Érick Oliveira Rodrigues

AUTOMATED SEGMENTATION OF EPICARDIAL AND MEDIASTINAL FATS USING INTERSUBJECT REGISTRATION AND CLASSIFICATION ALGORITHMS

Thesis presented to the Computing Graduate program of the Universidade Federal Fluminense in partial fulfillment of the requirements for the Master of Science degree on the field of Visual Computing.

Advisor: Prof. Dr. Aura Conci

Niterói

2015

Érick Oliveira Rodrigues

AUTOMATED SEGMENTATION OF EPICARDIAL AND MEDIASTINAL FATS USING INTERSUBJECT REGISTRATION AND CLASSIFICATION ALGORITHMS

Thesis presented to the Computing Graduate program of the Universidade Federal Fluminense in partial fulfillment of the requirements for the Master of Science degree on the field of Visual Computing.

Approved on February 2015 by

Prof. Dr. Aura Conci – Advisor

UFF

Prof. Dr. Alexandre Plastino de Carvalho

UFF

Prof. Dr. Leandro Augusto Frata Fernandes

UFF

Prof. Dr. Leonardo Kayat Bittencourt

UFF

Prof. Dr. Marcelo de Andrade Dreux

PUC-RIO

Prof. Dr. Leonardo Vieira Neto

UFRJ

Niterói

2015

ACKNOWLEDGEMENTS

I would like to thank my parents Gilson Mamede and Carla Valéria for the long-term support on my education. Even when virtually impossible they have surpassed the difficulties to grant their children an opportunity to study, one that both did not have.

I am grateful to my advisor Aura who have patiently demonstrated me a view of science that I did not know existed. Also, on not hesitating when it was necessary to highlight my mistakes.

I would like to thank every person directly or indirectly related to my university that cordially fulfilled their roles, a special thanks to Carlos Eduardo Cabral da Cunha.

My last acknowledgement goes to the financial support of the Universidade Federal Fluminense and CAPES.

"Take the risk of thinking for yourself, much more happiness, truth, beauty and wisdom will come to you that way."

Christopher Hitchens

ABSTRACT

The amount of fat on the surroundings of the heart is correlated to health risk factors such as carotid stiffness, coronary artery calcification, atrial fibrillation, atherosclerosis and other health conditions. Furthermore, cardiac fat deposits vary unrelated to the overall fat of the patient which, therefore, reinforces the idea of a direct quantitative analysis as being essential. However, manual quantification has not been widely employed in clinical practice due to the required human workload.

Clinical decision support systems are computer programs capable of evaluating data and providing a corresponding diagnosis or information to complement the physicians' analyses. The objective of this work is to propose a method capable of fully automatically segmenting two types of cardiac adipose tissues that stand apart from each other by the pericardium. The source for segmentation are CT images which, in turn, were obtained by the standard acquisition protocol used for coronary calcium scoring. Much effort was devoted to promote minimal user intervention and easiness of reproducibility.

The proposed segmentation methodology consists of an intersubject registration that roughly standardize images of distinct patients, an extraction of features of the registered images and, finally, an appliance of classification algorithms. The classification algorithm predicts if an incoming pixel corresponds to a certain type of cardiac fat based on the extracted features. Furthermore, we extensively evaluate the performance of several algorithms on this task and discuss which ones provided better predictive models. Experimental results regarding both epicardial and mediastinal fats have shown that the mean accuracy for the proposed method is 98.4% with a mean true positive rate of 96.2%. In average, the Dice similarity index has been equal to 96.8%.

Keywords: epicardial, mediastinal, segmentation, automatic, classification, Random Forest, cardiac, fat, adipose tissue, registration, intersubject

LIST OF FIGURES

Figure 1: Outermost layers of the heart [13]18
Figure 2: Mediastinal space [13]19
Figure 3: A registration process23
Figure 4: The basic components of the ITK registration framework23
Figure 5: Categories of transformations usually used in the medical registration26
Figure 6: A classical acquisition with an electrocardiogram-gated sequence [56]29
Figure 7: A features vector based on a pixel and its related information
Figure 8: An image within the (-200,-30) HU range37
Figure 9: Variation between four distinct patients on the same HU range
Figure 10: Four intersubjectly registered patients
Figure 11: Heart exceeding the boundaries of the image41
Figure 12: Highlighted retrosternal area of four distinct patients
Figure 13: Overall steps of the proposed registration43
Figure 14: A thresholding process44
Figure 15: Ten randomly chosen binary instances of the retrosternal area45
Figure 16: The atlas applied on the registration45
Figure 17: Evidence for occasional darker pixels preference (high fat patients)47
Figure 18: Successful recognitions (low fat patients)48
Figure 19: The hybrid mean difference similarity measure53
Figure 20: Successful recognitions on varied scales56
Figure 21: Binding of the two points pl and pr
Figure 22: The rectangular area and the central point of the atlas
Figure 23: Absence of continuous fat deposits in the retrosternal area59

Figure 24: Small gaps on the image provoked by rescaling	.63
Figure 25: An occasion where the point pl was displaced much more than pr	.63
Figure 26: Wrong recognition due to the absence of the isLongEnough variable	.64
Figure 27: The 4 instances, out of 52, where the recognition method failed	.67
Figure 28: The same 4 instances after the confirmation method	.68
Figure 29: An overview of the whole registration procedure	.70
Figure 30: Patient demographics	.71
Figure 31: A slice of one patient that compose the ground truth	.71
Figure 32: Overall steps required for the classified segmentation.	.72
Figure 33: Segmentation of an arbitrary DICOM file	.73
Figure 34: Neighborhood around the pixel <i>Py</i> , <i>x</i>	.74
Figure 35: Euclidian distance Kernel in a 7x7 window.	.77
Figure 36: Sup metric	.77
Figure 37: $d \infty$ distance in a 7x7 Kernel	.77
Figure 38: A Gaussian Kernel.	.78
Figure 39: Smooth variation Kernel in a 7x7 window	.79
Figure 40: CSV convolution	.79
Figure 41: Construction of a co-occurrence matrix	.81
Figure 42: Construction of a run length matrix.	.83
Figure 43: Accuracies (y-axis) of classifiers versus neighborhood sizes (x-axis)	.88
Figure 44: One instance segmented by algorithms trained on another patient	.91
Figure 45: Ground truth	.91
Figure 46: Manually (left) versus an automatically segmented slice (right)	.99
Figure 47: 3D fat model of an automatically segmented patient.	.99

LIST OF TABLES

Table 1: Values of Hounsfield units (HU) for (fan beam) medical CT scans	.36
Table 2: A comparison of the similarity measures using 52 images	.55
Table 3: Extracted features and used parameters.	.75
Table 4: Accuracies and convergence on a large dataset	.89
Table 5: Features ranked differently for each class	.94
Table 6: Random Forest algorithm on 66% split validation	.96
Table 7: Random Forest algorithm on 10-fold cross validation.	.96
Table 8: Comparison of the epicardial fat segmentation.	.97
Table 9: A comparison of the Dice similarity variation	.98

LIST OF ALGORITHMS

Algorithm 1: Hybrid mean similarity function.	.54
Algorithm 2: Confirmation method for the retrosternal area recognition.	.61
Algorithm 3: Steps to find the retrosternal area.	.65
Algorithm 4: Formally applying the confirmationMethod function	.65

LIST OF ABBREVIATIONS AND ACRONYMS

- AM Arithmetic Mean
- CDSS Clinical Decision Support Systems
- CPU Central Processing Unit
- CSV Coefficient of Smooth Variation
- CT Computed Tomography
- DICOM Digital Imaging and Communication in Medicine
- ECG Electrocardiography
- ED End-diastolic
- ES End-systolic
- FN False Negative
- FP False Positive
- H Entropy
- ITK Insight Segmentation and Registration Toolkit
- KNN k-Nearest Neighbor
- LV Left ventricle
- MD Mean Difference
- MI Mutual Information
- MLP Multi Layer Perceptron
- MR Magnetic resonance
- MRI Magnetic resonance imaging
- PET Positron emission tomography
- RAM Random Access Memory
- RLE Run Length Encoding

- RLM Run Length Matrix
- SPECT Single-proton emission computed tomography
- SURF Speed up robust features
- SMO Sequential Minimal Optimization
- SVM Support Vector Machine
- TN True Negative
- TP True Positive
- RBF Radial Basis Function
- RF Random Forest
- RVM Relevance Vector Machine
- WMI Weighted Mutual Information

SUMMARY

1	INT	ROD	DUCTION	15
	1.1	The	eme Definition	. 15
	1.2	Mc	tivation, Objectives and Contribution	15
	1.3	Тор	pics Arrangement	16
	1.4	Eth	ics Committee	17
2	LITE	LITERATURE REVIEW		
	2.1	Car	diac Anatomy	18
	2.2	Hea	alth Risks Associated to the Amount of Cardiac Fat	21
	2.3	Dat	ta Access	21
	2.4	Ima	age Registration	22
	2.4	.1	Types of Registration	24
	2.5	Seg	gmentation	29
	2.5	5.1	Classification Algorithms	30
	2.5	5.2	Applicability of the Classified Segmentation	32
	2.5	5.3	Segmentation of the Cardiac Adipose Tissue	33
3	PRC	OPO9	SED APPROACH	37
	3.1	Pro	pposed Registration	40
	3.1	1	Retrosternal Area Recognition	44
	3.1	2	Scale Invariance	55
	3.1	3	Confirmation Method	57
	3.1	.4	Applying the Transformation	68
	3.2	Pro	posed Segmentation	70
	3.2	2.1	Used Features	73
	3.2	2.2	Evaluation of the Convergence Time	84

	3.2	2.3 Evaluation of Neighbordhood Sizes	86
	3.2	2.4 Overfitting Analysis	90
	3.2	2.5 Feature Analysis	93
4	4 RESULTS		
	4.1	Comparing Results	97
	4.2	Visualization	98
5	COI	NCLUSION	100
	5.1	Further Improvements	101
	5.2	Final Remarks	102
R	EFERE	NCES	103
A	ppend	Jix A – Accessing CT Data from DICOM Files	116
	CT Ac	cquisition	117
	Libra	ıry	118
	Converting Raw Data to Housfield Units119		
	Displaying Images from Hounsfield Units121		
Distinct Thresholds of Hounsfield Ranges1			122
A	Appendix B – Rescaling Images According to DICOM125		
A	ppend	Jix C – Automatically Segmented Slices	127
	Segmented Patient		

1 INTRODUCTION

1.1 THEME DEFINITION

An increasing demand for medical diagnosis support systems has been observed jointly to the computational evolution on the last years. Systems of this kind speed up the tedious and meticulous analysis done by physicians or technicians on patients' medical data, where, in many cases, a huge amount of data have to be analyzed and, therefore, the diagnosis may lack precision and suffer noticeable inter and intra-observer variation [1].

The cardiac epicardial and mediastinal (also termed pericardial) fats are correlated to several cardiovascular risk factors [2]. At the present, three techniques (i.e., modalities) appear suitable for the quantification of these adipose tissues, namely Magnetic Resonance Imaging (MRI), Echocardiography and Computed Tomography (CT). Each one of these modalities have been used in several medical studies in the literature [3,4,5]. However, computed tomography provides a more accurate evaluation of fat tissues due to its higher spatial resolution if compared to ultrasound and MRI [6]. In addition, CT is also widely used for evaluating coronary calcium score [5].

This work is entirely based on the processing and evaluation of images that belong to the axial plane and compose the standard non-contrasted CT acquisition protocol. More information about CT and the axial plane can be found in Appendix A and B.

1.2 MOTIVATION, OBJECTIVES AND CONTRIBUTION

The automated quantitative analysis of the epicardial and mediastinal fats may add a prognostic value to cardiac CT trials, ensuring an improvement on its cost-effectiveness. Besides, that automation diminishes the variability introduced by different observers. In fact, as previously mentioned, quantifying these data by direct user interaction is highly prone to inter and intra-observer variability. Thus, evaluated samples may not be associated to a unified sense of segmentation. Iacobellis et al. [7] have shown that the epicardial fat thickness and coronary artery disease, for instance, correlate independently of obesity. This evidence

supports the individual segmentation and further quantification of the adipose tissues rather than merely and simply estimating that volume based on the overall fat of the patient.

The objective of this work is to develop a unified method capable of fully automatic segmenting both epicardial and mediastinal fats on CT images. To the extent of our knowledge, there is currently no unified method in the literature capable of both types of automatic segmentations. On the entire extent of the proposed methods for cardiac fat segmentation [6,8,9,10,11], the addressed issue is specifically the segmentation of the epicardial fat, whereas no work attempted to segment the mediastinal fat. Besides, we resolve this issue diverging greatly with respect to the basis of the approaches proposed by the authors of the current related works. This divergence is given by employing machine learning classification algorithms to produce the segmentation.

This work contributes mainly to the field of visual computing but also to the medical and machine learning field by, namely: (1) proposing an accurate intersubject registration for cardiac CT images, (2) proposing and analyzing a hybrid similarity measure that was applied within the registration procedure, (3) proposing a new feature based on the Gaussian Kernel, (4) corroborating on the appliance of classification algorithms for image segmentation, (5) analyzing the performance and accuracy of various classifiers for the problem, (6) creating a ground truth for cardiac fats available online and, finally, by (7) proposing a unified and fully automatic segmentation method for both epicardial and mediastinal fats on cardiac CT images.

1.3 TOPICS ARRANGEMENT

The topics of this work are distributed on the following manner. On the next Chapter we present an overview of the literature on the fields of cardiac anatomy, health related risks associated to the cardiac adipose tissue, cardiac fat segmentation, CT related information such as the adipose tissue range in Hounsfield Units (HU) and information on the DICOM standard. Furthermore, we address a brief review of image registration and classification algorithms, which are the two main steps employed in the proposed method.

The Chapter 3 consists of our entire methodology. At first, we address the proposed registration. Later on, the feature extraction as well as the appliance of the classification

algorithms to the problem are addressed on the second main section of the same chapter. The registration itself can be seen as a preprocessing step and the remaining as the actual segmentator. In our work, features can be viewed as characteristics of a certain pixel or surrounding area that help to define its class. Moreover, classification algorithms can be roughly viewed as learners that assemble predictive models trained on specific data, as to this work, on the extracted features. The demographics of the patients regarded in this work is also presented in the Chapter 3.

The Chapter 4 presents the achieved results with regard to the registration, features extraction and classification algorithms. Furthermore, on the same chapter, the achieved results are compared to the three mainly related works that have already attempted to semi or fully automatic segment the epicardial fat. In addition, the downsides of the current comparisons are discussed.

The final chapter addresses the conclusions, further improvements that can be applied and some technical details of the computers, libraries and programing languages regarded in this work.

1.4 ETHICS COMMITTEE

This work was approved by the ethics and research committee of HUCFF/UFRJ (protocol number 069/10) and of the Brazilian National Institute of Cardiology (protocol number 0324/04-04-2001). All patients were enlightened as to the objectives of this work and were required to sign a consent form.

2 LITERATURE REVIEW

In this chapter, we provide some literature-related information along with a more extensive introduction to our work. At first, we discuss about cardiac anatomy and highlight the structures that are significant for us. Later on, we discuss about health risks and diseases associated to the cardiac adipose tissue. Furthermore, we address other works in the literature that have already proposed segmentation approaches for the cardiac fat, comparing and discussing their methodology. Finally, a brief discussion about how CT data can be accessed and handled as well as a review on image registration and classification algorithms are addressed.

2.1 CARDIAC ANATOMY

The human heart is enclosed in the pericardium, a fibroserous sac comprising three concentric layers as shown in Figure 1. The outermost layer is a densely fibrous, tough and inelastic structure (fibrous pericardium). Inside the fibrous pericardium is the serous pericardium, which consists of two layers; the outer of these (which is firmly applied to the inner surface of the fibrous pericardium) is termed the parietal layer. This layer is reflected around the roots of the great vessels to become continuous with the visceral layer (epicardium), which covers the internal surface of the heart and is firmly applied to it [12].



Figure 1: Outermost layers of the heart [13].

Sacks et al. [14] define the epicardium or visceral layer of the pericardium as a population of mesothelial cells that migrate onto the surface of the heart from the area of the septum transversum (the embryological source of the diaphragm). Furthermore, they define that, in the normal adult, epicardial fat is concentrated in the atrioventricular and interventricular grooves and along the major branches of the coronary arteries and, to a lesser extent, around the atria, over the free wall of the right ventricle and over the apex of the left ventricle. In addition, the authors define pericardial fat as all the epicardial and the paracardial fat and, consequently, define that paracardial is the fat located on the external surface of the parietal pericardium (also within the mediastinum). They also highlight that the paracardial fat has been alternatively termed mediastinal fat in the literature. The mediastinal area is shown in Figure 2.



Figure 2: Mediastinal space [13].

Sicari et al. [3] define that the cardiac fat can be distinguished in two deposits: (1) the epicardial adipose tissue, which they describe exactly with the same words as the definition of Sacks et al. [14] and, (2) the pericardial adipose tissue, which they define as being the fat situated on the external surface of the parietal pericardium within the mediastinum (alternatively termed mediastinal or intrathoracic fat). Nevertheless, Rosito et al. [2], Dey et al. [11] and Nichols et al. [15], for instance, describe the pericardial fat as being any adipose tissue within the pericardial sac.

Marwam et al. [16], however, addressed the fact that the terminology used to define fat deposits surrounding the heart in the current literature is diverse and, to some extent, confusing. The authors state that the term pericardial fat is frequent in most of the published literature as referring to the adipose tissue enclosed within the pericardial sac. Despite of being a widely used definition, they state that the more accurate term would be epicardial fat, given its location on the internal epicardial surface of the heart. Shahzad et al. [8] also mentioned this confusion with relation to the terminology within the literature.

In summary, the majority of the published works [3,6,8,9,10,14,16,17,18,19] [20,21,22,23,24] agree on the epicardial fat terminology as being correct for the fat contained within the epicardium and, therefore, also within the pericardium. On the other hand, to the extent of our knowledge, the only work that has been out of that agreement on the epicardial fat definition is the one from Mahabadi's et al. [25]. Nevertheless, there is an accentuated disagreement on the pericardial fat terminology amongst various works. In fact, some works [2,8,11,15,26,27] support the idea that the pericardial fat is merely the fat that is enclosed by the pericardium sac, which is analogous to the epicardial fat definition. Others [3,18,21,28,29,30] support that the pericardial fat terminology defines the adipose tissue located on the external surface of the parietal pericardium, within the mediastinum. Moreover, some works [14,22,31,32,33] even define the pericardial fat as being equivalent to all the adipose tissues within the mediastinum, including the pericardial or epicardial fats.

In conclusion, due to the fact that the terminology for the cardiac fats is currently not properly established in the literature, we will define the fat located within the epicardium as epicardial, corroborating with the majority of published works, and by following the same "first outer anatomical container" logic, we conclude that mediastinal fat is the best definition for the fat located on the external surface of the heart or of the fibrous pericardium. In other words, by taking Figure 2 as a reference, the mediastinal fat is located within the pink, blue and green colored areas, and is mediastinal as long as it is not epicardial (i.e., as long as it is not located within the epicardium).

2.2 HEALTH RISKS ASSOCIATED TO THE AMOUNT OF CARDIAC FAT

Some studies [24,33] associate the amount of epicardial adipose tissue to the progression of coronary artery calcification. Schlett et al. [26], for instance, found that the epicardial fat volume is nearly twice as high in patients with high-risk coronary lesions as compared to those without coronary artery calcification. Several studies also correlate other cardiovascular risk factors and outcomes to the epicardial adipose tissue volume, such as diastolic filling [18], myocardial infarction [25], atrial fibrillation and ablation outcome [26], carotid stiffness [32], atherosclerosis [19,20,21], and many others [2,25,30,31,34]. Furthermore, Wei-Ta et al. [35] have also shown that high coronary artery calcium score is associated to a higher general cancer incidence.

In addition, some studies address the importance of the mediastinal fat (due to the previously discussed inconsistency some authors call it pericardial fat) and its correlation with pathogenic profiles, health risk factors and diseases [29,36,37]. Some [21,32] associate the mediastinal fat, along with the epicardial fat, to carotid stiffness. Others [21,33] associate them to atherosclerosis and coronary artery calcification. Sicari et al. [3] have also shown how mediastinal fat rather than epicardial fat is a cardiometabolic risk marker.

Moreover, a 16-year study [38] that assessed a total of 384 597 patients associated a rate of approximately 38.4% of death in the subsequent 28 days of individuals that have had their first major coronary event. The same study also concludes that fatal cases is slightly less associated to female individuals. Another study ranks cardiovascular accidents as the most common cause of sudden natural death [39]. Therefore, the practice of automatically evaluating the amount of fat related to the heart may contribute greatly to avoid such outcomes.

2.3 DATA ACCESS

The Digital Imaging and Communications in Medicine (DICOM) standard, originally published as the American College of Radiology – National Electrical Manufacturers Association Standard for Digital Imaging and Communications in Medicine, now maintained by the multi-specialty DICOM Standards Committee, specifies a nonproprietary data interchange protocol, digital image format, file structure for biomedical images and imagerelated information [40].

DICOM provides detailed engineering information that can be used to enable network connectivity among a variety of manufacturers' products. This standard also describes how to format and exchange medical images and associated information, both within and outside the hospital (e.g., tele-radiology, telemedicine). DICOM interfaces and protocols are available for sharing and storing any combination of the following categories of digital imaging devices: (a) image acquisition equipment (e.g., computed tomography, ultrasonography, and nuclear medicine scanners); (b) image archives; (c) image processing devices and image display workstations; (d) hard-copy output devices (e.g., photographic transparency film and paper printers) [40].

