Cláudia Santos Constantino

Bachelor of Science

Reproducibility Study of Tumor Biomarkers Extracted from Positron Emission Tomography Images with ^{18}F-Fluorodeoxyglucose

Dissertation submitted in partial fulfillment
of the requirements for the degree of

Master of Science in
Biomedical Engineering

Advisers: Francisco Oliveira, Researcher at Nuclear Medicine-Radiopharmacology, Champalimaud Foundation

Pedro Vieira, Assistant Professor, NOVA University of Lisbon

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE **NOVA** DE LISBOA

September, 2019

Reproducibility Study of Tumor Biomarkers Extracted from Positron Emission Tomography Images with ¹⁸F-Fluorodeoxyglucose

To my grandfather António

ACKNOWLEDGMENTS

On this page I want to thank to all those who left a mark in these five years journey of my life, whether professionally or personally.

First, I want to thank to Champalimaud Foundation for having me welcomed in the institution and for providing me the chance to realize my master thesis here.

Also, I would like to express my sincere gratitude to Professor Doctor Durval C. Costa for the opportunity to develop this fascinating project in the group of Nuclear Medicine-Radiopharmacology. Thank you for the help, continuous motivation and for the valuable transmitted knowledge. I will always be grateful.

To my adviser Doctor Francisco Oliveira, who has been an inspiration and role model in science since the beginning, I would like to thank him for the daily support, kindness, patience, and friendship. In addition to the unending knowledge learned from him, I found his passion for scientific research highly contagious. Thank you for all the incentives, knowledge transmission and guidance throughout the entire project.

To all my group of colleagues here, at the Champalimaud Foundation, I have to say that you made my days happier. I finish this journey with a lot of good and funny moments to remember.

To my longtime friends, thank you for the fundamental friendship, for the moments of true happiness, genuine laughs, and for being the listeners I needed in the hardest days.

To my partner, to whom I am very grateful for all the patience and care given to me at all times. Your friendship and love were essential and made me grow up a lot as person.

Finally, with all my heart, I want to thank to my family, my parents Célia, Silvino, sisters Alexandra and Lara and grandmother Maria, for always being my safe haven. Thank you for the support and encouragement that has permanently been provided to me. As you always say mom: "Com dedicação e trabalho árduo tudo é possível" ("dedication and compromise to hard work makes it all possible!").

ABSTRACT

Introduction and aim Cancer is one of the main causes of death worldwide. Tumor diagnosis, staging, surveillance, prognosis and access to the response to therapy are critical when it comes to plan and analyze the optimal treatment strategies of cancer diseases. ^{18}F-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography (PET) imaging has provided some reliable prognostic factors in several cancer types, by extracting quantitative measures from the images obtained in clinics. The recent addition of digital equipment to the clinical armamentarium of PET leads to some concerns regarding inter-device data variability. Consequently, the reproducibility assessment of the tumor features, usually used in clinics and research, extracted from images acquired in an analog and new digital PET equipment is of paramount importance for use of multi-scanner studies in longitudinal patient's studies. The aim of this study was to evaluate the inter-equipment reliability of a set of 25 lesional features commonly used in clinics and research.

Material and methods In order to access the features agreement, a dual imaging protocol was designed. Whole-body ^{18}F-FDG PET images from 53 oncological patients were acquired, after a single ^{18}F-FDG injection, with two devices alternatively: Philips Vereos Digital PET/CT (VE-REOS with three different reconstruction protocols- *digital)* and Philips GEMINI TF-16 (GEM-INI with single standard reconstruction protocol- *analog*). A nuclear medicine physician identified 283 ^{18}F-FDG avid lesions. Then, all lesions (both equipment) were automatically segmented based on a Bayesian classifier optimized to this study. In the total, 25 features (first order statistics and geometric features) were computed and compared. The intraclass correlation coefficient (ICC) was used as measure of agreement.

Results A high agreement (ICC > 0.75) was obtained for most of the lesion features pulled out from both devices imaging data, for all (GEMINI vs VEREOS) reconstructions. The lesion features most frequently used, maximum standardized uptake value, metabolic tumor volume, and total lesion glycolysis reached maximum ICC of 0.90, 0.98 and 0.97, respectively.

