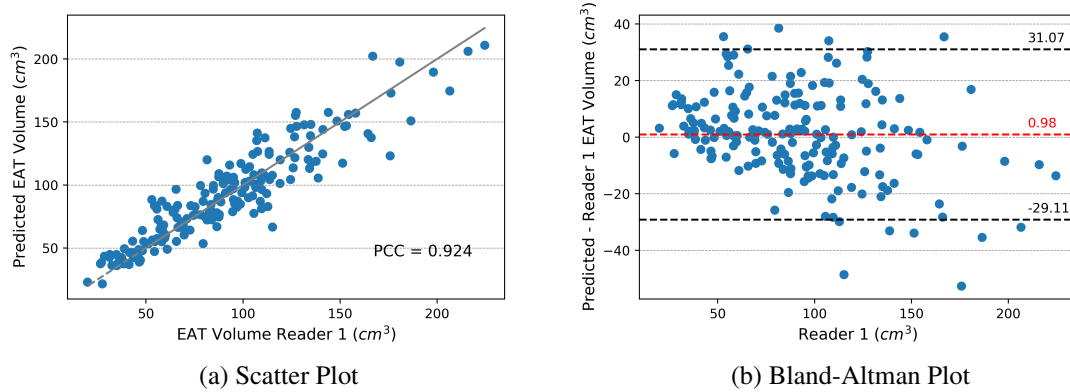


Figure 7.4: Automated volume quantifications versus the manual quantifications from the Reader 1. The gray line depicted in the Scatter plot illustrates the ideal scenario where the two measurements align perfectly, resulting in a PCC of 1.

study, including those with uncommon anatomical variations that have adversely impacted these findings.

Table 7.1 shows the segmentation results among the three experiments previously mentioned.

Table 7.1: Results of EAT segmentation metrics for intrareader, interreader and automatic approach. Results shown in mean  $\pm$  standard deviation (std).

Evaluation	DSC	Recall	Precision
Intrareader	$0.855 \pm 0.044$	$0.863 \pm 0.055$	$0.857 \pm 0.068$
Interreader	$0.798 \pm 0.046$	$0.788 \pm 0.067$	$0.817 \pm 0.062$
Automatic	$0.749 \pm 0.051$	$0.766 \pm 0.069$	$0.748 \pm 0.085$

Overall, the results for EAT segmentation are noticeably inferior compared to the pericardium ones. This discrepancy can be attributed to the smaller size, shattered and disjointed nature of the EAT, particularly concentrated along the pericardium borders. Thus, even a minor imprecision in segmenting a small part of the pericardium, which may contain a substantial amount of EAT, significantly impacts the overall accuracy of the EAT segmentation result.

Focusing on intrareader performance, it's evident that since both segmentations were carried out by Reader 1, these values are the highest. Nevertheless, none of these three metrics came close to 1, indicating a notable deviation from perfect agreement. As observed before with the quantification metrics, these results underscore the challenges associated with visualizing and accurately delineating the pericardium.

A substantial decline is noticeable across all three metrics when assessing the interreader outcomes in comparison to intrareader results. Specifically, the DSC shows a reduction exceeding 0.05, while the Recall experiences the most significant drop, with a decrease of approximately 0.09. These observations were slightly unexpected, given that the variability on the quantifications were not too pronounced. However, the variability of the segmentations, points to a consid-

erable disagreement between the readers in accurately defining the pericardium boundaries and consequently, EAT.

Finally, the automatic method shows the worst performance when compared to the intra and interreader variability. Among the metrics, recall exhibited the closest resemblance between the automatic method and interreader performance. This observation could indicate that the model's ability to accurately segment EAT is nearly equivalent to that of an independent reader. However, the lower precision value of approximately 0.07 in comparison to the interreader results suggests occasional instances where the model misidentifies other fats as EAT. The DSC achieved an intermediate value in relation to the other two metrics. Therefore, while our approach yielded satisfactory results, the model still cannot entirely replace an independent reader, neither reduce the interreader variability. However, it's important to note that the evaluation of the automatic method involved 190 patients, whereas the intra and interreader assessment involved only 20 patients.