The DICOM standard offers a comprehensive specification of information content, structure and encoding for electronic interchange of diagnostic and therapeutic images. In other words, DICOM is a complete specification "from top to bottom" of the elements required to achieve a practical level of automatic interoperation [40]. All the CT data accessed and processed in this work were gathered from DICOM files as 512x512 pixels-wide images. The acquisition procedure, the way that CT data is stored and accessed in a DICOM file are described within the Appendix A – Accessing CT Data from DICOM Files.

2.4 IMAGE REGISTRATION

Image registration can be defined as the process of matching characteristics from images in order to minimize the variation between overlapping pixels or areas of pixels [41]. Such processes are included in panoramas assemblages, medical images such as shown in Figure 3, time series [42,43] and many others processes. Registration is also alternatively seen as an optimization problem with the goal of finding the spatial mapping that brings images, parts of them, or even a combination of these parts into minimal variation. The image stitching terminology is also used in the literature [44,45] and consists of the same fundamentals from the registration yet usually associated to panoramas generation, which, in turn, usually involves more than just two images as input and some additional considerations [45].



Figure 3: A registration process.

The ITK framework [46] is a well-known and robust framework in the area of visual computing. It establishes its registration method based on the diagram depicted in Figure 4. The registration procedure is composed of a Fixed Image F(P) and a Moving Image M(P) as input data, where P represents a position in N-dimensional space. The Transform component T represents the spatial mapping of values from F(P) to M(P) [47].



Figure 4: The basic components of the ITK registration framework.

By a simple analogy we can suppose that the perfect registration or transformation T is achieved when the statement T(P) = F(P) becomes true, assuming that T(P) is the

transformed image. When comparing F(P) to T(P), the image T(P) should have been already interpolated. That is, the actual process of drawing pixels from the transformed image function T(P) to an image file or to the screen (i.e., to a raster) is often called rasterization and, for doing so, interpolations are required.

The interpolation can be done in two general ways. On the (1) forward method, each pixel from the transformed image function T(P) is directly mapped to a discrete pixel of the raster's grid. Therefore, the forward method can produce holes or overlaps on the output image due to rounding and ordering of the draws. On the (2) backward method, a single position of the grid is mapped to a set of points of T(P). In this case, neither holes nor overlaps occur on the outputted image. Therefore, as being the most concise, the backward method is the most commonly used type of interpolation [48].

In addition, the Measure component $S(F, M \circ T)$ provides a measure of how well the images F(P) and M(P) are matching. This measurement forms the quantitative criterion to be maximized by the Optimizer component. Thus, at each iteration of the loop, the Optimizer evaluates the obtained measurement of similarity and changes the parameters of the transformation in order to raise it, also deciding when and whether to stop [46].

2.4.1 TYPES OF REGISTRATION

The terms registration, as well as fusion, matching, integration, correlation, and others, appear polysemously in the literature, either referring to a single step or to the whole of the modality integration process [42]. Patients may undergo various MR, CT, SPECT and other studies for anatomical or general reference of a single organ, for instance. Thus, registration of images from practically any combination of modalities will benefit the physician's analysis. A second example concerns radiotherapy treatment, where both CT and MR can be employed. The former is needed to compute the radiation dose accurately, while the latter is usually better suited for delineation of tumor tissues. This type of registration is usually termed multimodality [42].

Besides the multimodality registrations, important application areas exist in monomodality registration (i.e., when only one modality or type of acquisition is considered). The multimodality refers to cases when a patient is scanned in successive intervals (time

series) or when an organ of several patients are scanned using a single modality, for instance. Other examples include treatment verification by comparison of ictal and inter-ictal SPECT images and growth monitoring (e.g., using time series of MR scans or termographic images on tumors, X-ray time series on specific bones and others) [43]. Due to the high degree of similarity between images of the same modality, proposing a registration method of this type is usually simpler than in multimodality applications [42].

According to Maintz et al. [42], the nature of registrations is divided in three main categories: (1) the extrinsic, (2) the intrinsic and (3) non-image based. On the extrinsic registrations, artificial objects (e.g., markers) are attached to the patient, which are designed to be well visible and accurately detectable in all of the pertinent modalities. As such, the registration of the acquired images is comparatively fast, easy, can virtually always be automated and, since the parameters of the registration can often be computed explicitly, registrations of this kind do not require complex optimization algorithms. On the (2) intrinsic registration, the corresponding methods rely only on the image-related content of the patient. The intrinsic registration can be based on a limited set of identified salient points (landmarks), on the alignment of segmented binary structures (segmentation based) yet it is most commonly based on object's surfaces or directly onto measures computed from the image grey values. Alternatively, it can also be (3) non-image based. Although it seems paradoxical for a multimodality registration to be non-image based it is actually possible if the imaging coordinate systems of the two scanners involved are somehow calibrated to each other. This type of registration usually requires the scanners to be brought into the same physical location, and to assume that the patient remains motionless between acquisitions.

2.4.1.1 SUBJECT TYPE

Matinz et al. [42] also explains that when all the involved images are acquired from a single patient this is referred to as intrasubject registration. If the registration is accomplished using two different patients (or a patient and a model), this is referred to as intersubject registration. If one image is acquired from a single patient and the other image is somehow constructed from a dataset obtained from imaging of many subjects, then it is called atlas registration. They also state that many registrations of a patient image to an image of a

'normal' subject are termed atlas registration. However, the authors prefer to define this type of registration as intersubject instead.

2.4.1.2 NATURE OF THE TRANSFORMATION

Rigid Transformation is a type of transformation where the distance between every pair of points is preserved [49,50]. Thus, the Rigid Transformation comprises operations like rotations, translations and reflections or a combination of these operations [42]. Affine Transformations are functions between affine spaces that preserve points, straight lines and planes [51,52]. Affine Transformations comprise operations such as translations, scaling, homogeneous and inhomogeneous dilations, similarity transformations, reflections, rotations, shear mapping and compositions of these operations [53]. The transformations usually employed within the medical registration field are divided in the categories depicted on the following Figure 5 [42]. The overall nature of the transformations can be either (1) global, i.e., applied to the entire image, or (2) local, i.e., applied to a certain area and further fused with other transformations or parts of the input image.



Figure 5: Categories of transformations usually used in the medical registration.

The definition of each type of transformations shown in Figure 5 can be simplified as (1) rigid, if only translations and rotations are allowed; (2) affine, if the transformation maps any parallel lines of the image into parallel lines; (3) projective, if it maps lines into lines or (4) curved or elastic, if it maps lines into curves [42,52].

In this work, we will just use affine transformations. Affine transformations are expressed in the form of a linear matrix multiplication. The Equation (1) demonstrates a transformation T where A is the coefficient matrix, B is the displacement matrix and $\begin{bmatrix} x \\ y \end{bmatrix}$ is the vector (or coordinates of the pixel) we want to transform [54,55,52].

$$T = \begin{bmatrix} x' \\ y' \end{bmatrix} = A \begin{bmatrix} x \\ y \end{bmatrix} + B \tag{1}$$

In \mathbb{R}^2 , the usual way to represent an affine transformation T is by using a 2x3 matrix with homogeneous coordinates as shown in Equation (2).

$$A = \begin{bmatrix} a_{0,0} & a_{0,1} \\ a_{1,0} & a_{1,1} \end{bmatrix}, B = \begin{bmatrix} b_{0,0} \\ b_{0,1} \end{bmatrix}.$$

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{0,0} & a_{0,1} & b_{0,0} \\ a_{1,0} & a_{1,1} & b_{0,1} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$
(2)

2.4.1.3 CARDIAC REGISTRATION

The work of Mäkelä et al. [56], for instance, presents a review of various image registration methods applied to cardiac imaging regarding different modalities (including CT). Every method assessed in that work is an intrasubject registration that, by definition, considers different slices of the same patient. Intrasubject registrations for several types of imaging modalities are already well developed with relation to the medical field, as the survey of Pluim et al. [57] indicates. Moreover, CT images have been registered to several other image modalities such as 2D video images [58,59], 2D fluoroscopy images [60,61,62] and portal

images [63,64,65]. Meyer et al. [66] and Koral et al. [67] registered CT and SPECT images focusing on the abdomen while Kagadis et al. [68] compared a surface-based and a mutual information based registration routine for the same matter [42,69,70].

Mäkelä et al. [56] separate the cardiac image registration in two main categories: (1) those based on Geometric Image Features and (2) those based on Voxel Similarity Measures. The methods based on the Geometric Features are divided in registration of a set of points and edges or registration of surfaces. Registration methods based on Voxel (or pixel) Similarity include moments and principal-axes methods, intensity difference, correlation and methods based on mutual information [56].

One of the branches of the Geometric Image Features type of registration is the Point-Based Registration. The Point-Based Registration mainly relies on matching points among the images. These points are often anatomical landmarks or external markers. On the external markers case, they are set before the data acquisition and are captured along with the main data during the acquisition process. Some methods consist of highlighting the landmarks or placing some critical points, posteriorly to the acquisition of the data such as on [6,71]. The advantages of the Point-Based Registration is that it can be applied to any imaging modality where markers or landmarks are always visible among every image. The registration approach proposed by this work can be defined as being derived from the Point-Based Registration, however, the landmark is found automatically.

The reason for applying intrasubject registrations on cardiac images is related to the movement of the heart, breathing and the actual displacement of the patient's body in relation to the acquisition device. All these occasions may induce a noticeable displacement and distortion between the retrieved slices. Cardiac intrasubject registrations are, for instance, a more complex problem than intrasubject brain registrations, in particular because of the non-rigid and mixed motion of the heart and thorax structures [56].

The cardiac data is usually acquired along with the aid of an electrocardiogram that will indicate the best moment during a cardiac cycle where the data should be recorded. The entire cardiac cycle is shown in Figure 6. A single slice is acquired at successive intervals of the cardiac cycle. However, due to the motion of the heart, the data does not remain consistent among these acquired slices. Moreover, several cardiac cycles are usually required to reconstruct a single slice. When possible, patients are asked to retain their breath (15–20 s) during the acquisition to reduce the thorax motion influence [56].



Figure 6: A classical acquisition with an electrocardiogram-gated sequence [56].

2.5 SEGMENTATION

Several approaches for image segmentation employ commonly used procedures such as thresholding [72], atlas or multi-atlas registration [8,73,74], clustering [75], edge detection [76], level set [77], active contour [78], region growing [79] and many others [80,81]. We define the act of segmenting an image using classification algorithms as classified segmentation. This procedure is also addressed in the literature as pixel classification or probability based segmentation [80]. We tend to think that classified segmentation is the best terminology since pixel classification may not be related to segmentation and probability based segmentation is an ambiguous term. The classified segmentation has been used in multispectral MR segmentation [82]. However, it is not a very commonly applied procedure for image segmentation.

The classified segmentation can be viewed as a simple iteration through a set of pixels or voxels of an image or 3D model where a set of characteristics related to the iterated pixel, voxel or surrounding area is extracted. These features are called features vector and each feature is illustrated as the variable b in Figure 7. Vectors of this type will usually compose a dataset that is provided as input to a classification algorithm, which, in turn, generates a predictive model that is used to define the class of incoming (unknown) pixels.



Figure 7: A features vector based on a pixel and its related information.

2.5.1 CLASSIFICATION ALGORITHMS

Machine learning algorithms are often divided in two main categories: (1) the supervised and (2) the unsupervised methods. The algorithm is categorized as supervised when it explicitly evaluates the class or label attribute of a training set as the predictive label desired to attach to an incoming unlabeled instance. Furthermore, when this assumption is formalized, the class attribute heavily induces the generated predictive model, since the algorithm usually minimizes the error of the predictive model based on this class. However, when not formalized, the algorithm is defined as unsupervised and the class plays no heavy influence but of a normal attribute, when it is not disregarded from the training. Classification algorithms are always categorized as supervised learning methods while clustering algorithms are often unsupervised.

Support vector machine (SVM) [83], multilayer perceptron (MLP) [84] and relevance machine (RVM) [85] are some of the most popular classification algorithms applied to create predictors. SVM and RVM make no assumptions about the data, they are able to find the global minimum of the objective function and can provide near optimal performance. Moreover, the complexity of these techniques depend on the number of support or relevance vectors, but not on the dimensionality of the input space. However, predictors based on these techniques provide too little insight on the importance of the variables involved on the prediction. The transparency of the predictive model is also very important in some areas, such as medical decision support and quality control [86]. SVM and MLP have been used for classifying diseases and providing diagnosis [87,88], segmentation [89,90], weather prediction [91] and many other tasks.

Decision and regression trees, in the other hand, are known for their transparency. However, they are rather sensitive to small perturbations on the learning set. It has been demonstrated that this problem can be reduced by applying bagging. Bagging predictors is a process of generating multiple predictive models based on a variation of the training set and using these to get an aggregated model [92]. One of the most acclaimed decision tree algorithms is the RandomForest (RF) proposed by Breiman [93]. RF is a combination of the random subspace method proposed by Ho [94] and bagging. RandomForests have been used for a large variety of tasks including identification of DNA-binding proteins, segmentation of video objects, classification of hyper-spectral data and many others [86].

Verikas et al. [86] demonstrates how RandomForest outperforms and is also outperformed by several machine learning algorithms on distinctive tasks. According to the No Free Lunch theorem there is no single classifier model that is the best for all problems [95]. An analysis of how and why RandomForest is usually included on the top of the ranking of classification algorithms, outperforming, for instance, SVM and MLP on several tasks is valuable instead.

Furthermore, Verikas et al. [86] also demonstrate how RF is a committee of weak learners for solving prediction problems. A decision tree is used as a weak learner in RF. When solving classification problems, the RF prediction is the un-weighted majority of class votes. As the number of trees in RF increases, the test set error rates converge to a limit, meaning that there is no overfitting in large RFs. In this work, we also complement this statement, in a slightly distinct fashion, with clear evidence of the non-overfitting characteristic of RF if compared to other classifiers.

Meyer et al. [96] assessed the SVM performance on a large-scale comparison including 16 classification and 9 regression techniques where the input parameters were carefully selected. The benchmark assessed 21 datasets, while 12 datasets were used for regression tests. The performance was evaluated using 10 times repeated 10-fold cross validation. Thus, on the classification tasks, SVM always ranked on the top 3 classifiers except for two datasets. However, SVM was outperformed in 10 out of 21 data sets. On the regression tasks, SVM was almost always on the top 3. Nevertheless, SVM was outperformed in all occasions with the exception of two.

Finally, neural networks are also powerful tools that can be used to approximate the complex nonlinear input-output relationships efficiently. On a classification problem, the objective is to learn the decision surface that accurately maps an input feature space to an output space of class labels. Among various architectures reported in the literature, Radial Basis Function (RBF) network is gaining attention due to its localization property of Gaussian function and has been widely used in classification problems [97].

2.5.2 APPLICABILITY OF THE CLASSIFIED SEGMENTATION

Image segmentation has been used in biomedical areas such as on the identification of lung diseases; on the automated classification of white blood cells; in the detection of cancerous cells and in chromosome karyotyping [81]. It has also been used in chest [98] and liver segmentation [80], in breast segmentation [99] and many others.

Rikxoort et al. [100] have proposed the use of the k-nearest neighbor algorithm for segmentation of the liver on CT images. The core of the method consists of a voxel labeling procedure: (1) for every voxel in the test set an amount of numerical values (a features vector) is computed and (2) a statistical classifier, trained on previously extracted features vectors, evaluates if the analyzed voxel is or is not part of the liver.

The features extracted on the approach of Rikxoort et al. [100] were: (1) the coordinates of each voxel, (2) the grey value of the voxel, (3) a Gaussian weighted neighborhood of the voxel and (4) three extra features provided by the use of an atlas. The three extra features were generated on the basis of an atlas and they define if an arbitrary voxel is above, behind or to the left of the same. Furthermore, the authors stated that the extraction of the features (1), (2) and (3) alone were not sufficient to achieve a satisfying segmentation. In conclusion, they mention that the obtained results are satisfying but there is still a lot of room for improvement. Furthermore, their computation time is relatively large. It took 3 minutes to register 12 training scans to a single patient scan and some extra minutes to classify and preprocess the images.

2.5.3 SEGMENTATION OF THE CARDIAC ADIPOSE TISSUE

Some of the first semi-automated segmentation methods for the epicardial fat were proposed around 2005. Dey et al. [101] apud Coppini et al. [6], for instance, apply a preprocessing step to remove all other structures apart from the heart by using a region growing strategy. Thereafter, an experienced user is required to scroll through the slices to place from 5 to 7 control points along the pericardium border, if visible. Therefrom, Catmull-Rom cubic spline functions are automatically generated to obtain a smooth closed pericardial contour. Finally, the epicardial fat is simply quantified by thresholding, since it is theoretically located within this generated contour. In Pednekar et al. [102], a method for the segmentation of abdominal adipose tissue was proposed. The work of Kakadiaris et al. [103] have further extended the method of [102] to the segmentation of the epicardial fat.

Coppini et al. [6] focused on reducing the user intervention. On their method, an expert is still necessary to scroll through the slices between the atrioventricular sulcus and the apex in order to place some control points on the pericardium. The amount of essential points is not clearly described. Nevertheless, the required amount of slices to be analyzed is apparently lesser than the ones required on the method proposed by Dey et al. [101]. Moreover, they also present their solution on a 3D space, and claim that Dey et al. [101] do not. The overall focus of their work was to describe their method mathematically. However, the work lacks on describing the general accuracy of their method.

Barbosa et al. [104] proposed a more automated segmentation method for the epicardial fat. They start using the same preprocessing method from Dey et al. [101] and further apply a high level step for identification of the pericardium. Their identification is done by the act of tracing lines originating from the heart's centroid to the pericardium layer and interpolating them with a spline. Although this approach may be interesting, of simple complexity, and highly applicable for virtually any proposed method for this issue, the reported results are not impressive. Only 4 out of 40 images were correctly segmented in a fully automatic way.

Shahzad et al. [8] proposed, as far as we know, the first fully automated method for epicardial fat segmentation in 2013. Their method uses a multi-atlas based approach to

segment the pericardium. The multi-atlas approach is based on registering several atlases (8 in this case) to a target patient and fusing these transformations to obtain the final result. The authors selected 98 patients for the test and reported a Dice similarity index of 89.15% to the ground truth, and a low rate of approximately 3% of unsuccessful segmentations. Notwithstanding, they did not provide any measurements of the overall processing time.

Ding et al. [9], in 2014, proposed an approach similar to the one from Shahzad et al. [8]. The authors segment the pericardium using an atlas approach, which consists of a minimization of errors after applying transformations to the atlas along with an active contour method. The mean Dice similarity coefficient was equal to 93% and they claim that their result was achieved in 60 seconds on a simple personal computer. Although their segmentation seems to be the most precise in the current literature, the reported computing time is a dubious and poorly described. In addition, 60 seconds may be considered too fast for segmenting and transforming an entire scan, which consists of roughly 50 images. They also present a work [73] that segmented the aorta instead of the pericardium, and compare their achieved time (60 seconds) to the 15 minutes of the former. If these 60 seconds correspond to just the time it takes for the algorithm to minimize the transformations, then this comparison is not feasible. Furthermore, they report that on their approach the atlases' images were pre-aligned to a standard orientation, therefrom, there is a comparison with only one of these atlases to speed up the process. The remaining pericardium contour will follow the pre-aligned pattern, which is a reported limitation. Besides, they did not describe how each one of these atlases is chosen as the correct one for each possible case.

Although the epicardial fat segmentation have been softly addressed in the literature, on the other hand, the mediastinal fat have not been addressed as a target for automatic segmentation despite the fact of being a cardiovascular risk marker as previously shown in Section 2.2.

2.5.3.1 ADIPOSE TISSUE IN HOUNSFIELD UNITS

Computed tomography (CT) holds two important advantages over conventional radiographs: three-dimensional image reconstructions and the capability to quantify X-ray attenuation. Attenuation is expressed in CT as Hounsfield Units (HU). The X-ray beams used

for diagnostic radiology are not monochromatic (i.e., consisting of photons with only one energy level), instead, they are composed of photons with a broad spectrum of energies [105]. Molteni [106] explains how HU may be different for a single type of material among distinct CT apparatus or even between the same machine model if different technical factors are applied such as distinct interpolators. In fact, each system and manufacturer incorporates a unique combination of X-ray source, detector array and projection geometry. Hence, when aiming to segment anything based on HU, this variation should be properly accounted.

For recent systems, the available range of CT is usually between 2¹²(4 048) and 2¹⁶ (65 536) for 12-bit and 16-bit types of acquisition, respectively. Thus, the maximal range for values in CT scans varies along with technical factors and is specified within each type of equipment. Hounsfield units correspond directly to grayscale values (i.e., the previously described attenuation). When the HU data is rendered to an image in order to be displayed on the screen, for instance, it should and is usually reduced to a representation within the interval of 256 values (8 bits). For that matter, the actual visualization of CT data has to be calibrated according to the range that corresponds to an arbitrary area of interest (i.e., fat, muscle, bones, etc), or even to an interpolation of the whole range, which, in the case of CT scans, induces a massive loss of data [106].

The values of Table 1 are present on the work of Molteni [106] and correlate a given substance to a mean of HU. In order to properly access the cardiac adipose tissue within CT scans, we need to consider an interval around -100 HU, which, in turn, corresponds to the overall fat tissue of the human body. Coppini et al. [6] and Shmilovich et al. [107] defined the cardiac adipose tissue interval as (-190,-30) while Spearman et al. [10] defined as (-195, -45) and Shahzad et al. [8] as (-200,-30). For this work, we will consider the largest proposed interval, which corresponds to the one used by Shahzad et al. All the addressed intervals fit properly on a 8 bits-depth image and no actual interpolation is required, avoiding a possible loss of data.