Conclusions Under controlled acquisition and reconstruction parameters, most of the features studied can be used for research and clinical work, whenever multiple scanner (e.g. VEREOS and GEMINI) studies, mainly during longitudinal patient evaluation, are used.

Keywords: Nuclear Medicine, Positron Emission Tomography, Lesions Quantification, Segmentation, Image Processing

RESUMO

Introdução e objetivo O cancro é uma das principais causas de morte no mundo. O seu diagnóstico, estadiamento, vigilância, prognóstico e acesso da resposta à terapia é essencial no planeamento e adoção da melhor estratégia de tratamento a seguir. Tomografia por emissão de positrões (PET) com ^{18}F-fluordeoxiglicose (^{18}F-FDG) tem demonstrado fornecer fatores de prognóstico fiáveis para diferentes tipos de cancro, a partir da extração de medidas quantitativas de imagens PET. A recente introdução no mercado de um PET digital traz algumas preocupações quanto à variabilidade dos dados entre equipamentos. Consequentemente, um estudo da reprodutibilidade das características lesionais frequentemente usadas em investigação e em clínica, extraídas de imagens PET de um equipamento analógico e do novo digital, é imprescindível para o uso de estudos *multi-scanner* em avaliações longitudinais de doentes. O objetivo deste estudo é avaliar a reprodutibilidade, entre equipamentos, de 25 características lesionais frequentemente usadas na clínica e investigação.

Materiais e métodos Para aceder à concordância das características, um protocolo de dupla aquisição foi implementado na clínica. Foram adquiridas imagens PET de corpo inteiro de 53 doentes oncológicos, após uma única injeção de ^{18}F-FDG, em dois equipamentos alternadamente: Philips Vereos Digital PET/CT (VEREOS – digital) com três protocolos de reconstrução diferentes e Philips GEMINI TF-16 (GEMINI - analógico) com um protocolo de reconstrução padrão. Um médico especialista em medicina nuclear identificou 283 lesões ávidas para ^{18}F-FDG. De seguida, todas as lesões foram segmentadas automaticamente por um algoritmo de classificação Bayesiana otimizado para este estudo. No total, 25 características (baseadas em estatísticas de primeira ordem e em geometria) foram calculadas e comparadas. O coeficiente de correlação intraclasse (ICC) foi usado para medir a concordância entre os valores das características correspondentes extraídas das imagens dos dois equipamentos.

Resultados Foi obtida uma elevada concordância (ICC > 0.75) entre a maioria das características retiradas das imagens dos dois equipamentos, para todas as reconstruções (GEMINI versus VEREOS). Para as características lesionais usadas mais frequentemente, valor máximo de captação padronizada, volume metabólico tumoral e glicólise total da lesão foi obtido um ICC máximo de 0,90, 0,98 e 0,97, respetivamente.

Conclusões Com parâmetros de aquisição e reconstrução controlados, a maioria das características estudadas podem ser usadas em investigação e na prática clínica em estudos que envolvam aquisições em ambos os equipamentos (VEREOS e GEMINI), em particular na avaliação longitudinal do doente.

Palavras-chave: Medicina Nuclear, Tomografia por Emissão de Positrões, Quantificação de Lesões, Segmentação, Processamento de Imagem.

PUBLICATIONS

This dissertation includes work reported in different abstracts accepted or submitted to international and national conferences:

Communication in international conferences

- C. S. Constantino, F. P. M. Oliveira, M. Silva, C. Oliveira, J. C. Castanheira, Â. Silva, S. C. Vaz, P. Vieira, D. C. Costa; "Reproducibility Study of Lesion Features Extracted from ^{18}F-FDG PET Images of the Same Patients Acquired on Two Philips PET/CT Scanners: Digital VEREOS versus GEMINI TF", EANM'19, Barcelona, Spain (2019). *(accepted)*

Communication in national conferences

- C. S. Constantino, F. P. M. Oliveira, M. Silva, C. Oliveira, J. C. Castanheira, Â. Silva, S. C. Vaz, P. Vieira, D. C. Costa; "Segmentação automática de lesões ávidas para ^{1}FDG em PET/CT baseada numa técnica de *clustering* bayesiana", XVII CNMN, Porto, Portugal (2019). *(submitted)*

- C. S. Constantino, F. P. M. Oliveira, M. Silva, C. Oliveira, J. C. Castanheira, Â. Silva, S. C. Vaz, P. Vieira, D. C. Costa; "Extração de características de lesões ávidas para FDG: estudo da reprodutibilidade entre Philips Vereos Digital e GEMINI TF16 PET/CT", XVII CNMN, Porto, Portugal (2019). *(submitted)*

As an output of this thesis, an article is also being prepared to be published during this year.