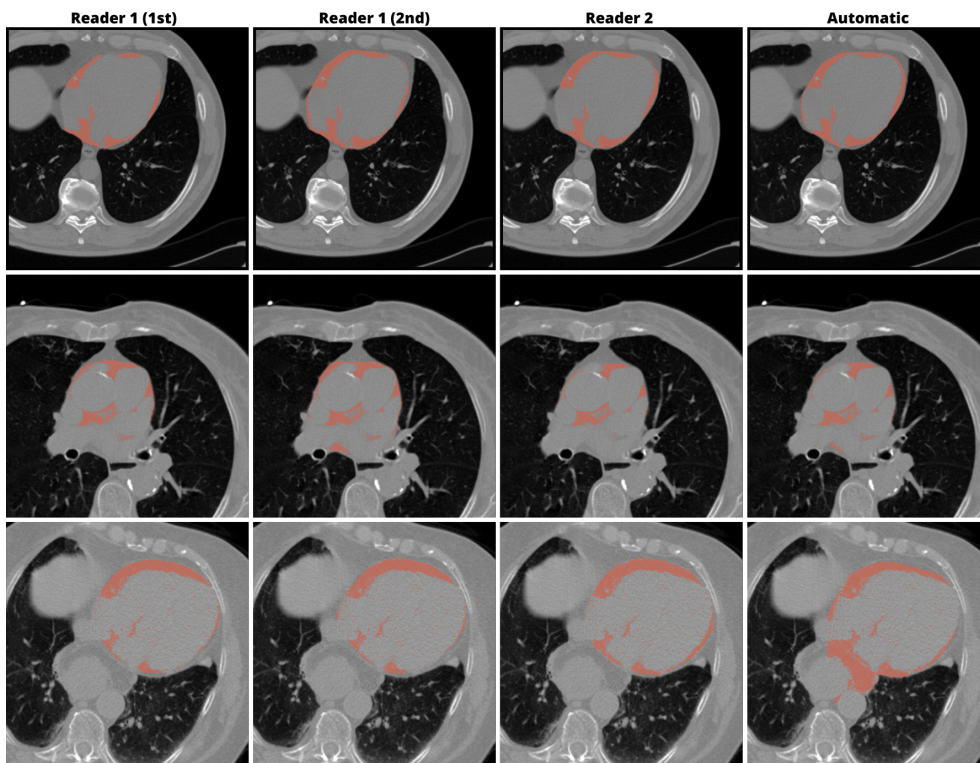


Figure 7.5: Examples of EAT segmentation performed by the two readers and the automatic method.

Figure 7.5 showcases three examples of EAT segmentation from different patients. The first example shows a patient where all segmentations appear quite similar. Nonetheless, upon closer and more detailed examination, it becomes apparent that in the second segmentation, Reader 1 identified a considerably greater volume of EAT than the first segmentation, indicating a higher intrareader variability. In this particular case, it appears that the segmentation of the automated approach is more similar to the one of the Reader 2. The second case represents a situation in which Reader 1 and Reader 2 carried out distinct EAT segmentations. Specifically, it can be seen that

Reader 1 identified more anterior and posterior fats as EAT, unlike Reader 2, leading to a greater interreader variability. The segmentation of the automatic method seems to be a balanced one between the interpretations of the two readers. The third example is a scenario where the automatic approach clearly fails to achieve accurate EAT segmentation. The example represents a lower slice with anatomical variations where another organ is present. It is possible that this represents an hiatal hernia which is a medical condition where a portion of the stomach protrudes through the diaphragm and into the chest cavity. Therefore the model incorrectly identified this portion as the pericardium, resulting in an inadequate and inaccurate EAT segmentation. This example shows the need to train the model using a more diverse range of data, including patients with anatomical variations that cannot be replicated through conventional data augmentation techniques.