Material	HU value	Notes
Air	-1000	In a vacuum for all practical effects
Fat	-100	
Water	0	Distilled, at standard temperature and pressure
Muscle	+40	
Blood	+40	
Bone	>400	Spans over a large range, to approximately +1200 and occasionally more
Aluminum	2640	At 60keV

Table 1: Values of Hounsfield units (HU) for (fan beam) medical CT scans.
3 PROPOSED APPROACH

As previously discussed, the overall CT image data encapsulated in a DICOM file, even when belonging to the same modality, comes in different manners and is, therefore, not standardized. Thus, considering two arbitrary scans of a single patient or distinct patients, there is variation on the following topics, namely: (1) position, regarding the supposed center of the heart; (2) scale, regarding the distance from where the patient was from the scanners' sensors or if it was rescaled along the acquisition process; and on the (3) texture. The texture difference is partly due to the fact that distinct manufacturers apply distinct formulas, interpolators and physical sensors on the acquisition and post-processing. To a fully extent, the differences in texture may also be related to the physical environment from where the data was acquired.

The images processed and analyzed in this work always had their pixel values belonging to the fat interval: (-200,-30) in Hounsfield units. We have addressed the details about accessing and exporting images within that given range in the Appendix A – Accessing CT Data from DICOM Files. The image shown in Figure 8 is an instance where its data, apart from the black tone, represent only the adipose tissue range of a CT-DICOM file. In other words, the black color (0) represents the background, that is, what is not within that given range.



Figure 8: An image within the (-200,-30) HU range.

The images in Figure 9 were taken from four distinct patients (on the axial-plane) and they were chosen due to being relatively close to the shoulders of each patient. By assuming that the images of the cardiac CT scans are aligned in a stack, we define their ordering as the first one being closer to the top (head) of the body, which is also known as craniocaudal direction. Furthermore, in this case, these images are represented on a greater HU range (-200,500) just for a matter of analysis and comparison.



Toshiba

Phillips

Figure 9: Variation between four distinct patients on the same HU range.

These images evidence a variation on scale, positioning and texture among distinct manufacturers. The first two images were acquired with the same scanner (although the texture seems not to change, there is a divergence on the scale and a slight divergence on the position). The third patient data was acquired with a second distinct scanner, and the fourth with a third distinct scanner. Hence, in total, there are three distinct manufacturers for these images. In Figure 10, the same images of Figure 9 after the appliance of the proposed registration are shown. This time, the images are within the adipose tissue range.



Toshiba



Figure 10: Four intersubjectly registered patients.

We can consider by the images in Figure 10 that they are much more aligned between themselves and within a corresponding scale than the ones shown in Figure 9. The texture difference, when it comes to the segmentation, will be partially accounted and processed by the classification step. Registrations virtually always neglect the textural variation, such as on intersubject registrations. The information about the object, certain content, shape or area is regarded instead. When the standardization of the texture is desired, it is usually done on a separate preprocessing step. However, the variation of the texture within the adipose tissue range of the CT images is almost unnoticeable, at least between the third and fourth images of Figure 10. We assume that the classification step of our approach will be able to properly predict the type of fat of the pixels even with remaining textural, positioning and scaling variations such as the ones that remain on the instances of Figure 10, variations that the proposed registration method alone is not able to standardize.

3.1 PROPOSED REGISTRATION

We have already defined that this work will be based on images created as from the adipose tissue range that are extracted from DICOM files with regard to computed tomography. Given these fat images of several patients, we are then trying to find a practicable registration to be applied, which will standardize the positioning and scale of the heart among several patients. It is important to highlight that, when working with fat-only images; the quantification of the fats comes even easier. After being segmented, it can be simplified as literally counting each segmented pixel.

In summary, the registration proposed in this work comprises two steps: (1) the scaling and (2) the translation. The DICOM structure stores information related to the scaling [108,109], and, therefore, we will address that standardization in the Appendix B – Rescaling Images According to DICOM. In summary, the images are stored in distinct scales within a DICOM file and an attribute of the same indicates by how much the image should be rescaled in order to transform it back to its natural proportion. Thus, our approach does not consider much of the scaling issue since it is a trivial operation and, therefore, it focus on the autonomous translation that should be applied to each patient.

Moreover, one may suppose that it is sufficient to apply a rescale operation based on the DICOM information on a patient image followed by a translation operation, which centers the rescaled image before being rasterized. However, apart from the variation on the positioning of every patient, when a rescaled slice is rasterized and rendered to the screen or to an image as Figure 11 shows, part of the heart may be cropped out on that process. This occasion, therefore, evidences the need of an intelligent translation to be applied jointly or after the rescaling.



Figure 11: Heart exceeding the boundaries of the image.

In average, there are 50 slices or images on every regarded cardiac CT scan. Our approach is based on registering a single slice of an arbitrary patient and further applying the same transformation to the remaining slices to reduce the computational time and simplify the method. Therefore, suitable slices for registration will always be the ones closer to the top of the body (on the craniocaudal direction) such as the one in Figure 11. Furthermore, since there is no external marker applied to each image of the scan, we need to rely on patient-related content only. This kind of registration would be categorized as of intrinsic nature [42]. We later merged a landmark approach with an atlas approach, conceiving the proposed method.

In order to autonomously perform the registration of several patients despite their scale and actual positioning of their heart, no manual placement of any common landmark is likely to be applied, such as several works have proposed [11,101,102,103]. Therefore, the remaining alternative is to automatically find the landmark. Thus, in other words, the parameters of our transformation are "searched for" and determined by finding an optimum of some function on the search space.

The subject type of our solution cannot be intrasubject. In fact, we do not want to align structures of a patient based just on its own information. Conversely, structures of several distinct patients have to be aligned to some extent. Hence, we define the proposed registration as a fusion of an intersubject and atlas registration. The proposed registration is defined as so due to the fact that we construct an atlas of a certain area based on a few patients, search for alignments for this atlas and align the patients to the chosen position. Therefore, our approach incorporates characteristics of a model and of an atlas. Moreover, the registration can be considered intersubject in a sense that it aligns distinct patients to a common position. Summarily, according to the definitions of Maintz et al. [42], our proposed registration approach is categorized as a registration of intrinsic nature where the parameters are search for and the transformation applied is affine.

Thus, the proposed registration consists of a combination of (1) a landmark approach and (2) an atlas approach. With respect to the landmark approach, regions that exhibit or represent a common pattern among the evaluated instances are the most eligible for the placement of a landmark. A common characteristic that is of relatively recognizable easiness is exhibited from the 1st to approximately the 20th slice of every patient, regarding the craniocaudal direction. That common characteristic is denominated retrosternal area and was selected to be automatically recognized. The retrosternal area is located on the back of the sternum represented as red spots in Figure 12 and does not vary greatly as other cardiac structures do. Besides, it always appears within the boundaries of the CT image, i.e., it is almost never accidentally cropped off. The slice shown on the image (e) of Figure 12, despite being the 20th slice of a patient, is still exhibiting the same pattern in relation to the images (a), (b) and (d), which, in turn, are slices of the top of the CT stack.



a) 4th slice

b) 5th slice



Figure 12: Highlighted retrosternal area of four distinct patients.

Summarily, the proposed registration heavily relies on this common characteristic denominated retrosternal area. After automatically recognizing the area, its central point is used as reference in order to align the images of the patients to a standard position. The Figure 13 below illustrates the overall steps of the registration. More details regarding the registration are further discussed in the following section.



Figure 13: Overall steps of the proposed registration.

3.1.1 RETROSTERNAL AREA RECOGNITION

The proposed recognition is based on similarity measures [70]. Thus, in order to apply similarity measures to our problem, a small atlas of the retrosternal area was constructed where its central point corresponds to the common landmark among every patient. This atlas is then, by a simple analogy, moved on top of a fixed image while associating every single position (x, y) of the atlas to a score of similarity between the pixel values of the fixed and the atlas (moving) images.

We have randomly selected 10 instances and manually aligned these images to compose the atlas as shown in Figure 15. Due to the texture variation among distinct manufacturers, we chose, prior to the atlas assemblage, to threshold the images as the instance depicted in Figure 14-(b) shows. In other words, each one of these 10 images received a threshold at level 0.2, which originates the binary images as the one shown in Figure 14-(c). The reason for choosing 0.2 as value for thresholding was to remove possible remnants of interpolations from the image, since the images have already been rescaled from the DICOM file to their natural proportions. However, the approach still behaves apparently the same if the chosen value for thresholding is lower than that.



a) Threshold operation at level 0.2. The graph represent the image's histogram.





c) Binary version of the image (b) after being thresholded at level 0.2.

Figure 14: A thresholding process.

After the thresholding operation, the randomly chosen images were overlapped and the pixel values of each position was summed and at last divided by the amount of selected instances (10 in the case). The Figure 15 contains each one of the instances that composes the atlas applied to this work. The resultant image (i.e., the atlas) is shown in Figure 16 along with its histogram.



Figure 15: Ten randomly chosen binary instances of the retrosternal area.



Figure 16: The atlas applied on the registration.

Therefrom, we need a measure to assess how much of the atlas is associated to an arbitrary part of a fixed image. A successful registration was defined as every single trial where

the algorithm-chosen retrosternal landmark coincided with the actual retrosternal area's center of the fixed image, regarding slightly positioning variations within this area. A total amount of 52 randomly chosen slices was provided as input to evaluate the following described similarity measures.

3.1.1.1 MEAN DIFFERENCE

The mean absolute difference, mean squared difference and mean cubic difference are very simple measures for similarity and can be obtained from the minimization of the Mean Difference (MD) in Equation (3) regarding g = 1, g = 2 and g = 3 respectively. Mdenotes our atlas (i.e., moving image) shown in Figure 16, h and w stand for the height and width of the atlas, respectively [110].

$$MD_{y,x}(F, M, g) = \frac{1}{hw} \left(\sum_{i=y}^{h+y} \sum_{j=x}^{w+x} |F_{i,j} - M_{i,j}|^g \right)$$
(3)

The chosen position (x, y) for the atlas corresponds to the minimum value among all the possible $MD_{y,x}$ positions of F. We have defined as possible all the positions available in Fplus a small extrapolation on the upper-left position. Thus, if a pixel that does not belong to Fis accessed, then this nonexistent pixel is treated as a background pixel. If the chosen position for the atlas corresponds to (x, y) then, alternatively, the chosen position for the landmark corresponds to the center of the retrosternal area that is equal to $(x + \frac{w}{2}, y + \frac{h}{2})$, which is also the center of the atlas. The rate of successful recognitions obtained with this measure was equal to 63.5%.

This measure, in an intrinsic fashion, gives more importance to certain types of pixels that are more abundant on the atlas image due to its equal-weighted nature. Darker pixels correspond to a great amount of pixels on the atlas and, due to that fact, the behavior depicted on the instances of Figure 17 is occasionally induced. In these occasions, the matching of darker pixels influence more on the final result due to the uncommon thorax and retrosternal structures of these patients. This behavior occurs more frequently when there is a thick layer of fat surrounding the heart of the patient, as opposed to the ones when there is not, which are shown in Figure 18. In addition, if the darker area of the atlas image is reduced (e.g., by resizing it), the brighter pixels are then favored and the atlas eventually ends up on the wrong position (usually on top of the abdominal fat) in a very similar fashion.

The following instances of Figure 17 and Figure 18 had their pixels of the atlas image shifted to the blue layer. The green (+) symbol illustrates the center of the atlas and the red (-) symbols indicate dark areas of the atlas image.



Figure 17: Evidence for occasional darker pixels preference (high fat patients).



Figure 18: Successful recognitions (low fat patients).

It is extremely difficult to non-empirically define the right proportion of the darker and brighter area on the atlas image. A complete empirical analysis is also impossible due to the restrict number of instances we have to rely on. Instead, our further proposed hybrid measure solves that hassle by dividing the image in two main areas and measuring the pixel deviation differently for each one of these parts (darker and brighter pixels). However, we have also evaluated the normalized cross correlation, mutual information and a weighted mutual information as similarity measures and compared the achieved results to the proposed measure.

3.1.1.2 NORMALIZED CROSS-CORRELATION

The main use for correlation was studying random-like processes that exhibit similarity in their behavior of occurrence. A good example would be the temperature of the air, which is certainly correlated with seasons [111]. The discrete equation for the cross-correlation coefficient applied to images is defined on the Equation (4). The equation represents the cross-correlation between the images F and M and its intensity values ($F_{i,j}$ and $M_{i,j}$) at the position (x, y) of the fixed image, where h and w still represent the height and width of the atlas, respectively. The variables μ_F and μ_M represent the mean intensity values of the fixed and moving images, also respectively [110]. The maximal value that $CC_{y,x}$ can achieve is 1, which would imply that the images are in alignment [110]. This measure allows to register objects whose intensity values are related by a linear transformation [46].

$$CC_{y,x}(F,M) = \frac{\left|\sum_{i=y}^{h+y} \sum_{j=x}^{w+x} (F_{i,j} - \mu_F) (M_{i,j} - \mu_M)\right|}{\sqrt[2]{\sum_{i=y}^{h+y} \sum_{j=x}^{w+x} (F_{i,j} - \mu_F)^2 \sum_{i=y}^{h+y} \sum_{j=x}^{w+x} (M_{i,j} - \mu_M)^2}}$$
(4)

The cross-correlation achieved a slightly lower rate of successful recognitions (55.7%) if compared to the similarity measures based on the difference of the related pixels. That low rate may be due to the fact that the atlas and the fixed images are not always related by a linear transformation. The relation between both images is generally more fuzzy than just a linear correlation. In theory, the mutual information measure would perform better because it considers the mutual dependence.

3.1.1.3 MUTUAL INFORMATION

Mutual information is a measure of how well one image explains the other [112]. When two images are composed by entirely distinct pixel values, mutual information is zero. The maximal value of mutual information is 1. The mutual information coefficient is defined as MIin Equation (5), where m is equal to the pixel values within the atlas image M, f is equal to every pixel value of the fixed image F that is within the atlas image size, i.e., between the topleft corner of the atlas and the bottom-right corner, respecting the (x, y) position of the atlas. The variables ρ_F and ρ_M are the marginal probability distributions, corresponding to intensities in the images F and M, respectively [110]. Furthermore, H(K) stands for the Shannon entropy of the histogram K [112] and ρ_{FM} stands for the joint probability.

$$MI_{y,x}(F, M, g) = H(F) + H(M) - H(F, M)$$

= $\sum_{f \in F} \sum_{m \in M} \rho_{FM}(f, m) \log_g \frac{\rho_{FM}(f, m)}{\rho_F(f)\rho_M(m)}$ (5)

The variable g usually receives 2, 10 or e as input. The achieved rate of successful recognitions was equal to 61.5%, which is also equal to the mean squared difference measure and higher than the rate obtained with the normalized cross-correlation.

3.1.1.4 WEIGHTED MUTUAL INFORMATION

In the traditional formulation of the mutual information, each event or object specified by (f, m) is weighted by the corresponding joint probability $\rho_{FM}(f, m)$. This assumes that all objects or events are equivalent apart from their probability of occurrence. However, in some applications, it may be the case that certain objects or events are more significant than others, or that certain types of associations are more semantically important than others [113].

For instance, the mapping $\{(9,9), (5,5), (2,2)\}$ may be viewed as stronger than the mapping $\{(9,2), (5,9), (2,5)\}$, although these relationships would yield the same mutual information. In other words, assuming that the image $E = \{9,5,2\}$ and $R = \{2,9,5\}$ are being measured by the *MI* equation, the result of the comparison between *E* and *E* or *E* and *R* would be the same. However, in our problem, it is clear that the images *E* and *E* are more similar than *E* and *R* although they hold the same mutual information. Thus, we propose a combined measure which considers both the mean difference measure and the mutual information, originating a weighted by the difference mutual information shown in Equation (6).

$$WMI_{y,x}(F,M,g) = \sum_{f \in F} \sum_{m \in M} \frac{1}{|f-m|+1} \rho_{FM}(f,m) \log_g \frac{\rho_{FM}(f,m)}{\rho_F(f)\rho_M(m)}$$
(6)

The factor $\frac{1}{|f-m|+1}$ defines the applied weight to the equation. When f = m each parcel of the sum of the *MI* equation will be multiplied by 1. Thus, in this case, the maximum value of *WMI* will still be 1 and that coefficient will be achieved when the images are the same and are perfectly aligned. The weighted mutual information measure achieved a successful rate of 70%, which is higher than all the results achieved by the other measures.

3.1.1.5 HYBRID MEAN DIFFERENCE

Although the weighted mutual information measure achieved a relatively high rate of successful recognitions, we did not consider that rate high enough. We focused on minimizing the error during the registration by the tightest amount possible. Thus, by evaluating the circumstances, we ended up with a Hybrid Mean Difference (HMD) measure capable of achieving a successful rate of recognitions of the retrosternal area higher than 70%.

The proposed measure is based on the premise of solving the issue introduced by the mean difference measure. At first, by thresholding the moving (atlas) image on the level t, we separate two main parts of the image: (1) the one with darker pixels and (2) the one with brighter pixels, as shown on the image (c) of Figure 19. Therefrom, the brighter moving image area will have its pixel values m inverted and subtracted of the values f from the fixed image (b), as the images (d) and (f) depict. If the difference between f and m is positive then this value is summed to compose a partial error score S_p , as shown at the bottom of the image (f).

Furthermore, the pixel values of the moving image that belong to the darker area, depicted as black on image (c), will be subtracted of the fixed image (b) to compose the image (e). The sum of the pixel values on the image (e) will compose the remaining partial error score S_n shown at the bottom of image (e). The final hybrid difference is given by $S_n - S_p$. By minimizing HMD, we automatically maximize S_p and minimize S_n altogether. When the images F and M are identical, HMD will return 0. However, with distinct images it will return negative values if $S_n < S_p$.

In summary, by applying this measure we are maximizing the intersection of the brighter pixels of the moving image to the ones of the fixed image while reducing errors that occur when brighter pixels of the atlas match darker pixels of the fixed image. This type of errors occur because the area of brighter pixels of the atlas is virtually always greater than the brighter pixels of the retrosternal area of the processed fixed image. As such, if the errors related to that surplus area are reduced, so is the recognition error provoked by the similarity measure. In addition, the intersection of the pixels is properly weighted by grey values of both images such as in MD. Finally, at the same time, we are minimizing the intersection of the darker pixels of the moving images to the brighter pixels of the fixed images.



Figure 19: The hybrid mean difference similarity measure.

The function in Algorithm 1 represents logical steps to achieve the hybrid mean difference, where t stands for the threshold level, F stands for the fixed image, M stands for the moving image, M.x and M.y stand for the position (x, y) within F where the upper-left pixel of M is placed. Furthermore, g stands for the exponent of the difference between the pixel values of F and M. Besides, we consider that the values $F_{i,j}$ and $M_{i,j}$ are normalized. The parameter t was empirically chosen to be 0 as well as g to be 3 (if r is an integer variable). The achieved accuracy of the proposed measure was equal to 92%. In addition, the Equation (7) mathematically illustrates the Algorithm 1.

```
1. hybridMeanDifference(Image F, Image M, Int g, Int t)
```

```
double score = 0
1.1.
          for i = 0; i < M. height; i + +
1.2.
             for j = 0; j < M.width; j + +
1.2.1.
                 double r = 1
1.2.1.1.
                 if M.x + j < M. width and M.y + i < M. height
1.2.1.2.
                     if M_{i,i} > t
1.2.1.2.1.
                        r = -\left(\max\left(F_{i,j} - \frac{1}{M_{i,j}}, 0\right)\right)^g
1.2.1.2.1.1.
                     else
1.2.1.2.2.
                        r = |F_{i,i} - M_{i,i}|^g
1.2.1.2.2.1.
                 score + = r
1.2.1.3.
1.3.
          return score
```



$$HDM_{y,x}(F, M, g, t) = \sum_{i=y}^{h+y} \sum_{j=x}^{w+x} \left\{ -\left(\max\left(F_{i,j} - \frac{1}{M_{i,j}}, 0\right) \right)^g , if M_{i,j} > t \\ |F_{i,j} - M_{i,j}|^g , otherwise \right\}$$
(7)

In summary, we have proposed an extension of an existing similarity measure called mean difference. This extension performed better than the normalized cross-correlation, mutual information and the proposed weighted mutual information as well. The highest rate of successful recognitions achieved by HMD was equal to 92% against 70% of the weighted mutual information measure. In other words, once the HMD was applied, the proposed recognition method was able to successfully locate the retrosternal area in 48 instances out of a total of 52. A comparison between the similarity measures is shown in Table 2.

Measures	Successful
	Recognitions
Hybrid Mean Difference*	92%
Weighted mutual information*	70%
Mean difference	63.5%
Mutual information	61.5%
Normalized cross-correlation	55.7%

Table 2: A comparison of the similarity measures using 52 images.

3.1.2 SCALE INVARIANCE

We have addressed the fact that CT slices can be trivially rescaled according to the information in their DICOM file. However, the proposed recognition, using the HMD measure, successfully works with slices of varied scales. That is, the recognition succeeded on images of pixel spacings within the (0.20,0.50) interval. In addition, approximately 95% of the patients we have checked had their associated pixel scaling data belonging to that interval. Hence, the proposed recognition is basically scale invariant for all the possible cardiac CT images. In other words, the method does not need to be adapted to work with slices of different scales. The Figure 20 shows 3 examples of successful recognitions of the retrosternal area regarding differences of 0.02, 0.1 and 0.15 on the pixel spacing information, respectively.