Comparing our results with the existing literature is difficult due to the prevalence of private datasets, inaccessible to us, which involve varying numbers of patients. Furthermore, a considerable number of studies employ the same datasets for both training and testing their models, leading, unsurprisingly, to better results than the approach outlined in this document. An example is the work of Commandeur et al. (2019) [34] who made use of approximately 614 CT scans gathered from four distinct cohorts. However, they employed the same cohorts for both training and testing their model. Furthermore, only about 60 CT scans were allocated for testing purposes, resulting in a DSC of 0.871. Other example is the study of Siriapisith (2021) [42] where they utilized 220 CT scans, but limited testing to just 20 CT scans, all originating from the same cohort, leading to a DSC of 0.901. In a similar vein, Molnar et al. (2021) [40] leveraged 411 CT scans, yet only 25 were designated for model testing, yielding a DSC of 0.90, and once again these scans were sourced from the same cohort. The same pattern is observed from Qu et. al [28], where from the 103 CT scans available, only 33 were employed for model testing, reaching a DSC value of 0.883. In contrast with our research, it can be seen that the majority of the outcomes in the literature surpass, by a considerable margin, the interreader performance of our study (DSC of 0.798), underscoring the disparities between the studies. Besides that, our testing set comprises 190 patients from a distinct dataset where the training was performed, a substantially larger cohort than what is predominantly encountered in existing literature.

### 7.3 Final Outcomes

This chapter illustrates the methodology employed for segmenting EAT within the pericardium. The automated approach utilized the segmented pericardium generated by the  $2.5D + pp$  model. Subsequently, a comprehensive evaluation of the automated EAT segmentation's performance was conducted by comparing it against intrareader and interreader variability. This evaluation involved the utilization of both volume quantification and segmentation metrics.

In conclusion, upon assessing intra and interreader variability in 20 patients from the CHVNGE dataset, it was observed that the automated approach achieved satisfactory outcomes. However, it is not yet viable for use as an autonomous reader, as both quantification and segmentation metrics yielded results inferior to the interreader variability identified between Reader

1 and Reader 2. It is important to highlight that the performance of the automatic approach was evaluated across 190 patients. Besides that, it was evident that segmenting the pericardium poses significant challenges, as indicated by the considerable variability in both volume quantification and segmentation metrics, not only between the same reader but even more so between the two readers. After analysing some examples where the model yielded unsatisfactory results, it was concluded that the automatic method fails to segment the EAT in patients with significant anatomical variability. This variability is absent in the public datasets utilized for training. As a result, there is a need for additional data variability in training to enhance the results.

## Chapter 8

# Conclusion

This document introduces a new DL fully automated approach to segment the pericardium. Consequently, EAT is segmented using the same fat HU range as employed by specialists.

Throughout this work it was possible to see the significance of contextual information in the segmentation of certain structures which are difficult to identify, like the pericardium. This was confirmed and validated through both 2D and 3D approaches. In fact, the techniques employed to improve the visibility of the pericardium in each 2D slice were found to be ineffective. Applying a distinct HU range to enhance the pericardium results in the exclusion of crucial information concerning not only the pericardium itself but also neighboring organs, model's performance. Thus, utilizing a larger HU range encompassing more values and comprehensive information has demonstrated to be beneficial. Similar findings were confirmed when using 3D information. Indeed, providing U-Net with three consecutive axial slices was shown to help the model to better recognize the pericardium in 3D, avoiding the segmentation of other organs, particularly in the lower slices.

After evaluating intra- and interreader variability in a group of 20 patients from the CHVNGE dataset, it was noted that the automated method achieved acceptable results. Nevertheless, its feasibility as an independent reader is currently limited, as the quantification and segmentation metrics produced results that were worse than the interreader variability observed between the two experts. Moreover, it would be beneficial to access the intra and intervariability in all the 190 patients, and not only the 20 from CHVNGE. Comparing our results with existing literature is complex due to inaccessible private datasets with varying patient counts. Many studies also achieve better results by using the same datasets for training and testing. On the other hand our approach aims on utilizing exclusively public data, accessible to everyone, to explore the limits of achievable results using publicly available datasets. Therefore, the goal was to attain the most favorable outcomes in an unprecedented private dataset, which encompasses numerous calcifications and certain anatomical variations. It is for this reason that the results are satisfactory.

However, much of the unsatisfactory results generated by the model were in patients with

significant anatomical variability, a scenario that common data augmentation techniques cannot replicate. Therefore, the main limitation of this work lies in the absence of a large quantity of high-quality publicly available data with more anatomical variability to train the model. Data engineering posed a significant challenge within this thesis, evident through the procedure of image registration executed for the Cardiac Fat dataset. Moreover, numerous CT scans exhibited inadequate labeling, and certain information within the DICOM files displayed inconsistencies, requiring additional efforts to prepare all the data for training. Thus, it would be extremely beneficial for data providers to exercise greater caution regarding these issues and improve the data's quality, particularly when dealing with great volumes of data. Besides the lack of data, another limitation was the inability to train models that require high computational demands. Indeed, training a 3D U-Net could significantly enhance the model's access to the proven valuable contextual information. However, experimenting with a 2.5D architecture that incorporates more slices beyond the typical 3 axial slices, or providing the model with coronal, axial, and sagittal views can be a solution. There is also the possibility to explore new architectures.

While numerous studies indicate a potential connection between EAT and CAD, the measurement and utilization of EAT as a means of risk stratification from CT scans are not yet integrated into the standard practices of clinical settings, due mainly to the laborious task of manual segmentation of EAT. Hence, an automated method for precise EAT segmentation could improve reproducibility and encourage the emergence of additional studies concerning the relationship between EAT and CAD. In addition to the evident benefit of enabling early-stage CAD screening, EAT has also been suggested as a potential therapeutic target in certain studies, potentially avoiding the need for future invasive angiography. While the predominant and typical method in clinical settings revolves around quantifying fat and establishing a direct link to CAD, the assessment of segmentation still holds considerable importance. In fact, when a highly accurate EAT segmentation is achieved radiomics can be used to extract a large number of features which could be used to predict CAD and subsequently initiating preventative measures to prevent future cardiovascular events.

Overall, despite the outlined advantages and potential, there is still a significant gap to bridge in order to achieve these goals. Indeed, medical data are private and sensitive, subject to stringent privacy regulations and ethical considerations, limiting the access to extensive datasets and consequently the generalization of DL models. Furthermore, the computational requirements of deep learning models can limit their applicability in real-time clinical scenarios. Additionally, there exists a degree of skepticism among medical professionals concerning the workings of DL models and the process by which they yield results. Hence, an ongoing collaboration remains essential between technical expertise and medical knowledge for the construction and validation of models. Additionally, in recent years, there has been progress in the development of explainable AI techniques.



# References

- [1] AK Malakar, D. Choudhury, B. Halder, P. Paul, A. Uddin, and S. Chakraborty. "A review on coronary artery disease, its risk factors, and therapeutics". *Journal of Cellular Physiology*, 234(10):16812–16823, 2019.
- [2] S. Safiri et al. "Burden of ischemic heart disease and its attributable risk factors in 204 countries and territories, 1990–2019". *European Journal of Preventive Cardiology*, 29(2):420–431, 2022.
- [3] The Cardiology Advisor. "The Emerging Role of Cardiac CT and Calcium Scoring in the Diagnosis of CAD". <https://www.thecardiologyadvisor.com/home/decision-support-in-medicine/cardiology/the-emerging-role-of-cardiac-ct-and-calcium-scoring-in-the-diagnosis-of-cad/>. (accessed Nov. 15, 2022).
- [4] G. Iacobellis. "Epicardial adipose tissue in contemporary cardiology". *Nature Reviews Cardiology*, pages 1–14, 2022.
- [5] CX Wong, AN Ganesan, and JB Selvanayagam. "Epicardial fat and atrial fibrillation: current evidence, potential mechanisms, clinical implications, and future directions". *European heart journal*, 38(17):1294–1302, 2017.
- [6] C. VanPutte, J. Regan, and A. Russo. *Seeley's anatomy & physiology: Cardiovascular System: The Heart*, pages 665–709. McGraw-Hill, 2017.
- [7] B. Gaborit, C. Sengenès, P. Ancel, A. Jacquier, and A. Dutour-Meyer. "Role of epicardial adipose tissue in health and disease: a matter of fat?". *Comprehensive physiology*, 7(3):317–p, 2017.
- [8] The Society of Thoracic Surgeons. "The Patient Guide to Heart, Lung, and Esophageal Surgery: Coronary Artery Disease". <https://ctsurgerypatients.org/adult-heart-disease/coronary-artery-disease>. (accessed Nov. 12, 2022).
- [9] RA Burch, TA Siddiqui, LC Tou, KB Turner, and M Umair. "The Cost Effectiveness of Coronary CT Angiography and the Effective Utilization of CT-Fractional Flow Reserve in the Diagnosis of Coronary Artery Disease". *Journal of Cardiovascular Development and Disease*, 10(1):25, 2023.
- [10] Hospital da Luz. "Coronary angiography". <https://www.hospitaldaluz.pt/en/health-dictionary/coronary-angiography>. (accessed Nov. 12, 2022).
- [11] J. Playán-Escribano, P. Martínez-Losas, and MA Cobos-Gil. "Electrocardiographic changes after angioplasty of the left anterior descending coronary artery". *Circulation*, 139(12):1550–1553, 2019.