Pixels Spacing k - 0.02



Pixel spacing: n

Pixel spacing: n - 0.15

Figure 20: Successful recognitions on varied scales.

As evidenced by Figure 20, the proposed method is virtually scale invariant, the size of the atlas does not need to change to match the scale of the fixed image. In other words, the size of the atlas does not depend on the size of the fixed image and can be thought as a constant. Therefore, the complexity of the Algorithm 1, which loops through the pixels of the atlas image only, may also be considered constant. Thus, the *hybridMeanDifference* function runs in O(1) time for each pixel of an arbitrary fixed image. Therefore, to assess the measure through an entire fixed image searching for the most suited position implies a O(wh) time, where w and h are now the width and the height of the fixed image, respectively.

3.1.3 CONFIRMATION METHOD

Although the proposed similarity measure (HMD) achieved a high rate of successful recognitions (92%), it was not successful on all trials. Thereafter, we further incremented the proposed recognition with a heuristical confirmation method. The proposed confirmation is responsible not only to increment the rate of successful recognitions but to prevent any further unaccounted variation that is not present in the patients regarded for this study. Such variations may include anatomical defects in the retrosternal area induced by surgery or genetic disposition, for instance.

Given a small area of pixels A at the center of the atlas (i.e., the center of the recognized retrosternal area), there should be two points p_l and p_r that continously move through the fat pixels on the left-bottom and right-bottom directions until they reach convergence. A fat pixel is defined as a pixel that is not background (i.e., black). The thin slanted white lines on the instances of Figure 21 illustrate the binding of, or the distance between these two points p_l and p_r after their convergence. The action of moving these points on the image is just a fragment of the whole confirmation method. The movement of these points will provide information for two logical conditions that reinforce the autonomously recognized position (x, y) for the atlas as correct or not. The two logical conditions are defined as such: (1) the traced line must be within a certain width (in the x-axis) and (2) both points must also be within a certain distance from each other and from the starting point as well. If this confirmation fails, the settlement of the atlas and the confirmation method must be redone jointly for all the positions in the fixed image.



Figure 21: Binding of the two points p_l and p_r .

After the retrosternal area recognition, we consider a rectangular area A where its central point corresponds to the central point of the atlas image. This central point is illustrated as the pixel pointed by the arrow in Figure 22 and the rectangular area is also illustrated in yellow in the same figure. The reason for considering a rectangular area instead of a single point is so that if a single point is selected, not always will the two points p_l and p_r be able to move through fat pixels because not every patient has a straight continuity of fat deposits in the retrosternal area as Figure 23 evidences.



Figure 22: The rectangular area and the central point of the atlas.



Figure 23: Absence of continuous fat deposits in the retrosternal area.

The aspect ratio of A was obtained based on observations of that area over several instances. There usually are more gaps on the horizontal direction than on the vertical and, therefore, the width of the rectangular area was set to be greater than its height. Following

that premise, we have empirically chosen the width to be equal to approximately 13% of the width of the fixed image and the height to be approximately 4% of its height.

Thereafter, from every pair of pixels within A there must be at least one pair p_l and p_r that satisfies the function shown in Algorithm 2, where Z(p) is a function that returns true if the pixel p passed as parameter is a fat pixel and false otherwise. The two dimensional vector dX stipulates the minimum and maximum distances that these points p_l and p_r must move on the x-axis. The parameter c stands for the convergence threshold, i.e., how many times both coordinates p_l and p_r will move on the image. Furthermore, l stands for how many background pixels can be skipped on a single movement of the coordinates p_l and p_r . Finally, the methods moveLeft(), moveDown(), moveRight() and moveUp() are self-explanatory, they will essentially move the coordinates of the arbitrary point p by 1 pixel on the orientation described by the name.

¹ confirmationMethod(**Image** *F*, **Vector2** *p*_{*l*}, **Vector2** *p*_{*r*}, **Vector2** *dX*, **Int** *c*, **Int** *l*) **vector2** $iP = \left(\frac{p_{l.}x + p_{r.}x}{2}, \frac{p_{l.}y + p_{r.}y}{2}\right) //\text{center of the initial pair of points}$ 1.1. for $c_a = 0$; $c_a < c$; $c_a + +$ 1.2. for $l_a = 1$; $l_a \le l$; $l_a + +$ 1.2.1. vector2 $p_a = p_1$ 1211 vector2 $p_b = p_r$ 1.2.1.2. if $Z(F_{p_1,y,p_1,x})$ 1.2.1.3. **do** l_a **times**: p_l . moveLeft() 1.2.1.3.1. if $\neg Z(F_{p_1,y,p_1,x})$ 1.2.1.3.2. **do** l_a **times**: p_l . moveRight() 1.2.1.3.2.1. **do** *l_a* **times**: *p_l*.*moveDown*() 1.2.1.3.2.2. $\mathbf{if} \neg Z(F_{p_1.y,p_1.x})$ 1.2.1.3.2.3. **do** l_a **times**: p_l . moveLeft() 1.2.1.3.2.3.1. if $\neg Z(F_{p_1,y_1,p_1,x})$ 1.2.1.3.2.3.2. **do** 2*l*_a **times**: *p*_l. *moveRight*() 1.2.1.3.2.3.2.1. if $\neg Z(F_{p_1,y,p_1,x})$ 1.2.1.3.2.3.2.2. **do** l_a **times**: p_l . moveLeft() 1.2.1.3.2.3.2.2.1. **do** *l_a* **times**: *p_l*.*moveUp*() 1.2.1.3.2.3.2.2.2. if $Z(F_{p_r,y,p_r,x})$ 1.2.1.4. **do** *l_a* **times**: *p_r*.*moveRight*() 1.2.1.4.1. if $\neg Z(F_{p_r,y,p_r,x})$ 1.2.1.4.2. **do** l_a **times**: p_r . moveLeft() 1.2.1.4.2.1. **do** *l_a* **times**: *p_r*.*moveDown*() 1.2.1.4.2.2. if $\neg Z(F_{p_r.y,p_r.x})$ 1.2.1.4.2.3. **do** l_a **times**: p_r . moveRight() 1.2.1.4.2.3.1. if $\neg Z(F_{p_r,y,p_r,x})$ 1.2.1.4.2.3.2. **do** $2l_a$ **times**: p_r . moveLeft() 1.2.1.4.2.3.2.1. if $\neg Z(F_{p_r,y,p_r,x})$ 1.2.1.4.2.3.2.2. **do** l_a **times**: p_r . moveRight() 1.2.1.4.2.3.2.2.1. do l_a times: p_r . moveUp() 1.2.1.4.2.3.2.2.2. if $p_a \neq p_1$ or $p_b \neq p_r$ 1.2.1.5. $l_a = l + 1$ //ends the l_a loop 1.2.1.5.1. if $p_r. x - p_l. x > dX. x$ and $p_r. x - p_l. x < dX. y$ 1.3. **boolean** isBalanced = $\left(\sqrt{(p_l.x - iP.x)^2 + (p_l.y - iP.y)^2} > \frac{\sqrt{(p_r.x - iP.x)^2 + (p_r.y - iP.y)^2}}{2}\right)$ and $\left(\sqrt{(p_r.x - iP.x)^2 + (p_r.y - iP.y)^2} > \frac{\sqrt{(p_r.x - iP.x)^2 + (p_r.y - iP.y)^2}}{2}\right)$ 1.3.1. $\sqrt{(p_l.x-iP.x)^2 + (p_l.y-iP.y)^2}$ **boolean** isLongEnough = $\sqrt{(p_l.x - iP.x)^2 + (p_l.y - iP.y)^2} >$ 1.3.2 dX.x and $\sqrt{(p_r.x - iP.x)^2 + (p_r.y - iP.y)^2} > dX.x$ return isBalanced and isLongEnough 1.3.3. 1.4. return false

Algorithm 2: Confirmation method for the retrosternal area recognition.

In summary, what the Algorithm 2 essentially does is, starting at the initial coordinates of p_l and p_r , it will try to displace these points only if there is an available fat pixel to move next, while respecting a set hierarchy. The l on p_l stands for "left pixel", which infers that p_l will move down and leftwards, whereas p_r will move right and downwards. The coordinates of p_l and p_r are changed only when the next moveable pixel is a fat pixel. Moreover, that displacement have to be strictly done respecting the following hierarchy: (1) at first, the point p should be moved on the horizontal direction (left or right), (2) if the associated pixel is not a fat pixel then p is moved on the vertical direction (down only) and (3) finally, if the pixel associated to the previous movement is also not a fat pixel then p is moved diagonally (left or right along with a downward movement).

That process of choosing which direction to move and verifying if the associated pixel represents fat is done at most l times for each interation of c. The l variable was introduced to overcome some interpolation problems introduced by the resizing operation. Although we have applied just the bicubic interpolation, sometimes, the resultant image contained small gaps between pair of pixels that were previously continuous. Therefore, the use of the l variable was essential. Finally, that whole process of moving the coordinates of p_l and p_r is done at most c times before the algorithm reaches convergence.

The instance in Figure 24 illustrates some of the gaps originated by the rescaling and interpolation (mainly on the external structures of the heart). That instance was originally acquired from a DICOM file in a small scale (high pixel spacing) and, therefore, had to be resized by a significant amount. Hence, that resizing left some small gaps between fat pixels despite the fact of applying a bicubic interpolation. The nearest neighbor and bilinear interpolations reinforced even more those gaps and, for that matter, were disregarded.



Figure 24: Small gaps on the image provoked by rescaling.

Finally, the boolean variable isBalanced verifies if both coordinates of p_l and p_r moved, to some extent, equally to each other. That is so to avoid situations such as the one depicted in Figure 25, where the distance from iP to p_r is much shorter than the distance from iP to p_l . Thus, the logic condition associated to the variable isBalanced, shown in Algorithm 2, verifies if the distance from iP of both points p_r and p_l . is at least half length of the other.



Figure 25: An occasion where the point p_l was displaced much more than p_r .

In a similar fashion, the boolean isLongEnough verifies if the points p_r and p_l are at least at a meaningful distance from the initial point iP in order to avoid situations depicted by the image in Figure 26, where the p_l and p_r are very close to the atlas image. The absence of this verification produced some unsuccessful recognitions. The slice featured in Figure 26, out of curiosity, can also be seen in Figure 12-(e) without any atlas blended to it.



Figure 26: Wrong recognition due to the absence of the isLongEnough variable.

Moreover, if the confirmation function in Algorithm 2 returns false then the recognition step is rerun whereas this time, at each evaluated position, the confirmation method is also run jointly. Thus, the chosen location will now be the one that minimizes the hybrid mean difference (Section 3.1.1.5) and that also returns true on the confirmation function. Therefore, there must exist a best position among every position μ that satisfies the call to the function $confirmPosition(F, \mu, w_A, h_A, c, l, dX)$ in Algorithm 4, where w_A and h_A stand for the width and height of the area A. The remaining parameters c, l and dX are just reassigned to the ones of the confirmationMethod function.

In summary, given a fixed image *F* (fat image) and the moving image *M*(atlas), the function *findRetrosternalArea*(*F*, *M*) in Algorithm 3 is run. At the end of this run the best position is retrieved and stored on μ . The function *confirmPosition*(*F*, μ , *w*, *h*, *c*, *l*, *dX*) in Algorithm 4 is then called, if true is returned, the position μ is reinforced as being accurate and the algorithm converges. Otherwise, the function *findRetrosternalArea*(*F*, *M*) is rerun but, this time, only evaluating positions *p* that respect the condition of returning true on the *confirmPosition*(*F*, *p*, *w*, *h*, *c*, *l*, *dX*) function. The set of empirically chosen parameters for these methods were: w = 0.13F.width, h = 0.04F.height, c = 0.6F.width, l = 0.003(F.width + F.height) and dX = (0.2F.width, 0.55F.width). The vector (e_x, e_y)

consists of the amount of extrapolation on the moving image position regarding the fixed image.

1. findRetrosternalArea(**Image** F, **Image** M) **double** bestScore = $+\infty$ 1.1. 1.2. **vector2** bestPosition = (0,0)**vector2** $(e_x, e_y) = (80, 10)$ 1.3. for $i = -e_y$; i < F. height $+ e_y$; i + +1.4. for $j = -e_x$; $j < F.width + e_x$; j + +1.4.1. (M.x, M.y) = (j, i)1.4.1.1. **double** hybridScore = hybridMeanDifference(F, M, 3) 1.4.1.2. **if** hybridScore < bestScore 1.4.1.3. bestScore = hybridScore 1.4.1.3.1. bestPosition = (j, i)1.4.1.3.2. return bestPosition 1.5.

Algorithm 3: Steps to find the retrosternal area.

```
1 confirmPosition(Image F, Vector2 \mu, Int w, Int h, Int c, Int l, Vector2 dX)

1.1. for every pair of points p_1 and p_2 within F

1.1.1. if (p_1. y \text{ and } p_2. y) \in \left[-\frac{h}{2} + \mu. y, \frac{h}{2} + \mu. y\right] and ((p_1. x \text{ and } p_2. x) \in \left[-\frac{w}{2} + \mu. x, \frac{w}{2} + \mu. x\right])

1.1.1. if confirmationMethod(F, p_1, p_2, dX, c, l)

1.1.1.1. return true

1.2. return false
```

Algorithm 4: Formally applying the confirmationMethod function.

3.1.3.1 COMPLEXITY

The complexity of the *findRetrosternalArea* function is considered equal to O(wh), as previously discussed in Section 3.1.2, where w and h is the width and the height of the fixed image, respectively. This function evaluates every point p_e of the fixed image for the placement of the atlas, and this is the direct reason for such complexity. However, the confirmation method requires an additional access of points. For each evaluated position p_e , we have to loop through every position of the rectangle area A. As such, in order to access every point within A, which is illustrated by the yellow rectangle depicted in Figure 22, a complexity of (0.12w0.04h) is required. As the constants and coefficients are not meaningful in a complexity measure, the worst case complexity for accessing every point in A will result

on O(wh). Nevertheless, we need to evaluate every combination of points within the area A. Thus, instead of just O(wh), the worst case complexity for this operation is $O(w^2h^2)$.

In addition, the confirmation method in Algorithm 2 has a loop based on the c variable, which, in turn, grows proportionally to the width of the fixed image. Hence, the complexity of that algorithm equals O(w). For every pair of points p_1 and p_2 selected by the Algorithm 4 (worst case complexity: $O(w^2h^2)$), the function *confirmationMethod*(F, p_1 , p_1 , dX, c, l), which is O(w), will be called. That fact results on a complexity of $O(w^3h^2)$ for the function confirmPosition, that is, $O(w)O(w^2h^2)$. By combining the complexity of the *confirmPosition* function, which is $O(w^3h^2)$, with the complexity of the *findRetrosternalArea* function, which is O(wh), we then have the worst-case complexity for the proposed recognition of the retrosternal area being equal to $O(w^4h^3)$. This is a relatively high computational complexity for real time processing. Nonetheless, the findRetrosternalArea function usually successfully recognizes the retrosternal area alone and does not need to be rerun along with the confirmation method. In fact, 92% of the 52 patients we have assessed had their retrosternal area successfully recognized without the confirmation method, as also shown in Section 3.1.1. Hence, the worst case complexity for the proposed recognition method is $O(w^4h^3)$. However, the complexity w^4h^3 occurs at just 8% of the time. At the remaining 92%, the complexity of the complete approach of finding the retrosternal area is simply $\Omega(wh)$.

3.1.3.2 SUCCESSFUL CORRECTIONS

Finally, the Figure 27 shows all the four instances where the *findRetrosternalArea* method unsuccessfully recognized the retrosternal area, whereas the Figure 28 shows the same images from Figure 27 after the appliance of the confirmation method. Therefrom, we have tested this recognition method with all the 52 previously selected images plus a set of 30 patients and the recognition approach (using also the confirmation method) did not fail to recognize the retrosternal area in none of them.



Figure 27: The 4 instances, out of 52, where the recognition method failed.





Figure 28: The same 4 instances after the confirmation method.

3.1.4 APPLYING THE TRANSFORMATION

Until now, we have: (1) built a simple atlas for the retrosternal area; (2) used this atlas along with a proposed similarity measure to recognize and select the retrosternal area location on an input image and (3) developed a heuristical confirmation method to reinforce the previously selected location and to correct eventual mistakes. However, we have not yet applied the actual translation to the images of the various patients.

We assume that each patient usually has from 30 to 70 slices in their cardiac CT scans, such as the ones regarded in this work. In summary, the whole registration method is comprised of the steps following described: at first, we select the $\delta(n)^{th}$ slice from the beginning of the stack (relatively next to the shoulders) according to the Equation (8), where the total number of slices of the CT scan is n. After selecting the $\delta(n)^{th}$ slice, we apply the transformation shown in Equation (9), where v_o stands for the original pixel spacing value while v_d stands for the desired standard pixel spacing value, which is equal to 0.35 in both directions. More details about pixel spacing and scaling are described in the Appendix A and B.

$$\delta(n) = \begin{cases} n-45, & \text{if } n > 45\\ 0, & \text{otherwise} \end{cases}$$
(8)

$$\begin{bmatrix} x'\\ y' \end{bmatrix} = \begin{bmatrix} \frac{v_o \cdot x}{v_d \cdot x} & 0\\ 0 & \frac{v_o \cdot y}{v_d \cdot y} \end{bmatrix} \begin{bmatrix} x\\ y \end{bmatrix}$$
(9)

T(x', y')The transformed image is then passed the to findRetrosternalArea function shown in Algorithm 3. After positioning the atlas on the retrosternal area, the confirmPosition function in Algorithm 4 is run and, if necessary, the *findRetrosternalArea* function is rerun along with the confirmation method. Finally, after the last placement of the atlas on the position ω , the image T is translated once again with the objective of aligning the found position ω of the retrosternal area to a standard centralized position φ . That standard position φ has been empirically chosen to be (0.14w, 0.23h). Therefore, the previously scaled image T is then translated by the Equation (10), originating $T_2(x'', y'')$. That step concludes the registration method and gives us the parameters of the transformation to be applied to every slice of the processed patient.

$$\begin{bmatrix} x''\\ y'' \end{bmatrix} = \begin{bmatrix} x'\\ y' \end{bmatrix} - \begin{bmatrix} \omega. x - \varphi. x\\ \omega. y - \varphi. y \end{bmatrix}$$
(10)

An overview of the proposed registration is shown in Figure 29. Visual results of our registration were shown in Figure 9 and Figure 10.



Figure 29: An overview of the whole registration procedure.

3.2 PROPOSED SEGMENTATION

In order to generate a concise predictive model we need to provide reliable data for the training step of the classification algorithm. Therefore, two specialists, one being a physician and the other being a computer scientist, have manually segmented the epicardial and mediastinal adipose tissues of 20 patients or CT scans (10 male and 10 female). These data compose a ground truth that contains 878 manually segmented cardiac CT images and is available at [114]. The demographics of these patients are shown in Figure 30.

-Population-	
Topulation	
Total of patients: 20 men: 10 women: 10	
Mean ages: $55.4 (\pm 22)$ median: 53	
Location: Rio de Janeiro, Brazil	
Mean epicardial fat volume (ml): 97.9	
min: 39.2 max: 203.4 median: 94.8	
Mean mediastinal fat volume (ml): 103.6	
min: 35.6 max: 226.2 median: 92.2	
Technical details	
Total of slices: 878	
Per patient mean: 43.9 median: 42	
Total of manufacturers: 2	
Siemens: 11 Phillips: 9	

Figure 30: Patient demographics.

It is important to highlight that, previously to the manual segmentation, the images were already registered by our proposed registration. The black value (0) of the segmented images represents the background. The produced ground truth conforms to the following standard: the epicardial fat is represented in red, the mediastinal fat is represented in green and the pericardium or a transitional area between the epicardial and mediastinal fats is depicted as blue. The grey textural information of the pixels was simply shifted to their respective color layer. The image in Figure 31 illustrates one of the manually segmented slices.



Figure 31: A slice of one patient that compose the ground truth.

Thereafter, some features should be extracted and algorithms should be chosen to train on the extracted information once the ground truth is established. The proposed classification approach consists of three main steps: (1) extracting the features, (2) training the predictive model and (3) further classifying an incoming CT scan. The steps (1) and (2) do not need to be redone every time a new incoming scan needs to be classified. In fact, if that were true, the method would take so long to converge that it would be unpractical. It do not take too much to perceive that the generated dataset is very big and that the step (1) is a very slow process. Thus, the step (3) is independent of (1) and (2) once they have already been processed. These three steps are represented by the "Initial Points" in Figure 32.



Figure 32: Overall steps required for the classified segmentation.

The first step consists of iterating every pixel of the images that compose the ground truth and extracting features related to the iterated pixel. The features vector of each pixel will represent a line on the extracted database and, therefore, various pixels of various images will compose the final dataset that is used to train predictive models for further segmentation.
There is a huge variety of features that can be extracted from images; some of them are specifically valuable for each type of image and arrangement of data. On the following section, we define and select some features that could be extracted for our problem. Moreover, we decided not to normalize the features after the extraction. The reason is so that without normalization, we are able to generate various predictive models from different training sets and aggregate these predictive models with little effort.

The second and third steps, i.e., the steps to be applied after extracting the features, are totally related to the classification algorithms. On this aspect, the incoming pixel can be assigned to 3 different classes (mediastinal fat, epicardial fat or pericardium). It should be remembered that black pixels (0) are set as background and are not classified by the predictive model. Moreover, some classification algorithms work only with binary classes. Therefore, we divided the 3 possible classes into a binary mapping for each class. The classes are disposed in three columns on the dataset; each one represents a class and receives true or false as value. When training the algorithm for the epicardial fat, the columns that represent the remaining classes should be removed prior to the training in order to avoid predictive models trained using these classes, which are the desired output. The Figure 33 illustrates the overall steps for the segmentation of an incoming DICOM file.