- [12] A. Chua, R. Blankstein, and B. Ko. "Coronary artery calcium in primary prevention". *Australian Journal for General Practitioners*, 49(8):464–469, Aug. 2020.
- [13] Johns Hopkins Medicine. "Coronary Computed Tomography Angiography (CCTA)". <https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/coronary-computed-tomography-angiography-ccta>. (accessed Nov. 12, 2022).
- [14] KM Abdelrahman et al. "Coronary Computed Tomography Angiography From Clinical Uses to Emerging Technologies: JACC State-of-the-Art Review". *Journal of the American College of Cardiology*, 76(10):1226–1243, 2020.
- [15] N. Ahmadi et al. "Mortality incidence of patients with non-obstructive coronary artery disease diagnosed by computed tomography angiography". *The American journal of cardiology*, 107(1):10–16, 2011.
- [16] Epicardial and pericardial fat analysis on ct images and artificial intelligence: A literature review. *Quantitative Imaging in Medicine and Surgery*, 12:2075–2089, Mar. 2022.
- [17] M. Goeller et al. "Epicardial adipose tissue density and volume are related to subclinical atherosclerosis, inflammation and major adverse cardiac events in asymptomatic subjects". *Journal of cardiovascular computed tomography*, 12(1):67–73, 2018.
- [18] E. Eisenberg et al. "Deep learning–based quantification of epicardial adipose tissue volume and attenuation predicts major adverse cardiovascular events in asymptomatic subjects". *Circulation: Cardiovascular Imaging*, 13(2):e009829, 2020.
- [19] G. Iacobellis, E. Lonn, A. Lamy, N. Singh, and AM Sharma. "Epicardial fat thickness and coronary artery disease correlate independently of obesity". *International journal of cardiology*, 146(3):452–454, 2011.
- [20] A. Yerramasu et al. "Increased volume of epicardial fat is an independent risk factor for accelerated progression of sub-clinical coronary atherosclerosis". *Atherosclerosis*, 220(1):223–230, 2012.
- [21] G. Iacobellis. "Epicardial fat: a new cardiovascular therapeutic target". *Current opinion in pharmacology*, 27:13–18, 2016.
- [22] L. Zhang, J. Sun, B. Jiang, L. Wang, Y. Zhang, and X. Xie. "Development of artificial intelligence in epicardial and pericoronary adipose tissue imaging: a systematic review". *European journal of hybrid imaging*, 5(1):1–13, 2021.
- [23] G. Iacobellis and HJ Willens. "Echocardiographic epicardial fat: a review of research and clinical applications". *Journal of the American Society of Echocardiography*, 22(12):1311–1319, 2009.
- [24] P. Daudé et al. "Deep-Learning Segmentation of Epicardial Adipose Tissue Using Four-Chamber Cardiac Magnetic Resonance Imaging". *Diagnostics*, 12(1):126, 2022.
- [25] KG Kahl et al. "Pericardial, intra-abdominal, and subcutaneous adipose tissue in patients with major depressive disorder". *Acta psychiatrica scandinavica*, 130(2):137–143, 2014.

- [26] O. Rodrigues, FFC Morais, NAOS Morais, LS Conci, LV Neto, and A Conci. "A novel approach for the automated segmentation and volume quantification of cardiac fats on computed tomography". *Computer Methods and Programs in Biomedicine*, 123:109–128, Jan. 2016.
- [27] M. Marwan and S. Achenbach. "Quantification of epicardial fat by computed tomography: why, when and how?". *Journal of cardiovascular computed tomography*, 7(1):3–10, 2013.
- [28] J. Qu et al. "Deep Learning-Based Approach for the Automatic Quantification of Epicardial Adipose Tissue from Non-Contrast CT". *Cognitive Computation*, 14(4):1392–1404, 2022.
- [29] M. Benčević, I. Galić, M. Habijan, and A. Pižurica. "Recent Progress in Epicardial and Pericardial Adipose Tissue Segmentation and Quantification Based on Deep Learning: A Systematic Review". *Applied Sciences*, 12(10):5217, 2022.
- [30] D. Dey et al. "Automated algorithm for atlas-based segmentation of the heart and pericardium from non-contrast CT". *Proc SPIE Int Soc Opt Eng.*, Mar. 2010, 7623,:762337, doi: 10.1117/12.844810.
- [31] X. Ding, D. Terzopoulos, M. Diaz-Zamudio, DS Berman, PJ Slomka, and D. Dey. "Automated pericardium delineation and epicardial fat volume quantification from noncontrast CT". *Medical Physics*, 42:5015–5026, Sep. 2015.
- [32] O. Rodrigues, LO Rodrigues, LSN Oliveira, A. Conci, and P. Liatsis. "Automated recognition of the pericardium contour on processed CT images using genetic algorithms". *Computers in Biology and Medicine*, 87:38–45, Aug. 2017.
- [33] F. Commandeur et al. "Deep Learning for Quantification of Epicardial and Thoracic Adipose Tissue from Non-Contrast CT". *IEEE Transactions on Medical Imaging*, 37:1835–1846, Aug. 2018.
- [34] F. Commandeur et al. "Fully automated CT quantification of Epicardial adipose tissue by deep learning: A multicenter study". *Radiology: Artificial Intelligence*, 1, Nov. 2019.
- [35] C. Priya and S. Sudha. "Adaptive Fruitfly Based Modified Region Growing Algorithm for Cardiac Fat Segmentation Using Optimal Neural Network". *Journal of Medical Systems*, 43, May 2019.
- [36] Z. Li, L. Zou, and R. Yang. "A neural network-based method for automatic pericardium segmentation". *Proceedings of the 2nd International Conference on Computer Science and Software Engineering*, pages 45–49, May 2019.
- [37] A. Kazemi et al. "Automated Segmentation of Cardiac Fats Based on Extraction of Textural Features from Non-Contrast CT Images". In *2020 25th International Computer Conference, Computer Society of Iran (CSICC)*, Jan. 2020, pp. 1-7.
- [38] Q. Zhang, J. Zhou, B. Zhang, W. Jia, and E. Wu. "Automatic Epicardial Fat Segmentation and Quantification of CT Scans Using Dual U-Nets with a Morphological Processing Layer". *IEEE Access*, 8:128032–128041, 2020.
- [39] A. Hoori, T. Hu, J. Lee, S. Al-Kindi, and DL Wilson. "Deep Learning Segmentation and Quantification Method for Assessing Epicardial Adipose Tissue in CT Calcium Score Scans". *Sci Rep*, 12, 2276, Feb. 2022, doi: 10.1038/s41598-022-06351-z.