Figure 33: Segmentation of an arbitrary DICOM file.

3.2.1 USED FEATURES

The features we have selected for extraction are divided in three main categories: (1) the primary features, which are directly related to the pixel information, (2) the secondary,

which are related to the image or to a neighborhood window and (3) the tertiary, which are related to data that was already derived from the image or from a neighborhood window.

Thus, the possible and also the ones that were extracted as primary features are the pixel value along with its x, y and z coordinates. For secondary features we extracted the x and y coordinates of the pixel with respect to the center of gravity of the image $(x - x_g, y - y_g)$, where (x_g, y_g) is shown in Equation (18) and further addressed in Section 3.2.1.2. Besides, a $\tilde{i}x\tilde{j}$ neighborhood of pixels around the iterated pixel $P_{y,x}$ was considered for extracting information, where $\tilde{i} = \tilde{j} = 2q + 1 | q > 0$ as shown in Figure 34. If any pixel of this neighborhood extrapolates the boundaries of the image then it is treated again as a black pixel.

$P_{y-\left \frac{\tilde{l}}{2}\right ,x-\left \frac{\tilde{j}}{2}\right }$		$P_{y-\left \frac{\tilde{l}}{2}\right ,x}$		$P_{y-\left \frac{\tilde{\iota}}{2}\right ,x+\left \frac{\tilde{J}}{2}\right }$
$P_{y,x-\left\lfloor \frac{\tilde{j}}{2} \right\rfloor}$		$P_{y,x}$		$P_{y,x+\left[\frac{\tilde{j}}{2}\right]}$
	:		:	
$P_{y+\left\lfloor\frac{\tilde{l}}{2}\right\rfloor,x-\left\lfloor\frac{\tilde{j}}{2}\right\rfloor}$		$P_{y+\left[\frac{\tilde{l}}{2}\right],x}$		$P_{y+\left\lfloor\frac{\tilde{i}}{2}\right\rfloor,x+\left\lfloor\frac{\tilde{j}}{2}\right\rfloor}$

Figure 34: Neighborhood around the pixel $P_{v,x}$.

From this neighborhood, the following secondary features were extracted: (1) a simple arithmetic mean of the grey values, (2) the geometric moments M(0,1), M(1,0) and M(1,1) and (3) the proposed Coefficient of Smooth Variation (CSV) of the grey values. The arithmetic mean is obtained from the Equation (11) and the coefficient of smooth variation is addressed in the following section.

$$AM = \frac{1}{\tilde{\iota}\tilde{j}}\sum_{i}^{\tilde{\iota}}\sum_{j}^{\tilde{j}}P_{i,j}$$
(11)

Furthermore, the (1) moments $M_{1..4}$ at the distances (0,1), (1,0) and (1,1) of the cooccurrence matrix were extracted as tertiary features. The reason for selecting those distances is so that there are images where the layer comprising the heart is of approximately one or two pixels of thickness. Moreover, based on the run length matrix, the (2) run percentage and (3) grey level non-uniformity were also extracted as tertiary features respecting the directions $\theta = \{0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}\}$. All these tertiary features were obtained from the neighborhood window and are addressed on the following sections. The extracted features are summarized in the Table 3 along with the applied parameters. By accounting the number of applied parameters with relation to each feature, a total of 31 features were extracted from the images. The texture is considered an important regional descriptor for segmentation and classification of various types of medical images. Thus, in this case, the reason for extracting texture-based features (most of the secondary and tertiary features) was mainly due to the hypothesis that the epicardial and mediastinal fat yield a slight difference on their texture that can be partially accounted by the analytical process of these features.

Туре	Feature	Parameters
Primary	Grey value	-
	x	-
	У	-
	$m{z}$ (slice number)	-
Secondary	x relative to the center of gravity	-
	$m{y}$ relative to the center of gravity	-
	Arithmetic Mean	-
	Coefficient of Smooth Variation	-
	Geometric Moments	M(0,1), M(1,0), M(1,1)
Tertiary	Moments of the Co-occurrence Matrix	$(\Delta x, \Delta y)$ = {(0,1), (1,0), (1,1)} combined to { M_1, M_2, M_3, M_4 }
	Run Percentage	$\theta = \{0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}\}$
	Grey level non-uniformity	$\theta = \{0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}\}$

Table 3: Extracted features and used parameters.

3.2.1.1 COEFFICIENT OF SMOOTH VARIATION

In this work, a new feature based on the Gaussian 1D filter and in the sup metric is proposed, which is denominated Coefficient of Smooth Variation. In this coefficient, a convolution operation is performed using a Kernel of which weights are based on a unidimensional Gaussian filter computed in terms of the distance r from the point (x_i, y_i) of

the Kernel to its center (x_c, y_c) . However, the sup metric is used as the distance in \mathbb{R}^2 . That is, instead of using the traditional Euclidian metric shown in the Equation (12) [115]:

$$d(x_c, y_c; x_j, y_i) = \sqrt{(x_j - x_c)^2 + (y_i - y_c)^2}$$

$$d_2 = (|x_j - x_c|^2 + |y_i - y_c|^2)^{\frac{1}{2}}$$
(12)

The metric d_{∞} is used, where in the Equation (12), instead of 2, ever increasing numbers are considered as shown in Equation (13):

$$d_n = (|x_j - x_c|^n + |y_i - y_c|^n)^{\frac{1}{n}}$$
(13)

In a way that *n* tends to infinity, resulting in the metric known as sup, which is actually part of the class of metrics d_n , where n = 1,2,3,4,5, ... 100, ... 1000, ... Thus, for ever increasing high values of *n*, one of the parcels within the modulus would become so greater than the other in a way that it finally becomes the only relevant parcel, which results on the known sup metric shown in Equation (14) [115]:

$$d_{\infty} = max(|x_{j} - x_{c}|, |y_{i} - y_{c}|)$$
(14)

The Figure 35 represents the Euclidian distance of each pixel with relation to the central one in a 7x7 window. For this case, there is a total of 9 groups of distances, excluding the central one.

$\sqrt{18}$	$\sqrt{13}$	$\sqrt{10}$	3	$\sqrt{10}$	$\sqrt{13}$	$\sqrt{18}$
$\sqrt{13}$	$\sqrt{8}$	$\sqrt{5}$	2	$\sqrt{5}$	$\sqrt{8}$	$\sqrt{13}$
$\sqrt{10}$	$\sqrt{5}$	$\sqrt{2}$	1	$\sqrt{2}$	$\sqrt{5}$	$\sqrt{10}$
3	2	1	0	1	2	3
$\sqrt{10}$	$\sqrt{5}$	$\sqrt{2}$	1	$\sqrt{2}$	$\sqrt{5}$	$\sqrt{10}$
$\sqrt{13}$	$\sqrt{8}$	$\sqrt{5}$	2	$\sqrt{5}$	$\sqrt{8}$	$\sqrt{13}$
$\sqrt{18}$	$\sqrt{13}$	$\sqrt{10}$	3	$\sqrt{10}$	$\sqrt{13}$	$\sqrt{18}$

Figure 35: Euclidian distance Kernel in a 7x7 window.

The first step to determine the Kernel of the proposed coefficient is to compute the distance of the pixels using the sup metric. In this case, the highest distance between the vertical and horizontal directions is what defines the distances of the pixels. Thus, the distance increase in a rectangular fashion before the modulus operation as shown in Figure 36.



Figure 36: Sup metric.

The distance Kernel based on the sup metric is shown in Figure 37:

3	3	3	3	3	3	3
3	2	2	2	2	2	3
3	2	1	1	1	2	3
3	2	1	0	1	2	3
3	2	1	1	1	2	3
3	2	2	2	2	2	3
3	3	3	3	3	3	3

Figure 37: d_{∞} distance in a 7x7 Kernel.

Instead of 9 groups of distances given by the Euclidian distance, in the case of the proposed coefficient there are only 3. Assuming again that the Kernel size is $n \times n$, then the number of groups of Euclidian distances is given by the Equation (15):

$$\left[\frac{n^2}{8}\right] + \frac{n-3}{2} + 1 \tag{15}$$

As opposed to $\lfloor n/2 \rfloor$ of the proposed coefficient. The possible types of distances to be computed are significantly less. In a real-time scenario, that difference greatly influences on the performance. Therefrom, regarding a 7x7 window, instead of using the traditional Gaussian Kernel shown in Figure 38:



1	6	15	20	15	6	1
6	36	90	120	90	36	6
15	90	225	300	225	90	15
20	120	300	400	300	120	20
15	90	225	300	225	90	15
6	36	90	120	90	36	6
1	6	15	20	15	6	1

Figure 38: A Gaussian Kernel in a 7x7 window.

We apply the coefficient of smooth variation, which looks like the Kernel in the Figure

56	56	56	56	56	56	56
56	215	215	215	215	215	56
56	215	3162	3162	3162	215	56
56	215	3162	107	3162	215	56
56	215	3162	3162	3162	215	56
56	215	215	215	215	215	56
56	56	56	56	56	56	56

Figure 39: Smooth variation Kernel in a 7x7 window.

When the entire neighborhood window is regarded, the convolution is given by the Equation (16). The β is a constant and should be adjusted to avoid overflows. In our approach, we have considered β to be 10^7 . The weight $d^{\infty+1}\sqrt{\beta}$ is multiplied by the pixel value $P_{i,j}$ of the window. The *CSV* is itself one of the features extracted from the neighborhood window.

$$CSV = \sum_{i}^{\tilde{\iota}} \sum_{j}^{\tilde{J}} \sqrt[d_{\infty}+1]{\beta} P_{i,j}$$
(16)

If the CSV was outputted to an image, the result of the convolution on a single pixel with a 5x5 Kernel would be the one shown in Figure 40.



Figure 40: CSV convolution.

3.2.1.2 GEOMETRIC MOMENTS

Geometric moments provide important statistical descriptors of an image and are applicable for texture analysis [88,116]. The geometric moment M(m, n) of order m + n is defined by Equation (17) where $P_{i,j}$ are pixel values, j and i are pixel coordinates and m, n are integer exponents that define the moment order. The center of gravity (x_g, y_g) can be obtained by the Equation (18) [117].

$$M(m,n) = \sum_{i} \sum_{j} j^{m} i^{n} P_{i,j}$$
(17)

$$(x_g, y_g) = \left(\frac{M(1,0)}{M(0,0)}, \frac{M(0,1)}{M(0,0)}\right)$$
(18)

3.2.1.3 CO-OCCURRENCE MATRIX

The co-occurrence matrix associates the number of co-occurrences of a grey level of a pixel p_a to a grey level p_b on an image H at a given distance $(\Delta x, \Delta y)$, where $p_n \in H$ and the maximal n is the amount of distinct grey values that exist in H. The function $C_{\Delta x,\Delta y}(p_a, p_b)$ in Equation (19) denotes the co-occurrences between the pixel values p_a and p_b at a distance $(\Delta x, \Delta y)$ [88]. An example of the construction of the matrix is shown in Figure 41.

$$C_{\Delta x,\Delta y}(p_a, p_b) = \sum_{i} \sum_{j} \begin{cases} 1, & if \ H(i,j) = p_a \ and \ H(j + \Delta x, i + \Delta y) = p_b \\ 0, & otherwise \end{cases}$$
(19)



Figure 41: Construction of a co-occurrence matrix.

The sequence $\{1,2\}$ on the direction (1,0) appears 4 times in the input image. Therefore, the cell (1,2) of the co-occurrence matrix must be populated with the number of co-occurrences of the sequence $\{1,2\}$, which is 4. This procedure is repeated for each possible co-occurrence on the input image, respecting the given distance.

The probability of a grey level p_a co-occurring with p_b at a certain distance is given by the times these two grey values co-occur divided by all the co-occurrences of every pair of grey values (k, l) at the same distance. Hence, the probability of p_a co-occurring with p_b is given by the function $P_{\Delta x, \Delta y}(p_a, p_b)$ on the Equation (20) below. The moment of the cooccurrence matrix is given by the function M_g in Equation (21), where g is the moment's degree [88].

$$P_{\Delta x, \Delta y}(p_a, p_b) = \frac{C_{\Delta x, \Delta y}(p_a, p_b)}{\sum_k \sum_l C_{\Delta x, \Delta y}(k, l)}$$
(20)

$$M_g = \sum_k \sum_l P_{\Delta x, \Delta y}(k, l)(k-l)^g$$
(21)

3.2.1.4 RUN LENGTH MATRIX AND DERIVATIVES

The run percentage and grey level non-uniformity are features calculated over the run length matrix of an image. The run length matrix, in turn, is assembled based on the run length encoding principle. Moreover, the run length encoding is a simple technique used to encode and compress information [118]. On a run length encoded message, the identical symbols that previously appeared continuously n times are traded for a unique representation of the symbol plus a concatenation of the number n.

The run length matrix follows a similar principle in relation to the encoding. Given a direction θ , the number of identical grey values of an image in that same direction can be accounted to compose a matrix. Thus, one axis of this matrix will represent all the possible grey values comprised on the image and the other axis will represent the length of the run, i.e., all the possible times the grey values appear continuously on the direction θ . The four conventional possible directions for θ are 0°, 45°, 90° and 135°. The 0°, for instance, represent the horizontal direction. When computing the matrix for the 0° direction all the lines of the image are accessed and when any grey level p is continuously repeated l times, the value of run length matrix on the position (p, l) is increased by one.

The run length matrix construction can be recursively programmed. For instance, a function responsible for jumping one pixel where the length of the run is passed as parameter may be created and, for every first pixel of a line or diagonal on the image it could be triggered. The called function evaluates if the next pixel value is equal to the current and keeps calling itself and incrementing the length variable on each call. The stopping condition is when the next value is not equal to the current value or if it is the end of the image. At that stopping condition the function increments the run length matrix at the according position. Finally, at the end of the triggering, the matrix will be correctly populated. The Figure 42 illustrates how a run length matrix on the direction 0° is constructed.



Figure 42: Construction of a run length matrix.

This matrix is, in principle, very different from the co-occurrence matrix. For instance, in the input image, the pixel value 0 occurred consecutively four times on the 0° direction, which means that the cell (0,4) of the run length matrix should be populated with the occurrence 1. As for the pixel value 1, there were two occurrences where it appeared continuously 2 times. Therefore, the run length matrix at the position (1,2) should be populated with the number of occurrences 2.

We will assume that the run length matrix for an arbitrary image is R_{θ} , where $R_{\theta}(p, l)$ represents the times that the grey value p appears continuously l times on the image respecting the direction θ . Thus, the grey level non-uniformity feature is computed by solving the Equation (22). The summations are done for every possible p and l of the matrix [88].

$$G_{\theta} = \frac{\sum_{p} (\sum_{l} R_{\theta}(p, l))^{2}}{\sum_{p} \sum_{l} R_{\theta}(p, l)}$$
(22)

Furthermore, the run percentage RP_{θ} is of easy computation. It consists of summing all the elements of the run length matrix and dividing it by the area *S* of the image, as shown in Equation (23). The areas of the neighborhood window are constant and, therefore, that division is irrelevant in this case. Nevertheless, to maintain the usual formulation we kept the division [88]. The sum of the run length matrix appears to be a relevant feature since it could show how uniform is an arbitrary part of the image.

$$RP_{\theta} = \frac{\sum_{p} \sum_{l} R_{\theta}(p, l)}{S}$$
(23)

3.2.2 EVALUATION OF THE CONVERGENCE TIME

For the classification tasks we have used the Weka library [119]. Weka is an opensource collection of machine learning algorithms maintained by the University of Waikato and is entirely programmed in Java. The Weka usage is twofold, it has its own graphical interface that can be used on several types of graphical analysis and the library can be directly imported and used in Java code as well.

Before proceeding to the classification tasks, there is a need to define the meaning of accuracy. In machine learning, accuracy is defined as the sum of the true positive and true negative occurrences divided by the total population, as shown in Equation (24). With respect to the proposed solution of this work, the total of distinct classes are three. From here on, when only one accuracy rate is provided, we define it as being the arithmetic mean of the three classes (red, green and blue).

Furthermore, the true positive (TP) rate stands for the percentage of pixels that were correctly classified as true. The true negative (TN), in turn, stands for the percentage of pixels that were correctly classified as false. The false positive and false negative rates stand for the two types of errors that can occur on a classification problem. The Type I error is the false positive (FP), which occurs when the algorithm classifies an instance as true and the correct is false. Thus, the Type II error is called false negative (FN) and occurs when the algorithm classifies as false when the label should be true [120].

$$Accuracy = \frac{\sum True \ positive + true \ negative}{\sum Total \ population}$$
(24)

In order to reduce the amount of classification algorithms to be assessed, we have extracted the features of a single patient (which has approximately 50 images and, in average, a total of 512×512×50 features vectors are extracted) and evaluated the time that each algorithm took to construct the predictive model based on two thirds of the patient data plus the time it took to evaluate the model on the remaining one third (66% split) in a simple personal computer. It is important to reduce the computational time spent on the training, not only to shrink the whole analysis but also to speed up the final classifying process. In fact, there is a need to evaluate distinct algorithms in several aspects and a fast convergence is strictly necessary.

Therefore, for that single patient, we have tested all the classification algorithms present in Weka on its version 3-6-11. Some of these algorithms are, namely, the Support Vector Machine (SVM), Sequential Minimal Optimization (SMO), Naïve Bayes, Radial Basis Function Network (RBFNetwork), Random Trees, C4.5 (or J48), Primal Estimated Sub-Gradient Solver for SVM (SPegasos), REPTree, iBk (k-NN), Multilayer Perceptron and others. Among all the tested algorithms, we have selected for further analysis the ones that converged within 200 seconds. The parameters of each algorithm were based on their standards with some adjustments and the best result was selected. On this evaluation, the size of the neighborhood window was 5x5 (relatively small in order to speed up the processing). The comparison of the achieved results is shown in Table 5 below.

Algorithm	Accuracy	Time (sec)	Acc/Time
J48Graft	99.0%	132.86	0.75
RandomForest	98.9%	112.57	0.88
REPTree	98.9%	10.34	9.56
J48	98.9%	151.23	0.65
SimpleCart	98.9%	108.78	0.91
SMO	98.3%	58.66	1.68
RandomTree	97.5%	8.0	12.19
RBFNetwork	96.8%	3.48	27.82
SPegasos	96.8%	15.77	6.14
DecisionStump	96.8%	52.34	1.85
HyperPipes	94.8%	0.04	2370.0
NaiveBayes	86.0%	55.48	1.55

Table 5: Accuracies and convergence time on a single patient.

At first, the REPTree algorithm appears to be the best choice since it achieved a great accuracy in a relatively fast convergence time if compared to the other algorithms. Besides, the HyperPipes result was rather interesting. In this case, it returned a very good accuracy "almost instantaneously". However, when applied to a bigger dataset and a set of images of distinct patients, the results obtained through the HyperPipes algorithm start to differ drastically.

3.2.3 EVALUATION OF NEIGHBORDHOOD SIZES

In order to avoid unfair comparisons of classifiers we have tested some of them along with a variation on the neighborhood size. Hypothetically, one may consider that some classifiers perform better on a neighborhood of certain size. Some charts that represent the accuracy of each classifier on the y-axis and the variation of the neighborhood size in pixels on the x-axis are displayed in Figure 43. The accuracies are divided per class in three colors, each one respecting the ground truth definition. The red color represents the epicardial fat, green represents the mediastinal and blue the pericardium or a transitional space between the epicardial and mediastinal fats.

The accuracies on the charts were also achieved using the 66% random-selected split method as test mode from the data of 20 patients (878 images). However, some slices were skipped from the processing to decrease the convergence time. The reason for choosing the split method was also to reduce the huge convergence time due to the great amount of data. The extracted features composed datasets of approximately 1.5 gigabytes for each neighborhood size that were provided to the classifier. The period to train and evaluate the model lasted, in some cases, up to 20 minutes for each combination of neighborhood size and classification algorithm.











Figure 43: Accuracies (y-axis) of classifiers versus neighborhood sizes (x-axis).

The five algorithms shown in Figure 43 were the ones who performed faster on this large dataset. The HyperPipes was the fastest and always converged virtually within 0.5 seconds but, in this case, the best accuracy that it could achieve was around 70%. We can state by this evidence that this algorithm is not generalizable. In other words, it fails to generate a "universally applicable" predictive model from the moment that more than one patient is regarded for the training. However, it is a rather simple algorithm and, in fact, the achieved low accuracy was somehow supposed to happen.

Due to the convergence time issue, we were not able to extensively assess all the possible sizes for the neighborhood for all the algorithms selected on the convergence time test. The REPTree algorithm did not converge remarkably faster than the remaining decision tree algorithms on this large dataset, even though fast convergence is probably the main characteristic of the REPTree algorithm. The RBFNetwork was faster than RandomForest and the SPegasos was slower but both returned lower accuracies.

The RandomTree and DecisionStump were just a little faster than RandomForest and REPTree but returned significant lower accuracies and, therefore, were disregarded due to the massive presence of decision tree algorithms on the previous convergence time experiment. The J48Graft returned similar accuracies if compared to the RandomForest but its convergence was approximately 1.4 times slower and, due to that matter, it was impracticable to evaluate all the sizes for this classifier. All the other algorithms shown in Table 5 that were not quoted here took more time to converge on a large dataset than the SPegasos and, therefore, they could not be precisely evaluated.

Furthermore, the accuracy of almost every algorithm started to slightly decrease after the size of 25x25 pixels and to sharply decrease after 39x39. In fact, the more the neighborhood size converges to the size of the image, more information is lost. This behavior was expected, however, that analysis was done in order to set an arguably perfect size for the neighborhood window, for posteriorly refining the results. Moreover, the higher the size of the neighborhood window, the higher is the time the algorithm takes to extract the features from the image and, consequently, to converge. Therefore, by evaluating this tradeoff we decided that 25x25 was the most suited size for the neighborhood window.

Finally, we also conclude that the decision tree algorithms were the ones that best performed over a large dataset. The Table 4 compares mean values produced by the algorithms evaluated on this benchmark over all the classes and evaluated sizes of the neighborhood shown on the previous charts. Furthermore, in this comparison, we have included the SimpleKMeans algorithm with the objective of relating its performance. That was so because the issue addressed by this work appears to be, at first, more suitable to clustering than to classification.

Algorithm	Accuracy	Time (sec)
RandomForest	96.72%	548.67
J48Graft	95.11%	763.90
REPTree	94.02%	544.63
SPegasos	83,5%	826.45
RBFNetwork	75,32%	387.92
SimpleKMeans	60.20%	2000.32
HyperPipes	57.37%	0.55

Table 4: Accuracies and convergence on a large dataset.

3.2.4 OVERFITTING ANALYSIS

Overfitting is usually defined as a lack of generalization or a super adjustment to the training dataset induced by some machine learning algorithms during the training pace. For instance, an algorithm that achieves high accuracies evaluating its predictive model on the training set but fails to get high accuracies on similar datasets can be considered overfitted to the training data [121].

To complement our work we modestly analyzed the overfitting degree of the RandomForest, REPTree, J48, J48Graft, SMO and SPegasos algorithms. Our overfitting analysis consisted of training classifiers with data of a single patient ($50 \times 512 \times 512$ features vectors) and using the generated predictive model to segment another patient scan of a different manufacturer. The algorithm who visually performed better, i.e., the one that segmented the images more similarly to the ground truth was defined as having the lowest overfitting. The parameters for the features were the same for all the algorithms (5x5 pixels for the size of the neighborhood window) and the parameters of the classifier were empirically chosen. Results of a single instance segmented by the classification algorithms are shown in Figure 44. For this trial, we have disregarded the blue layer (i.e., the pericardium or the unknown area).



RandomForest



REPTree



SMO and SPegasos

RBFNetwork

Figure 44: One instance segmented by algorithms trained on another patient.



Figure 45: Ground truth.

We concluded that the RandomForest was the algorithm that overfitted the least. In fact, the segmented result is the most similar if compared to the ground truth in Figure 45. The REPTree, J48 and J48Graft algorithms performed a less sparse segmentation but it is clear that the construction of their predictive model was heavily induced by the features derived from the spatial disposition of the pixels (x and y coordinates) and not much on the texture. That fact is evidenced by the lines and columns on the segmentation produced by these algorithms.

The SMO and SPegasos, both algorithms based on the SVM, outputted the segmented image entirely in yellow. The yellow color is a combination of the colors red and green. Therefore, what the algorithm essentially did was classifying all the bright pixels as epicardial and mediastinal fat. Thus, we can infer that the SMO and SPegasos algorithms lack heavily on generalization, at least for this evaluation. They took an easy path assuming that every pixel is both epicardial and mediastinal fats. This may be associated to the weakness of the classifier in a sense that, if the generated model does not predict well, then it is better to just assume that everything is true or everything is false to raise the accuracy by some margin.

The predictive model generated by the RBFNetwork was similar to the SMO and SPegasos to some degree. With regard to the mediastinal fat (green), the algorithm classified every pixel as false. On the other hand, regarding the epicardial fat (red), it classified almost everything but the outline as true. The result of the RBFNetwork segmentation is a little better than assuming that every pixel belongs to both classes, but still being a bad result.

Among the decision tree algorithms, we can say that the RandomForest was the one that least overfitted, directly followed by the J48Graft algorithm. The J48Graft converges slower than the RandomForest and produces similar accuracies but a very different type of segmentation (less sparse). The J48Graft converged faster than the J48 algorithm and produced better segmentations (and accuracies). Thus, the J48 was probably the one that most suffered overfitting between these four.

The decision tree algorithms performed better in this trial. The RandomForest, along with the J48Graft, were selected as the ones that overfitted the least, according to their accuracies and the actual segmentation shown in Figure 44. We tend to think that RandomForest is the best choice due to the fact that on the segmentation produced by this

algorithm, there is still a clear possibility for improvement on a post-processing step. Some techniques could be applied so that the sparse characteristic of the segmentation could be reduced and, consequently, the overall result may improve.

3.2.5 FEATURE ANALYSIS

Until now we have evaluated various classifiers using all the 31 previously extracted features. However, although all the extracted features are valuable for the distinct types of fat, when it comes to evaluating the parameters of the features, there must be one set of parameters that will be better than the others. Therefore, that was the main reason for selecting a couple of different parameters for the features based on the co-occurrence and run length matrixes. Thus, we considered the premise that, after an extended evaluation of these features, some of them could be disregarded with no significant impact in the outcome of the classification.

An incipient ranked evaluation of these features using the Linear Forward Selection as search method, the Attribute Subset Evaluator along with the REPTree algorithm and the same dataset used in Section 3.2.3 confirmed that the features have very distinct degrees of importance for each class involved in our problem. The Table 5 compares the position of the feature in the ranking (the lower the better) for the three possible classes. If the feature is not ranked with numbers on the Table 5 it means that the feature was not considered sufficiently relevant by the algorithm. We intend to do in the near future a more extensive analysis using distinct search methods and the RandomForest algorithm.

Туре	Feature	Epicardial	Mediastinal	Pericardium
Primary	Grey value	1	-	-
	x	2	1	-
	У	3	2	-
	z (slice number)	-	-	19
Secondary	x relative to the center of gravity	4	3	1
	y relative to the center of gravity	5	4	2
	Arithmetic Mean	6	-	4
	Coefficient of Smooth Variation	-	5	3
	Geometric Moments	14-15	11	15-18
Tertiary	Moments of the Co-occurrence Matrix	7-10	6-7	5-11
	Run Percentage	13	8-10	-
	Grey level non-uniformity	11-12	-	12-14

Table 5: Ranked features per class.

For the epicardial fat, the parameters of the moments of the co-occurrence matrix that went better than the others were at the distance (0,1) and g = 4, followed by the distance (1,0) and g = 3. For the mediastinal fat the best parameters were (0,1) and g = 4 followed by (1,1) and g = 4. Moreover, for the pericardium, (0,1) and g = 1 followed by (0,1) and g = 2 were the ones most valuable.

Furthermore, with relation to the features based on the run length matrix, the orientations 0° and 90° were the most valuable for the epicardial fat, where 90° was the only valuable orientation for the run percentage feature. For the mediastinal, the most valuable orientations were 45°, 90° and 135°, in this order. However, for the pericardium, the most valuable orientations were, respectively, 0°, 45° and 135°. Conclusively, in our problem, it is extremely difficult to set parameters for extracting the features. As we can see, the importance of the features are very distinct for each type of fat. The run percentage is not significant for the pericardium due to the fact that it appears just as a small contour of fat around the heart where the run length matrix cannot extract much information.

Moreover, a few features were counter-intuitively ranked. For instance, z was basically irrelevant among all of the three classes. However, theoretically, it appears to be a valuable feature since the arrangement of the types of fat varies greatly from one slice to the

other (on the *z* index). Nevertheless, if the RandomForest algorithm was regarded in this analysis, maybe the results would change significantly with relation to this feature. Another surprise was the ranking of the run percentage and grey level non-uniformity features, both are based on the run length matrix but none could be more significant than the moments of the co-occurrences in any of the three occasions. Since the mediastinal fat is dense we thought that the two features based on the run length matrix would perform better than the co-occurrence moment for this type of fat. However, that was not what occurred.

The coefficient of smooth variation performed much better than the arithmetic mean on the case of the mediastinal fat, the mean arithmetic feature was not even considered relevant. The CSV performed better than the arithmetic mean on the pericardium as well. However, it lost for the pixel value and for the mean arithmetic on the epicardial fat. Nevertheless, we tend to think that repeating this evaluation over the RandomForest would at least slightly change that panorama. That is, as we have seen in the previous Section 3.2.4, the segmentation produced with the REPTree relies much more on attributes such as the *x* and *y* than on the texture-based attributes.

The tertiary features are unarguably the most expensive to be extracted. Thus, even by this simple analysis, we aim to remove the features based on the RLM matrix due to their cost-effectiveness coefficient, and to remove the geometric moments as well. The geometric moments were considered relevant in all of the three cases but in any of them they were significantly important for the predictive model. Moreover, some set of parameters that performed worse than others are to be removed as well. In a real-time scenario, that cleaning could greatly impact on the speed of the conceived system.

4 RESULTS

Until now, we have (1) proposed and evaluated a registration method for standardizing every patient, (2) extracted some features from their registered images, (3) provided these features to classification algorithms in order to train predictive models that are used to classify an incoming instance, (4) evaluated the results of these segmentations over a significant number of classifiers and (5) analyzed the importance of the extracted features.

RandomForest was the algorithm that performed the best if the speed, accuracy, overfitting and segmentation analyses are considered. To refine our results we generated a dataset of approximately 2.5 gigabytes originated from 16 patients of the ground truth. In this occasion, no slice were skipped during the training set. The extracted dataset is also available directly on the Weka's arff format at [114]. The 10-fold cross validation and the 66% split evaluation were both regarded as test modes. The difference between the two is not expressive due to the huge amount of instances (pixels) in the dataset. The Table 6 contains the accuracies and the confusion matrixes of the RandomForest algorithm with standard parameters (which are: -I 10 -K 0 -S 1) over 16 patients, using a neighborhood of 25x25 pixels and obtained through the 66%-split test mode, whereas Table 7 contains the values obtained through the 10-fold cross validation. The 25x25 size for the neighborhood was chosen on the basis of the results shown in Figure 43.

Tissue	Accuracy	TP Rate	TN Rate	FP Rate	FN Rate
Epicardial fat ^a	98.3%	98.1%	98.4%	1.6%	1.5%
Mediastinal fat ^b	98.0%	92.9%	98.8%	1.1%	1.1%
Pericardium ^c	97.7%	81.6%	98.9%	1.0%	1.0%
Total of regarded instances (a,b,c): 1 373 079 Total of positives: a: 269 343, b: 188 222, c: 96 540					

Table 6: Random Forest algorithm on 66% split validation.

Tissue	Accuracy	TP Rate	TN Rate	FP Rate	FN Rate
Epicardial fat ^a	98.5%	98.3%	98.5%	1.4%	1.4%
Mediastinal fat ^b	98.4%	94.2%	99%	1%	0.9%
Pericardium ^c	98.0%	81.9%	99.1%	0.9%	0.9%
Total of reaarded instanc	es (a.b.c): 4 038	469 Total of	positives: a: 79	4 297. b: 552 0	78. c: 284 305

Table 7: Random Forest algorithm on 10-fold cross validation.

4.1 COMPARING RESULTS

The Dice index is one of the most commonly used parameters in the literature for comparison of the automatic segmentation of cardiac fats. Nevertheless, it can be addressed in a slightly different manner by each author, producing distinct results. This work is the first to propose the use of classification algorithms for cardiac fat segmentation. The nature of the classification algorithms implies that they are tightly related to confusion matrixes and to the accuracy index. In fact, this is usually the standard way to evaluate the performance of a classification. However, just one of the three main related works evaluates their segmentation on the basis of accuracies and confusion matrixes. Furthermore, another usually done comparison is on the rate of successful segmentations. However, this rate is usually observed by the authors and is, therefore, highly subjective. In other words, one can simply state that a segmentation was successful, but this evaluation is prone to high variability.

The Table 8 relates the results of these three main related works. If it is the case that values are not provided by the authors, the respective cells were left blank. The work of Kakadiaris et al. [103] is semi-automated, while the works of Shahzad et al. [8] and Ding et al. [9] are fully automatic. All these three works proposed methods for segmenting just the epicardial fat and, therefore, we compare just our epicardial fat segmentation in this table.

Author	Dice Index	TP Rate
Kakadiaris et al.	-	85.6%
Shahzad et al.	89.15%	-
Ding et al.	93.0%	-
This work (epicardial)	97.9%	98.3%

Table 8: Comparison of the epicardial fat segmentation.

The Dice index can be computed differently to some extent. It can be achieved using the Equation (25) where G stands for the ground truth and H stands for the segmented image. In our approach, we have shifted the segmented pixels to their fat-respective colors. Taking our scheme as reference, $G \cap H$ represents the amount of matches of colored values of a segmented image to the ground truth and |G| + |H| represents the sum of matches of colored pixels to colored pixels and colored pixels to grey pixels.

$$D = \frac{2|G \cap H|}{|G| + |H|}$$
(25)

However, the Dice index can also be interpreted differently, that is, the $G \cap H$ may not account the matches of grey pixels if the authors assume that they want to quantify the similarity of the cardiac fat (colored) pixels only and disregard the similarity of the remaining (background and others, such as the thorax fat). When accounting the match of grey values as correct (a grey pixel of the ground truth and a grey pixel of the segmented image is considered correct) it is said that our method achieved a mean Dice index of 96.8%. However, when including the matching of black pixels as correct, it reached 97.5%. Accounting the similarity of the black pixels makes no sense since it is set as the background in our approach, however, one could take advantage of it to report a controversial improved index. However, when only pairs of colored pixels are regarded as correct matchings (i.e., a match of a colored pixel to a grey or black pixel is considered wrong), the mean Dice similarity index was equal to 80.1%. Table 9 compares the achieved mean Dice index over this reported variation. Moreover, for this evaluation, the blue contour of the pericardium was considered epicardial and mediastinal fat at the same time. In addition, the Dice index was achieved from the data of 6 randomly chosen patients.

Tissue	Only colored	Colored and grey	Colored, grey
		(excluding black)	and black
Epicardial	79.7%	97.9%	98.2%
Mediastinal	80.6%	94.3%	96.9%
Mean	80.1%	96.8%	97.5%

Table 9: A comparison of the Dice similarity variation.

4.2 VISUALIZATION

The images in Figure 46 are a comparison of a single manually segmented slice to the result of the proposed automatic segmentation. The green color denotes the mediastinal fat, red represents the epicardial and blue corresponds to the pericardium. All the colored pixels represent pixels within the fat range of a CT and, therefore, that is the reason for some discontinuities evidenced on the images. It is important to highlight that the blue color is not

present on the automatically segmented images due to the fact that it was interpreted as a transitional area between the epicardial and mediastinal fats. Thus, the pericardium classification was slightly different than the other two. That is, if a pixel was classified as red or green before being classified as blue, then it remains at its previously classified color. Otherwise, the pixel is painted yellow (both epicardial and mediastinal) instead of blue. The two images in Figure 47 correspond to the same patient shown in Figure 46 but this time reconstructed on a 3D model (all slices) after its automated segmentation. Although the quality of the model may not be the best, it is possible to perfectly distinguish the contour of the heart on the epicardial fat (red color). The automatically segmented slices of a second patient are also shown in Appendix C – Automatically Segmented Slices.



Figure 46: Manually (left) versus an automatically segmented slice (right).



Figure 47: 3D fat model of an automatically segmented patient.

5 CONCLUSION

In this chapter we present a brief conclusion and discuss further improvements for the method proposed in this work. Furthermore, we address details about the implementation, used equipment and infrastructure.

We started this work with the intent of automatically segmenting two types of fats from very distinct CT slices, with regard to the same patient and to others. Thus, applying classification algorithms to this problem seemed very plausible from the beginning. Otherwise, distinct approaches would have to consider such a great amount of variation that could make them unpractical, while machine learning algorithms are able to automatically learn and overcome portions of this variation. The achieved results are satisfactory but the current approach needs to be adapted in order to be applied in real time. Currently, with a huge set of extracted features and not sufficient effort applied to optimizing the code, the algorithm still takes approximately from 24 to 48 hours to fully segment a single patient scan. However, a more recent optimization have been indicating that it may be possible to segment from 3 to 4 patients a day, without loosing much accuracy.

A couple of works have already proposed semi and fully automated segmentations for the epicardial fat, which were addressed and compared in the previous chapter. However, every currently proposed method quantifies the epicardial fat based on the pericardium border, which makes the quantification of the mediastinal fat unpractical, since it is not located within the pericardium contour. Besides, in contrast to the other methods, the proposed approach is easily reproducible and of adaptive nature with regard to various types of modalities and medical images, apart from the fact of apparently producing better results.

As to the correctness extent, the epicardial fat was automatically segmented in a more reliable way than the mediastinal fat. In fact, the mediastinal fat is prone to a higher variation with relation to their spatial disposition and their volume among distinct patients and, therefore, that fact could induce these worse results. Furthermore, the caudal slices of each patient were also segmented in a worse fashion than the remaining. However, these slices contain extremely confuse disposition of the organs due to the positioning of the liver and the stomach with relation to the heart that varies on each patient, which makes them hard to segment even to a human specialist.

5.1 FURTHER IMPROVEMENTS

In the Chapters 3 and 4, we have seen that the segmentation produced with the RandomForest algorithm is sparse. That is actually an interesting outcome, since it creates a clear possibility for a post-processing improvement. A simple heuristic could optimize the segmentation by a significant amount by connecting the colored sparse points, for instance. Furthermore, since the epicardial fat follows the internal elliptical contour of the heart, a post processing such as an elliptical container could be adjusted to the segmented epicardial fat in order to reduce errors. Unfortunately, we were not able to implement and evaluate these hypotheses.

The dataset used in our work regarded CT images produced by different scanners from mainly two different manufacturers (Siemens and Philips). We did not include features related to the scanner model and manufacturer into the prediction due to the fact that, for every pixel of every slice of each patient (i.e., each features vector), these features would be equally repeated, significantly increasing the size of the extracted dataset. However, we believe that if these features were regarded or if distinct predictive models were generated for each model and manufacturer the results would also improve.

Ensemble methods use multiple learning algorithms to obtain better performance [122]. In this work, we have used a single classification algorithm to segment the images instead of a combination of them. The J48Graft could be combined to the RandomForest and perhaps to the REPTree algorithms to increase the overall performance of the predictive model. The RandomForest is itself considered an ensemble method. Still, it can properly be applied to other classification algorithms, in a sense, generating a more complex ensemble method.

In summary, the proposed methodology could have its convergence time improved if (1) areas are classified instead of pixels, which would generate a faster but worse result, if (2) a more robust selection of features is performed or if (3) computers with a great amount of cores are used in order to distribute the workload, processing a slice at each core (parallelism). On the other hand, the segmentation could be improved if the following statements were regarded: (1) considering more patients in order to generate a more generalizable predictive model during the training phase, (2) using ensemble methods instead of just one classifier. Furthermore, by (3) developing a post-processing technique to enhance the segmentation such as dilations of the colored pixels and/or applying elliptical containers to the epicardial fat, aiming to, for instance, automatically fill some of the produced gaps, or (4) by training the classification algorithm with CT images originated from a single model of scanner and manufacturer and applying the predictive model to scans of the same model. We believe that the considerations addressed in this section would significantly improve the efficiency and accuracy of the proposed approach.

5.2 FINAL REMARKS

All the processing rates provided by this work including all the time, accuracy and efficiency analysis were obtained from relatively common personal computers. The Oscar cluster from the Universidade Federal Fluminense, which fits in this definition, was also used in the process [123]. In summary, the number of CPU cores of the used computers varied from 2 to 8, the available amount of RAM varied from 4 to 8 gigabytes and no dedicated graphical card was used. All the processing was done on the CPU apart from the 3D rendering. The operating systems varied from Windows 8 to Linux distributions such as Fedora and Ubuntu and all the code used in this work was programmed in Java and run over Oracle's JVM. Although we have based our registration on the ITK scheme, we have not used the ITK library itself. The 3D modeling was done on LibGDX and the remaining was provided by the Oracle's JDK or programmed from scratch.

We finally conclude that the appliance of classification algorithms on image segmentation, as long as the right features are selected, is highly prone to success and may surpass many usual segmentation methods on several aspects. RandomForest is commonly rated as one of the best decision-tree algorithms and proved its efficiency in our analysis. We have also concluded that decision tree algorithms provided much better performance over neural networks and function-based classification algorithms. The achieved mean accuracy for the epicardial and mediastinal fats was 98.4% with a mean true positive rate of 96.2%. The mean Dice similarity index was 96.8%. Every registration was considered successful and every segmentation could also be considered successful in a sense that no major error or unpredicted behavior occurred.