- [40] D. Molnar et al. "Artificial intelligence based automatic quantification of epicardial adipose tissue suitable for large scale population studies". *Scientific Reports*, 11, Dec. 2021.
- [41] M. Bencevic, M. Habijan, and I. Galic. "Epicardial Adipose Tissue Segmentation from CT Images with A Semi-3D Neural Network". volume 2021-September, pages 87–90. Croatian Society Electronics in Marine - ELMAR, Sep. 2021.
- [42] T. Siriapisith, W. Kusakunniran, and P. Haddawy. "A 3D deep learning approach to epicardial fat segmentation in non-contrast and post-contrast cardiac CT images". *PeerJ Computer Science*, 7, 2021.
- [43] CAL Neto and HFP Oliveira. "Automatic Detection of Structures in CT Images to Improve Radiotherapy Treatment". M.S. thesis, FEUP, UP, Porto, Portugal, 2019. [Online]. Available: <https://hdl.handle.net/10216/123032>.
- [44] Q. Li, L. Zhang, J. You, D. Zhang, and P. Bhattacharya. "Dark line detection with line width extraction". In *2008 15th IEEE International Conference on Image Processing*, pages 621–624, 2008.
- [45] F. Greco, R. Salgado, WV Hecke, RD Buono, PM. Parizel, and CA Mallio. Epicardial and pericardial fat analysis on ct images and artificial intelligence: a literature review. *Quantitative Imaging in Medicine and Surgery*, 12(3), 2021.
- [46] HG Santos, LS Ochi, EH Marinho, and LMA Drummond. "Combining an evolutionary algorithm with data mining to solve a single-vehicle routing problem". *Neurocomputing*, 70(1-3):70–77, 2006.
- [47] AM Bucher et al. "Influence of technical parameters on epicardial fat volume quantification at cardiac CT". *European Journal of Radiology*, 84(6):1062–1067, 2015.
- [48] S. Mirjalili, SM Mirjalili, and A. Lewis. "Grey wolf optimizer". *Advances in engineering software*, 69:46–61, 2014.
- [49] W Pan. "A new fruit fly optimization algorithm: taking the financial distress model as an example". *Knowledge-Based Systems*, 26:69–74, 2012.
- [50] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [51] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation". volume 9351, pages 234–241. Springer Verlag, 2015.
- [52] W. Sun, Z. Gao, J. Cui, B. Ramesh, B. Zhang, and Z. Li. "Semantic Segmentation Leveraging Simultaneous Depth Estimation". *Sensors*, 21(3):690, 2021.
- [53] O. Oktay et al. "Attention U-Net: Learning Where to Look for the Pancreas". *ArXiv: 1804.03999*, Apr. 2018.
- [54] Z. Fang, Y. Chen, D. Nie, W. Lin, and D. Shen. "Rca-u-net: Residual channel attention u-net for fast tissue quantification in magnetic resonance fingerprinting". In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 101–109. Springer, 2019.

- [55] LC Chen, G. Papandreou, F. Schroff, and H. Adam. "Rethinking atrous convolution for semantic image segmentation". *arXiv preprint arXiv:1706.05587*, 2017.
- [56] LR Dice. "Measures of the amount of ecologic association between species". *Ecology*, 26(3):297–302, 1945.
- [57] G. Rote. "Computing the minimum Hausdorff distance between two point sets on a line under translation". *Information Processing Letters*, 38(3):123–127, 1991.
- [58] D Nettleton. "Selection of Variables and Factor Derivation.". In *Commercial Data Mining*, pages 79–104, Jan. 2014.
- [59] R. Baeza et al. "A Generalization Study of Automatic Pericardial Segmentation in Computed Tomography Images". In *International Conference on Wireless Mobile Communication and Healthcare*, pages 157–167. Springer, 2022.
- [60] Visual Computing Lab - UFF. (2015). "Cardiac Fat Database - Computed Tomography". <http://visual.ic.uff.br/en/cardio/ctfat/index.php>. (accessed Feb. 20, 2022).
- [61] Kónya et al. (2020). *Lung segmentation dataset*. <https://www.kaggle.com/sandorkonya/ct-lung-heart-trachea-segmentation>. (accessed Feb. 20, 2022).
- [62] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. "ORB: An efficient alternative to SIFT or SURF". In *2011 International conference on computer vision*, pages 2564–2571. IEEE, 2011.
- [63] X. Li, Y. Liu, Y. Wang, and D. Yan. "Computing homography with RANSAC algorithm: a novel method of registration". In *Electronic Imaging and Multimedia Technology IV*, volume 5637, pages 109–112. SPIE, 2005.
- [64] CB. Barber, DP. Dobkin, and H. Huhdanpaa. "The quickhull algorithm for convex hulls". *ACM Transactions on Mathematical Software (TOMS)*, 22(4):469–483, 1996.