REFERENCES

- [1] Aspen Publishers, "Web-based diagnosis support system seeks to reduce delays in medical diagnoses," Managed care outlook, vol. 23, no. 13, p. 8, 2010.
- [2] G. A. Rosito, J. M. Massaro, U. Hoffman, F. L. Ruberg, A. A. Mahabadi, R. S. Vasan, C. J. O'Donnel, C. S. Fox, "Pericardial fat, visceral abdominal fat, cardiovascular disease risk factors, and cardiovascular calcification in a community-based sample: the framingham heart study," Circulation, vol. 117, no. 605-613, 2008.
- [3] R. Sicari, A. Maria Sironi, R. Petz, F. Frassi, D. M. Chubuchny, V. Positano, M. Lombardi, E. Picano, A. Gastaldelli, "Pericardial Rather Than Epicardial Fat is a Cardiometabolic Risk Marker: An MRI vs Echo Study," Journal of the American Society of Echocardiography, vol. 24, no. 10, pp. 1156-1162, 2011.
- [4] G. Iacobellis, M. C. Ribaudo, F. Assael, E. Vecci, C. Tiberti, A. Zappaterreno, U. Di Mario, F. Leonetti, "Echocardiographic epicardial adipose tissue is related to anthropometric and clinical parameters of metabolic syndrome: A new indicator of cardiovascular risk," The Journal of clinical endocrinology and metabolism, vol. 88, no. 11, pp. 5163-5168, 2003.
- [5] T. S. Polonsky, R. L. McClelland, N. W. Jorgensen, D. E. Bild, G. L. Burke, A. D. Guerci,
 P. Greenland, "Coronary Artery Calcium Score and Risk Classification for Coronary
 Heart Disease Prediction," Journal of the American Medical Association, vol. 303, no.
 16, pp. 1610-1616, 2010.
- [6] G. Coppini, R. Favilla, P. Marraccini, D. Moroni, G. Pieri, "Quantification of Epicardial Fat by Cardiac CT Imaging," The Open Medical Informatics Journal, vol. 4, pp. 126-135, 2010.
- [7] G. Iacobellis, E. Lonn, A. Lamy, N. Singh, A. M. Sharma, "Epicardial fat thickness and coronary artery disease correlate independently of obesity," International Journal of Cardiology, pp. 452-454, 2011.
- [8] T. Walsum, R. Shahzad, D. Bos, C. Metz, A. Rossi, H. Kirili, A. Lugt, S. Klein, J.
 Witteman, P. Feyter, W. Niessen, L. Vliet, "Automatic quantification of epicardial fat

volume on non-enhanced cardiac CT scans using a multi-atlas segmentation approach," Medical Physics, vol. 40, no. 9, 2013.

- X. Ding, D. Terzopoulos, M. Diaz-Zamudio, D. S Berman, J. Piotr, "Automated Epicardial Fat Volume Quantification from Non-Contrast CT," Medical Imaging 2014: Image Processing, vol. 9034, 2014.
- [10] J. V. Spearman, F. G. Meinel, U. J. Schoepf, P. Apfaltrer, J. R. Silverman, A. W. Krazinski, C. Canstein, C. N. De Cecco, P. Costello, L. L. Geyer, "Automated Quantification of Epicardial Adipose Tissue Using CT Angiography: Evaluation of a Prototype Software," European Radiology, vol. 24, no. 2, pp. 519-526, 2014.
- [11] D. Dey, N. D. Wong, B. Tamarappo, R. Nakazato, H. Gransar, V. Y. Cheng, A. Ramesh, I. Kakadiaris, G. Germano, P. J. Slomka, D. S. Berman, "Computer-aided non-contrast CT-based quantification of pericardial and thoracic fat and their associations with coronary calcium and metabolic syndrome," Atherosclerosis, vol. 209, pp. 136-141, 2010.
- [12] V. Mahadevan, "Anatomy of the heart," Surgery (Oxford), vol. 30, no. 1, pp. 5-8, 2012.
- [13] StudyBlue. (2015) Cardiovascular systems. [Online].
 https://www.studyblue.com/notes/note/n/cardiovascular-system/deck/3238038
- [14] H. S. Sacks, J. N. Fain, "Human epicardial adipose tissue: A review," The American heart journal, vol. 153, no. 6, pp. 907-917, 2007.
- [15] J. H. Nichols, B. Samy, K. Nasir, C. S. Fox, P. C. Schulze, F. Bamberg, U. Hoffmann, "Volumetric measurement of pericardial adipose tissue from contrast-enhanced coronary computed tomography angiography: A reproducibility study," Journal of cardiovascular computed tomography, vol. 2, no. 5, pp. 288-295, 2008.
- [16] M. Marwan, S. Achenbach, "Quantification of epicardial fat by computed tomography: Why, when and how?," Journal of cardiovascular computed tomography, vol. 7, no. 1, p. 3, 2013.
- [17] I. S. Yarza, G. V. Ramíez, R. S. Mendiguren, P. J. S. Roblero, M. Paramo, G. Bastarrika,
 "Feasibility of epicardial adipose tissue quantification in non-ECG-gated low-

radiation-dose CT: comparison with prospectively ECG-gated cardiac CT," Acta radiologica, vol. 53, no. 5, pp. 536-540, 2012.

- [18] S. Dabbah, H. Komarov, A. Marmor, N. Assy, "Epicardial fat, rather than pericardial fat, is independently associated with diastolic filling in subjects without apparent heart disease," NMCD. Nutrition Metabolism and Cardiovascular Diseases, pp. 1-6, 2014.
- [19] A. Yerramasua, D. Dey, S. Venuraju, D. V. Anand, S. Atwal, R. Corder, D. S. Berman, A. Lahiri, "Increased volume of epicardial fat is an independent risk factor for accelerated progression of sub-clinical coronary atherosclerosis," Atherosclerosis, vol. 220, no. 1, pp. 223-230, 2012.
- [20] R. Djaberi, J. D. Schuijf, J. M. Werkhoven, G. Nucifora, J. W. Jukema, J. J. Bax,
 "Relation of Epicardial Adipose Tissue to Coronary Atherosclerosis," The American Journal of Cardiology, vol. 102, no. 12, pp. 1602-1607, 2008.
- [21] T. Choi, N. Ahmadi, S. Sourayanezhad, I. Zeb, M. J. Budoff, "Relation of vascular stiffness with epicardial and pericardial adipose tissues, and coronary atherosclerosis," Atherosclerosis, vol. 229, no. 1, pp. 118-123, 2013.
- [22] E.L. Thomas, J.A. Fitzpatrick, S.J. Malik, S.D. Taylor-Robinson, J.D. Bell, "Whole body fat: Content and distribution," Progress in nuclear magnetic resonance spectroscopy, vol. 73, no. 56-80, 2013.
- T. Oka, H. Yamamoto, N. Ohashi, T. Kitagawa, E. Kunita, H. Utsunomiya, R. Yamazato,
 Y. Urabe, J. Horiguchi, K. Awai, Y. Kihara, "Association between epicardial adipose tissue volume and characteristics of non-calcified plaques assessed by coronary computed tomographic angiography," International journal of cardiology, vol. 161, no. 1, pp. 45-49, 2012.
- [24] A. A. Mahabadi, N. Lehmann, H. Kälsch, T. Robens, M. Bauer, I. Dykun, T. Budde, S. Moebus, K. H. Jöckel, R. Erbel, S. Möhlenkamp, "Association of Epicardial Adipose Tissue With Progression of Coronary Artery Calcification Is More Pronounced in the Early Phase of Atherosclerosis : Results From the Heinz Nixdorf Recall Study," JACC: Cardiovascular Imaging, vol. 7, no. 9, pp. 909–916, 2014.

- [25] A. A. Mahabadi, M. H. Berg, N. Lehmann, H. Kälsch, M. Bauer, K. Kara, N. Dragano, S. Moebus, K. H. Jöckel, R. Erbel, S. Möhlenkamp, "Association of Epicardial Fat With Cardiovascular Risk Factors and Incident Myocardial Infarction in the General Population," Journal of the American College of Cardiology, vol. 61, no. 13, pp. 1388-1395, 2013.
- [26] C. L. Schlett, M. Ferencik, M. F. Kriegel, F. Bamberg, B. B. Ghoshhajra, S. B. Joshi, J. T. Nagurney, C. S. Fox, Q. A. Truong, U. Hoffmann, "Association of pericardial fat and coronary high-risk lesions as determined by cardiac CT," Atherosclerosis, vol. 222, no. 1, pp. 129-134, 2012.
- [27] M. O. Al Chekakie, C. C. Welles, R. Metoyer, A. Ibrahim, A. R. Shapira, J. Cytron, P. Santucci, D. J. Wilber, J. G. Akar, "Pericardial fat is independently associated with human atrial fibrillation," Journal of the American College of Cardiology, vol. 56, no. 10, pp. 784-788, 2010.
- [28] M. Granér, R. Siren, K. Nyman, J. Lundbom, A. Hakkarainen, M. O. Pentikainen, K. Lauerma, N. Lundbom, M. Adiels, M. S. Nieminen, M. Taskinen, "Cardiac Steatosis Associates With Visceral Obesity in Nondiabetic Obese Men," The Journal of clinical endocrinology and metabolism, vol. 98, no. 3, pp. 1189-1197, 2013.
- [29] F. Atalar, S. Gormez, B. Caynak, G. Akan, G. Tanriverdi, S. Bilgic-Gazioglu, D. Gunay,
 C. Duran, B. Akpinar, U. Ozbek, A. S. Buyukdevrim, Z. Yazıcı, "Mediastinal adipose tissue expresses a pathogenic profile of 11 β-hydroxysteroid dehydrogenase Type 1, glucocorticoid receptor, and CD68 in patients with coronary artery disease,"
 Cardiovascular pathology, vol. 22, no. 3, pp. 183-188, 2013.
- [30] R. Taguchi, J. Takasu, Y. Itani, R. Yamamoto, K. Yokoyama, S. Watanabe, Y. Masuda,
 "Pericardial fat accumulation in men as a risk factor for coronary artery disease,"
 Atherosclerosis, vol. 157, no. 1, pp. 203-209, 2001.
- [31] P. Raggi, "Epicardial Adipose Tissue as a Marker of Coronary Artery Disease Risk,"Journal of the American College of Cardiology, vol. 61, no. 13, pp. 1396-1397, 2013.
- [32] T.E. Brinkley, F.-C. Hsu, J.J. Carr, W.G. Hundley, D.A. Bluemke, J.F. Polak, J. Ding, "Pericardial fat is associated with carotid stiffness in the Multi-Ethnic Study of

Atherosclerosis," NMCD. Nutrition Metabolism and Cardiovascular Diseases, vol. 21, no. 5, pp. 332-338, 2011.

- [33] P. M. Gorter, A. M. de Vos, Y. Graaf, P. R. Stella, P. A. Doevendans, M. F. L. Meijs, M. Prokop, F. L. J. Visseren, "Relation of Epicardial and Pericoronary Fat to Coronary Atherosclerosis and Coronary Artery Calcium in Patients Undergoing Coronary Angiography," The American Journal of Cardiology, vol. 102, no. 4, pp. 380-385, 2008.
- [34] P. Raggi, P. Alakija, "Epicardial adipose tissue: A long-overlooked marker of risk of cardiovascular disease," Atherosclerosis, vol. 229, no. 1, pp. 32-33, 2013.
- [35] W. Chen, J. Huang, M.-H. Hsieh, Y.-J. Chen, "Extremely high coronary artery calcium score is associated with a high cancer incidence," International Journal of Cardiology, vol. 155, pp. 474-475, 2012.
- [36] O. Chenn, I. Ahmad, B. Hua, J. A. Sockolow, I. Klem, T. Sacchi, J. F. Heitner,
 "Correlation of Pericardial and Mediastinal Fat With Coronary Artery Disease,
 Metabolic Syndrome, and Cardiac Risk Factors," Journal of Cardiovascular Magnetic
 Resonance, vol. 53, no. 10, pp. A283-A283, 2009.
- [37] S. Bonapace, G. Perseghin, G. Molon, G. Canali, L. Bertolini, G. Barbieri, G. Targher,
 "Nonalcoholic Fatty Liver Disease Is Associated With Left Ventricular Diastolic
 Dysfunction in Patients With Type 2 Diabetes," Cardiovascular and Metabolic Risk,
 vol. 58, pp. S542-S543, 2013.
- [38] K. Dudas, G. Lappas, S. Stewart, A. Rosengren, "Trends in Out-of-Hospital Deaths Due to Coronary Heart Disease in Sweden (1991 to 2006)," Circulation, vol. 123, no. 1, pp. 46-52, 2011.
- [39] C. T. Escoffery, S. E. Shirley, "Causes of sudden natural death in Jamaica: A medicolegal (coroner's) autopsy study from the University Hospital of the West Indies," Forensic science international, vol. 129, no. 2, pp. 116-121, 2002.
- [40] W. D. Bidgood, S. C. Horii, F. W. Prior, D. E. Syckle, "Understanding and Using DICOM, the Data Interchange Standard for Biomedical Imaging," Journal of American Medical Informatics Association, vol. 4, no. 3, 1997.

- [41] S. Suen, E. Lam, K. Wong, "Photographic stitching with optimized object and color matching based on image derivatives," Optics Express, vol. 15, no. 12, pp. 7689– 7696, 2007.
- [42] J. B. A. Maintz, M. A. Viergever, "A survey of medical image registration," Medical Image Analysis, vol. 2, no. 1, pp. 1-36, 1998.
- [43] L. F. Silva, G. O. S. Olivera, S. Galvão, J. B. Silva, A. A. Santos, S. M. D. Augusto, D. C. Muchaluat-Saade, A. Conci, "Analise de Séries Temporais de Sinais Térmicos da Mama para Detecção de Anomalias," WIM - XIV Workshop de Informática Médica – Anais CSBC, pp. 1818-1827, 2014.
- [44] A. Gooßen, T. Pralow, R. R. Grigat, "Automatic stitching of digital radiographies using image interpretation," Medial Image Understanding and Analysis, Proceedings of the Twelfth Annual, pp. 204-208, 2008.
- [45] S. Samsudin, S. Adwan, H. Arof, N. Mokhtar, F. Ibrahim, "Development of Automated Image Stitching," Journal of Digital Imaging, vol. 26, pp. 361-370, 2013.
- [46] ITK. (2013, December) The ITK Software Guide, Third Edition. [Online].http://www.itk.org/ItkSoftwareGuide
- [47] E. Schreibmann, B. Thorndyke, T. Li, J. Wang, L. Xing, "Four-dimentional image registration for image-guided radiotherapy," Journal of Radiation Oncology, vol. 71, no. 2, pp. 578-586, 2008.
- [48] B. Zitová, J. Flusser, "Image registration methods: a survey," Image and Vision Computing, vol. 21, no. 11, pp. 977–1000, 2003.
- [49] J. M. McCarthy, Introduction to Theoretical Kinematics.: MIT Press. ISBN 0262132524, 1990.
- [50] R. Galarza, A. Irene, J. Seade, Introduction to classical geometries.: Birkhauser, 2007.
- [51] M. Gabrani, O. J. Tretiak, "Elastic transformations," Signals, Systems and Computers, vol. 1, pp. 501-505, 1996.
- [52] A. Zisserman, R. Hartley, Multiple View Geometry in Computer Vision. UK: Cambridge University Press, 2004.
- [53] P. K. Schneider, D. H. Eberly, Geometric Tools for Computer Graphics.: Morgan Kaufmann. ISBN 978-1-55860-594-7, 2003.
- [54] OpenCV. (2014) OpenCV Documentation. [Online]. http://docs.opencv.org/doc/tutorials/imgproc/imgtrans/warp_affine/warp_affine.ht ml
- [55] Oracle. (2014) Oracle Java Documentation. [Online].http://docs.oracle.com/javase/7/docs/api/java/awt/geom/AffineTransform.html
- [56] T. Mäkelä, P. Clarysse, O. Sipilä, N. Pauna, Q. C. Pham, "A Review of Cardiac Image Registration Methods," IEEE Transactions on Medical Imaging, vol. 21, no. no. 9, 2002.
- [57] J. P. W. Pluim, J. B. A. Maintz, M. A. Viergever, "Mutual information based registration of medical images: a survey," IEEE Transactions on Medical Imaging, vol. 22, no. 8, pp. 1-36, 2003.
- [58] M. J. Clarkson, D. Rueckert, A. P. King, P. J. Edwards, D. L. G. Hill, D. J. Hawkes, "Registration of video images to tomographic images by optimising mutual information using texture mapping," Medical Image Computing and Computer-Assisted Intervention, vol. 1679, pp. 579–588, 1999.
- [59] M. J. Clarkson, D. Rueckert, D. L. G. Hill, D. J. Hawkes, "Registration of multiple video images to pre-operative CT for image guided surgery," Medical Imaging: Image Processing, vol. 3661, pp. 14–23, 1999.
- [60] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, D. J. Hawkes, "A comparison of similarity measures for use in 2D-3D medical image registration," IEEE Transactions on Medical Imaging, vol. 17, no. 4, pp. 586–595, 1999.
- [61] L. Zo"llei, E. Grimson, A. Norbash, W. Wells, "2D-3D rigid registration of X-ray fluoroscopy and CT images using mutual information and sparsely sampled histogram estimators," Computer Vision and Pattern Recognition - IEEE Computer Society Press, vol. 2, pp. 696–703, 2001.

- [62] J. Kim, J. A. Fessler, K. L. Lam, J. M. Balter, R. K. Ten Haken, "A feasibility study on mutual information based set-up error estimator for radiotherapy," Medical Physics, vol. 28, no. 12, pp. 2507–2517, 2001.
- [63] D. Plattard, M. Soret, J. Troccaz, P. Vassal, J.-Y. Giraud, G. Champleboux, X. Artignan,
 M. Bolla, "Patient set-up using portal images: 2D/2D image registration using mutual information," Computer Aided Surgery, vol. 5, no. 4, pp. 246–262, 2000.
- [64] R. Bansal, L. Staib, Z. Chen, A. Rangarajan, J. Knisely, R. Nath, J. Duncan, "A minimax entropy registration framework for patient setup verification in radiotherapy,"
 Computer Aided Surgery, vol. 4, no. 6, pp. 287–304, 1999.
- [65] D. Sarrut, S. Clippe, "Geometrical transformation approximation for 2D/3D intensitybased registration of portal images and CT scan," Medical Image Computing and Computer- Assisted Intervention, vol. 2208, pp. 532– 540, 2001.
- [66] J. L. Boes, B. Kim, P. H. Bland, K. R. Zasadny, P. V. Kison, K. Koral, K. A. Frey, R. L. Wahl, C. R. Meyer, "Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations," Medical Image Analysis, vol. 1, no. 3, pp. 195–206, 1997.
- [67] K. F. Koral, S. Lin, J. A. Fessler, M. S. Kaminski, R. L. Wahl, "Preliminary results from intensity-based CT-SPECT fusion in I-131 anti-B1 monoclonal-antibody therapy of lymphoma," Cancer, vol. 80, no. 12, pp. 2538–2544, 1997.
- [68] G. C. Kagadis, K. K. Delibasis, G. K. Matsopoulos, N. A. Mouravliansky, P. A. Asvestas,
 "A comparative study of surface and volume-based techniques for the automatic registration between CT and SPECT brain images," Medical Physics, vol. 29, no. 2, pp. 201–213, 2002.
- [69] P. A. Elsen, E. J. D. Pol, M. A. Viergever, "Medical image matching A review with classification," IEEE Eng. Med. Biol., vol. 12, no. 2, pp. 16–39, 1993.
- [70] D. L. G. Hill, P. G. Batchelor, M. H. Holden, D. J. Hawkes, "Medical image registration," Phys. Med. Biol., vol. 46, no. 1, pp. 1-45, 2001.

- [71] M. A. Wirth, C. Choi, A. Jennings, "Point-to-point registration of nonrigid medical images using local elastic transformation methods," IEEE Int. Conf. Image Processing and Its Applications, pp. 780–784, 1997.
- [72] P. D. Sathy, R. Kayalvizhi, "Modified bacterial foraging algorithm based multilevel thresholding for image segmentation," Engineering applications of artificial intelligence, vol. 24, no. 4, pp. 595-615, 2011.
- [73] I. Isgun, "Multi-Atlas-Based Segmentation With Local Decision Fusion—Application to Cardiac and Aortic Segmentation in CT Scans," IEEE Transactions on Medical Imaging, vol. 28, no. 7, pp. 1000-1010, 2009.
- [74] H. A. Kirili, M. Schaap, S. Klein, S. L. Papadopoulou, M. Bonardi, C. H. Chen, A. C. Weustink, N. R. Mollet, E. J. Vonken, R. J. Geest, T. Walsum, W. J. Niessen, "Evaluation of a multi-atlas based method for segmentation of cardiac CTA data: alarge-scale, multicenter, and multivendor study," Medical Physics, vol. 37, no. 12, pp. 6279-6291, 2010.
- [75] Xiaohong Zhao, Yanhua Qu, Hong Zhang, "Sports Video Segmentation using Spectral Clustering," Journal of Multimedia, vol. 9, pp. 873-873, 2014.
- [76] "Edge detection by spherical separation," Computational management science, vol.11, no. 4, pp. 517-530, 2014.
- [77] A. Belaid, "Phase-based level set segmentation of ultrasound images," IEEE
 Transactions on Information Technology in Biomedicine, vol. 15, no. 1, pp. 138-147, 2011.
- [78] L. Xiangyun, Y. Zhiyong, Z. Qi, Y. Qian, Z. Dong, Z. Jianhui, "Multi-scale and shape constrained localized region-based active contour segmentation of uterine fibroid ultrasound images in HIFU therapy," PLoS ONE, vol. 9, no. 7, 2014.
- [79] P. N. Happ, "A region-growing segmentation algorithm for GPUs," IEEE Geoscience and Remote Sensing Letters, vol. 10, no. 6, pp. 1612-1616, 2013.
- [80] A. M. Mharib, A. R. Ramli, S. Mashodor, R. B. Mahmood, "Survey on liver CT image segmentation methods," The Artificial Intelligence Review, vol. 37, no. 2, pp. 83-95, 2012.

- [81] K.S. Fu, J. K. Mui, "A survey on image segmentation," Pattern Recognition, vol. 13, no. 1, pp. 3-16, 1981.
- [82] J. C. Bezdek, L. O. Hall, P. Clarke, "Review of MR image segmentation techniques using pattern recognition," Medical Physics, vol. 20, pp. 1033-1048, 1993.
- [83] C. Cortes V. Vapnik, "Machine Learning," Support-vector networks, vol. 20, no. 3, 1995.
- [84] F. Rosenblatt, Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Washington, DC: Spartan Books, 1961.
- [85] E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," Journal of Machine Learning Research, vol. 1, pp. 211-244.
- [86] A. Verikas, A. Gelzinis, M. Bacauskiene, "Mining data with random forests: A survey and results of new tests," Pattern Recognition, vol. 44, pp. 330-349, 2011.
- [87] A. Abdulkadir, B. Mortamet, P. Vemuri, C. R. Jack, G. Krueger, S. Kloppel, "Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier," NeuroImage, vol. 58, no. 3, pp. 785-792, 2011.
- [88] E. O. Rodrigues, A. Conci, T. B. Borchartt, A. C. Paiva, A. C. Silva, T. MacHenry,
 "Comparing results of thermographic images based diagnosis for breast diseases," in
 Proceedings of Systems, Signals and Image Processing (IWSSIP), 2014.
- [89] NeuroImage, "A prior feature svm-mrf based method for mouse brain segmentation," vol. 59, no. 3, pp. 2298-2306, 2012.
- [90] S. Amiri, M. M. Movahedi, K. Kazemi, H. Parsaei, "An Automated MR Image Segmentation System Using Multi-layer Perceptron Neural Network," Journal of Biomedical Physics and Engineering, vol. 3, no. 4, pp. 115-22, 2013.
- [91] C. Voyant, P. Randimbivololona, M. L. Nivet, C. Paoli, M. Muselli, "Twenty four hours ahead global irradiation forecasting using multi-layer perceptron," Meteorological applications, vol. 21, no. 3, pp. 644–655, 2014.
- [92] T. G. Dietteric, "An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization," Machine Learning, vol. 40, no. 2, pp. 139–157, 2000.

- [93] L. Breiman, "Random forests," Machine Learning, vol. 45, pp. 5-32, 2001.
- [94] T. K. Ho, "The Random Subspace Method for Constructing Decision Forests," EE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, vol. 20, no. 8, 1998.
- [95] R. O. Duda, P. E. Hart, D. G. Stork, Pattern Classification (2nd edition). New York: John Wiley & Sons, 2001.
- [96] D. Meyer, F. Leisch, K. Horni, "The support vector machine under test," Neurocomputing, vol. 55, no. 2, pp. 169–186, 2003.
- [97] G. S. Babu, S. Suresh, "Meta-cognitive RBF Network and its Projection Based Learning algorithm for classification problems," Applied soft computing, vol. 13, no. 1, pp. 654-666, 2013.
- [98] B. Ginneken, M. B. Stegmann, M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," Medical Image Analysis, vol. 10, pp. 19-40, 2006.
- [99] R. S. Marques, A. Conci, T. Machenry, "On the use of Hausdorff Distance on Evaluation of Breast Image segmentations," in Biomat, Toronto, 2013.
- [100] E. Rikxoort, Y. Arzhaeva, B. Ginneken, "Automatic segmentation of the liver in computed tomography scans with voxel classification an atlas matching," 3D Segmentation in The Clinic: A Grand Challenge, pp. 101-108, 2007.
- [101] D. Dey, Y. Suzuki, S. Suzuki, M. Ohba, P. J. Slomka, D. Polk, L. J. Shaw., D. S. Berman, "Automated quantitation of pericardiac fat from noncontrast CT," Investigative Radiology, vol. 43, no. 2, pp. 145-153, 2008.
- [102] A. Pednekar, "Automatic Segmentation of Abdominal Fat from CT Data," Application of Computer Vision, pp. 308–15, 2005.
- [103] I. Kakadiaris, A. Bandekar, K. Gul, S. S. Mao, J. Takasu, M. Budoff, M. Naghavi,
 "Automated pericardial fat quantification in CT data," Proceedings of Journal of the American College of Cardiology, vol. 47, no. 4, p. 264A, 2006.

- [104] J. G. Barbosa, B. Figueiredo, N. Bittencourt, J. M. R.S. Tavares, "Towards automatic quantification of the epicardial fat in non-contrasted CT images," Computer Methods in Biomechanics and Biomedical Engineering, vol. 14, no. 10, pp. 905-914, 2011.
- [105] T. D. Ruder, Y. Thali, S. T. Schindera, S. A. D. Torre, W. Zech, M. J. Thali, S. Ross, G. M.
 Hatch, "How reliable are Hounsfield-unit measurements in forensic radiology?,"
 Forensic Science International, vol. 220, pp. 219-233, 2012.
- [106] R. Molteni, "Prospects and challenges of rendering tissue density in Hounsfield units for cone beam computed tomography," vol. 116, no. 1, pp. 105-119, 2013.
- [107] H. Shmilovich, D. Dey, V. Y. Cheng, R. Rajani, R. Nakazato, Y. Otaki, R. Nakanishi, P. J. Slomka, L. E.J. Thomson, S. W. Hayes, N. D. Wong et al., J. D. Friedman, H. Gransar, "Threshold for the upper normal limit of indexed epicardial fat volume: Derivation in a healthy population and validation in an outcome-based study," The American Journal of Cardiology, vol. 108, no. 11, pp. 1680–1685, 2011.
- [108] NEMA. (2009) The Dicom Standard. [Online]. http://medical.nema.org/standard.html
- [109] P. O. Pianykh, Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide, 2nd ed.: Springer, 2009.
- [110] J. Bozek, M. Grgic, J. A. Schnabel, "Validation of Rigid Registration of mammographic images," ELMAR Proceedings, pp. 11-16, 2011.
- [111] J. S. Hyde, A. Jesmanowicz, "Cross-correlation: An fMRI signal-processing strategy," NeuroImage, vol. 62, no. 2, pp. 848-851, 2012.
- [112] H. Yuanyuan, F. Xunli, L. Jianghong, "Image retrieval using mutual information," Journal of Theoretical and Applied Information Technology, vol. 50, no. 1, pp. 284-290, 2013.
- [113] A. C. Pocock, Feature Selection Via Joint Likelihood. Manchester, 2012.
- [114] Visual Lab. (2014) A Computed Tomography Cardiac Fat Dataset. [Online]. http://visual.ic.uff.br/en/cardio/ctfat/index.php
- [115] R. E. Woods, R. C. Gonzalez, Processamento de Imagens Digitais. São Paulo, Brazil: Edgard Blucher, 2000.

- [116] A. Kumar, R. Paramesran, "Geometric moment extraction equipment for image processing applications," Review of scientific instruments, vol. 85, no. 4, 2014.
- [117] Intel Developer Zone. (2014) Image Statistic Functions Moments. [Online]. https://software.intel.com/sites/products/documentation/hpc/ipp/ippi/ippi_ch11/c h11_image_moments.html
- [118] Y. K. Liu, B. Zalik, P. Wang, D. Podgorelec, "Directional difference chain codes with quasi-lossless compression and run-length encoding," Signal Processing: Image Communication, vol. 27, no. 9, pp. 973-984, 2012.
- [119] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, "The WEKA Data Mining Software: An Update," SIGKDD Explorations, vol. 11, no. 1, 2009.
 [Online]. http://www.cs.waikato.ac.nz/ml/weka/
- [120] J. P. A. Ioannidis, I. Hozo, B. Djulbegovic, "Optimal type i and type II error pairs when the available sample size is fixed," Journal of clinical epidemiology, vol. 66, no. 8, pp. 903-e2, 2013.
- [121] T. M. Khoshgoftaar, E. B. Allen, "Controlling overfitting in classification-tree models of software quality," Empirical software engineering, vol. 6, no. 1, pp. 59-79, 2001.
- [122] D. Optiz, R. Maclin, "Popular ensemble methods: An empirical study," Artificial Intelligence Research, vol. 11, pp. 169-198, 1999.
- [123] Universidade Federal Fluminense. (2014) Oscar Cluster. [Online]. http://suporte.ic.uff.br/index.php/servicos/posgrad/oscarusar

APPENDIX A — ACCESSING CT DATA FROM DICOM FILES

The DICOM acronym stands for Digital Imaging and Communications in Medicine, which consists of a set of protocols, encoding and values that define how medical data should be stored, accessed and transmitted. The Section 2.3 of this work provides a general literature review of the DICOM standard.

DICOM was firstly introduced for the purpose of standardizing the various formats, encoding and transmission protocols of medical images and related data. Such topics were formerly defined distinctively by each single manufacturer, therefore, hardening the reproducibility of the data. In general, DICOM files contain more than just image-related information. They may also hold patient-related information. Among these related data there virtually, namely, the full name of the patient, sex and birth date. Furthermore, they also contain important acquisition-related data such as the type of equipment used and a couple of related settings.

A DICOM object (i.e., a DICOM file) is comprised of DICOM elements or, alternatively, DICOM attributes. Every DICOM element has a tag, a data type (VR – acronym for value representation), length and value. The value is what we usually aim to retrieve by accessing an element when possessing its identification tag. In other words, the value is the wanted information itself. The Figure A.1 illustrates DICOM elements in a data stream.



Figure A.1: Encoding of a DICOM element in a DICOM data stream [DICOM Standard -

Chapter 5].

Every DICOM element has an identification tag that uniquely defines the element and its properties. Moreover, a DICOM tag is comprised of two hexadecimal values of four digits each (e.g., 0028,0012). The first value is defined as the group index and the second one as element index. DICOM tags that are related to one another sometimes belong to the same group, but not strictly always. For instance, the tag (0028,0010) represents the number of rows of an image and (0028,0011) represents the number of columns. The data type (VR) is represented as a two characters code. The VR field may receive: (1) US for unsigned short, (2) UI for unique identifier, (3) CS for coded string, (4) OB for other byte, etc. Although this is not what the Integrating the Healthcare Enterprise (IHE) recommends, the VR data is omitted in some cases for being redundant. By being redundant we mean that, for instance, the number of columns and rows of an image will always be an integer and, hence, there is a redundancy on declaring the data type.

The DICOM files are of binary format, differing from textual-based file formats such as XML and HTML. Due to that fact, each element must declare its length for storage, compression and transfer purposes. The length on the DICOM standard is always even. When the element's value is a single character string like patient sex data (0010,0040) that is either 'F' for female, 'M' for male or 'O' for other, the length of that element must be 2 and the value will be padded by a space character (ASCII: 0x20). String types (like CS and UI) are padded by space whereas binary and US types are padded by null (0x0).

CT ACQUISITION

The bases of the CT acquisition are better described on the Section 2.5.3.1 of this work. The mechanics and specific information will be summarily described here. A CT scanner is composed of an apparatus usually with a circular aperture in its center. The patient aperture is usually from 60 cm to 70 cm in diameter. Inside the covers of the CT scanner is a rotating frame which has an x-ray tube mounted on one side and the curved (i.e., "banana-shaped") detector mounted on the opposite side as shown in Figure A.2.



Figure A.2: A CT scanner apparatus.

The x-ray tube rotates around the patient as the following Figure A.3 shows. One or usually several of these rotations compose a single slice, depicted by a section of the object on the left of the image. The output image is generated by using the information taken by the detector on every single position of one cycle of rotation.



Figure A.3: A moment of the acquisition of a single slice that composes the whole CT scan.

LIBRARY

In order to access the DICOM data we have used the dcm4che library along with the dcm2xml implementation available at the dcm4che website. The dcm2xml implementation

takes a DICOM file as input and outputs a XML file containing all the DICOM information. The difference is that the XML file is directly readable (textual format), whereas the DICOM is not. Thus, in order to retrieve any information, we access the XML file, search for the wanted identification tag and gather the content stored on the related value field. The DICOM elements are distributed as exemplified in the Figure A.4 below.

```
1. <?xml version="1.0" encoding="UTF-8"?><dicom>
2. <attr tag="00020000" vr="UL" len="4">212</attr>
3. <attr tag="00020001" vr="0B" len="2">>00\01</attr>
4. <attr tag="00020002" vr="UI" len="26">>1.2.840.10008.5.1.4.1.1.2</attr>
5. <attr tag="00020010" vr="UI" len="18">1.2.840.10008.1.2</attr>
6. <attr tag="00020013" vr="SH" len="6">>OSIRIX</attr>
7. <attr tag="00020016" vr="AE" len="10">pacswebFIR</attr>
8. <attr tag="00020100" vr="UI" len="12">iMac-de-CDPI</attr>
9. <attr tag="00020100" vr="UI" len="10">ISO_IR 100</attr>
10. ...
11.</dicom>
```

Figure A.4: A fragment of a DICOM converted to a XML file using the dcm2xml implementation.

CONVERTING RAW DATA TO HOUSFIELD UNITS

Occasionally, pixel values on DICOM files, such as in CT images, are not stored directly on the Hounsfield scale (more details in Section 2.5.3.1). The DICOM standard allows the possibility of storing values of a conversion (e.g., such as a linear function) instead of the raw values (Hounsfield units in this case), along with the method of the conversion. There is, to the extent of our knowledge, two types of transformations: (1) a linear transformation or (2) a modality look up table. All the patients we have worked with had their data stored as a linear transformation, and we presume that this is always true for CT data. Due to that fact, we will intentionally disregard the modality look up table conversion and address just the linear one. The absence of the Modality LUT tag (0028,3000) on the DICOM file ensures that the values are stored as a linear transformation.

The Pixel Data tag (7FE0,0010) contains the actual values of the pixels stored as a linear vector. For instance, the Pixel Data tag on the XML file looks like the piece of code shown in Figure A.5 below.

1. <attr tag="7FE00010" vr="OW" len="524288">35\41\44\42\45\52\53\43\33\26\33\45\55\54\46\41\42\41\43\46\46\ 45\40\46\51\48\43\42\47\45\46\47\45\39\39\47\56\57\55\59\66\77\84\84\74\70\64 \60\63\71\77\78\74\70\68\74\79\75\76\84\97\104\103\97\91\91\93\93\93\100\104\ 101\99\104\108\106\97\87\80\80\86\91\89\85\82\84\87\90\78\62\59\65\72\77\77\ 67\56\50\52\58\60\61\64\64\62\57\56\57\59\63\64\59\52\49\45\41\39\36\41\52\59 \60\64\71\73\62\45\35\33\32\38\46\50\46\40\38\36\26\17\16\25\35\40\41\37\35\3 9\50\55\46\28\20\26\36\40\40\37\32\27\25\22\23\25\26\33\39\38\35\35\39\46\ 46\42\42\43\39\32\32\39\45\39\26\18\18\17\15\19\30\36\38\42\47\50\48\...</attr>

Figure A.5: A fragment of a pixel data attribute tag.

In order to reconstruct an image with the dimensions of the original one from the values stored on the Pixel Data tag we will call each value of the vector P(i) or P(x, y), where x equals the rest of the division of i from the width of the image (x = i % w) and y equals the modulus of the division of i by the width of the image ($y = \left|\frac{i}{w}\right|$). The width equals the number of columns on the image and can be obtained by accessing the Columns tag (0028,0010).

For every P(i) there must be two conversions until the returned value corresponds to the raw value of the pixel in HU. The first conversion is associated to the Pixel Representation tag (0028,0103). When this tag is equal to 1 it means that every value P(i) is stored as a two's complement and, hence, it must be converted back from this notation. If the tag is equal to 0, it means that the data is stored as unsigned and the value P(i) can be accessed directly. The Bits Stored tag (0028,0101) is also important for containing the amount of bits allocated to store each pixel value. In order to simplify the explanation, we will call the reconverted P(i)from the two's complement notation as $P_2(i)$ and depict the process on the following equation shown in Algorithm A.1, where \neg inverts the binary number, r stands for the pixel representation tag value (0 or 1), and b for the amount of allocated bits.

$$P_{2}(i) = \begin{cases} P(i) & \text{, if the leftmost (out of b) bit of } P(i) = 0 \\ \neg (P(i) + 1) & \text{, otherwise} \end{cases}$$
, if $\neg r = 0$
P(i) , otherwise

Algorithm A.1: Conversion from complement of two.

We have already made the first conversion. The second conversion is the linear one. After obtaining $P_2(i)$, two tags from the DICOM should be accessed: (1) the Rescale Slope (0028,1053) and the Rescale Intercept (0028,1052), which represent the angular and linear coefficient of the linear function, respectively. The corresponding Hounsfield value, represented by H(i) is further obtained by simply solving the Equation (A.1) disposed below.

$$H(i) = P_2(i) * Rescale Slope + Rescale Intercept$$
(A.1)

DISPLAYING IMAGES FROM HOUNSFIELD UNITS

Every DICOM file viewer internally implements the windowed view to display DICOMacquired images on the screen. Some provide the option of varying the window range and some do not. Moreover, some possess a function that paints the pixels within a given HU range (ImageJ and Osirix). Nevertheless, this painting do not preserve the texture-related information, just the spatial disposition of the colored pixels. Although all DICOM viewers display windowed data we did not found an option to export to images a given window in any of the following DICOM editors: RadiAnt Viewer, Invesalius, ImageJ and Osirix. Due to this fact, we implemented this feature in a small framework following the instructions below.

By assuming that the inferior limit of the image is always 0, we define that H_i and H_s represent the inferior and superior limit of the Hounsfield scale, respectively. O_i and O_s represent the inferior and superior limit of the output image O, also respectively, and that $\Delta H = H_s - H_i$ as well as $\Delta O = O_s - O_i$. Finally, we conceive a function O(i) that, from the corresponding value H(i), converts it into a value within the respective output image range as shown in Algorithm A.2.

$$O(i) = \begin{cases} round\left(\frac{\Delta O(H(i) - H_s)}{\Delta H} + O_s\right), & if \ H(i) \in [O_i, O_s] \\ 0 & , otherwise \end{cases}$$

Algorithm A.2: Function responsible for converting the values of a Hounsfield range $[H_i, H_s]$ to an image range $[O_i, O_s]$. In that case, the 0 value on the output image will be considered background as well as the H_i value. If not desired, that matching between the background and H_i can be avoided by setting the O_s variable to $O_s - 1$ and by summing 1 to the O(i) function. By doing that the value 0 will only be associated to the background only.

DISTINCT THRESHOLDS OF HOUNSFIELD RANGES

The following Figure A.6 depicts the difference of a single cardiac CT slice drawn to a 8-bit depth image when varying the Hounsfield range (H_i and H_s). Besides, every pixel that was out of the given interval [H_i , H_s] was being painted black.





[-200,500]



[-400,1000]

[-200,1000]



[-500,1500]



[-200,300]

[-500,500]

Figure A.6: Distinct thresholding of HU ranges.

The images we have been using on this work look like the Figure A.7, which stands within the [-200,-30] range. This range, in turn, corresponds to the Hounsfield range of the adipose tissues.



Figure A.7: Adipose tissue thresholding.

If we paint the information that is higher than H_s as white, even if regarding a smaller ΔH , then we have a more continuous image (i.e., without a sudden variation from white to black) like the one shown in the Figure A.8 below.



Figure A.8: Values higher than H_s being painted as white instead of black.

APPENDIX B — RESCALING IMAGES ACCORDING TO DICOM

The DICOM standard stores information about the size of the object acquired on 3 dimensions. This rescaling information should be conceived as the distance that pixels or voxels within the Pixel Data tag (7FE0,0010) should be drawn apart from each other when rendered on the screen. The Pixel Spacing tag (0028,0030) stores that distance on the x and y axes in mm, respectively. Moreover, the Slice Spacing tag (0018,0088), if present, subtracted by the Slice Thickness tag (0018,0050) stores the distance between the slices (i.e., on the z-axis), also in mm. Therefore, when rendered to the screen, every possible pair of pixels p_1 and p_2 should respect the following *distance* in Equation (A.2), where v is a three dimensional volume vector that has its x equals to the Pixel Data tag first element, y to the second element of the Pixel Data tag and z to the Slice Spacing tag element.

$$\Delta p = (p_1 - p_2)v$$

$$distance = \sqrt[2]{\Delta p. x^2 + \Delta p. y^2 + \Delta p. z^2}$$
(A.2)

In order to output the data of a single slice to a single image we just need the information of the Pixel Data tag. By manual rescaling and analysis we have defined that a good value for the Pixel Spacing on cardiac CT data was 0.35. That value was chosen when avoiding the output image of being cropped or of being too small (loosing too much information) after rescaled. Thus, we map the pixel data by applying the following rescaling function in Equation (A.3), where v_a represents the slice's actual Pixel Spacing, v_d the desired scaling we want to achieve, x and y the positions of the pixels on the slice's image and x' and y' the mapped rescaled image.

$$\begin{bmatrix} x'\\ y' \end{bmatrix} = \begin{bmatrix} \frac{v_a \cdot x}{v_d \cdot x} & 0\\ 0 & \frac{v_a \cdot y}{v_d \cdot y} \end{bmatrix} \begin{bmatrix} x\\ y \end{bmatrix}$$
(A.3)

It is not sufficient to just rescale the slice, the image should be centered after or before the rescalement. There are basically two ways of proceeding on this: (1) before being rescaled the image should be translated to the origin or (2) after being rescaled the image should be translated by the difference of the bottom-right pixel position after rescaling to the one before rescaling.

In this work, we have also observed that, even when rescaling the images to a common scale, there is still a noticeable difference on the same (i.e., in a sense that some images are actually not within the same scale even after being standardized). Further searches into that topic defined what may be causing that variation: some devices automatically make assumptions about the geometry magnification of the data being acquired. In many cases, without documenting the nature of the correction, what triggers it (e.g., body part), and how to suppress it [A.1]. Therefore, that induces a general variation on the spatial measures of the DICOM standard.

APPENDIX C — AUTOMATICALLY SEGMENTED SLICES

The following images of a single patient represent slices automatically segmented by our approach. All of them were segmented by the same predictive model. The slices on the left column are the ones that were automatically segmented, whereas the slices on the right represent the ground truth of the slice at its left.

SEGMENTED PATIENT































Figure A.9: Automatically segmented patient and its ground truth.