CONTENTS

xvi

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

18F-FDG   18F-Fluorodeoxyglucose

PET   Positron Emission Tomography

CT   Computed Tomography

NM   Nuclear Medicine

PMT   Photomultiplier tubes

DPC   Digital Photon Counter

EANM   European Association of Nuclear Medicine

EARL   EANM Research Ltd

ROI   Region-of-Interest

MTV   Metabolic Tumor Volume

TLG   Total Lesion Glycolysis

LOR   Line-of-Response

TOF   Time-of-Flight

EM   Expectation Maximization

FCM   Fuzzy C-means

FLAB   Fuzzy Locally Adaptive Bayesian

CCC   Champalimaud Clinical Centre

FOS   First Order Statistics

SA   Surface Area

AFOV   Axial Field Of View

OSEM   Ordered-Subset Expectation Maximization

FWHM   Full Width at Half Maximum

NIFTI   Neuroimaging Informatics Technology Initiative

SUV   Standardized Uptake Value

ICC   Intraclass Correlation Coefficient

IQR   Interquartile Range

INTRODUCTION

1.1 Tumor ^{18}F-fluorodeoxyglucose avidity

The ^{18}F-fluorodeoxyglucose (^{18}F-FDG) is a radiopharmaceutical labeled with the radioisotope fluorine-18 generated in a cyclotron that produces positrons and has a brief half-life (109.7 min). ^{18}F-FDG is a glucose analogue and is absorbed via living cells by cell membrane glucose transporters. Then, it is incorporated into the normal glycolytic pathway.

In great number of tumors, the consumption of glucose increases mostly due to the overexpression of the facilitated diffusion glucose transporters and augmented hexokinase activity [1]. So, these tumor cells are highly metabolic active and favor more inefficient anaerobic pathway adding to the already increased glucose demands [2]. These mechanisms together allow tumor cells to absorb and retain high levels of ^{18}F-FDG when compared with normal tissues. Its accumulation in tissue is proportional to the quantity of glucose utilization. The uptake of ^{18}F-FDG is substantially increased in most tumors as is the case of esophageal, ovarian, stomach, head and neck, lung, cervical, colorectal and breast cancers, as well as melanoma and most types of lymphoma [3]. It is important to be aware that ^{18}F-FDG will accumulate in areas with high levels of metabolism and glycolysis, and not just in tumor cells. Increased uptake is probable to appear in sites of hyperactivity, as muscular or nervous, active inflammation, tissue repair, and others. For example, kidneys and urinary bladder are common physiologic uptake areas of ^{18}F-FDG since they are excretion organs.

In order to follow the metabolic activity using this radiopharmaceutical, it is required patient injection. The injection of this biologically important material should fulfill the ALARA

principle for minimizing radiation exposure [4]. The injected dose should be the lower possible for the patient ensuring a reasonable image quality in which tracer uptake in target structures (lesions tissue) are discernible.

1.2 Context and motivation

Tumor diagnosis, staging, surveillance, prognosis and access to the response to therapy are critical when it comes to plan the optimal treatment strategies of malignant diseases. For tumor diagnosis, precise tumor location and description of the form, intensity, and heterogeneity may be of paramount importance.

The anatomy and functionality of the human body have been subject of many studies and can be mapped in a non-invasive way by positron emission tomography (PET) integrated with X-ray computed tomography (CT), facilitating malignancies diagnosis, treatment planning and monitoring. In particular, PET molecular imaging reveals biological processes, providing insight into tumor metabolism and its effects on tissue function. CT provides valuable anatomic information about the tumor and its involvement with adjacent organs and vasculature. The lack of a high spatial resolution of PET equipment is compensated with the integration of CT equipment, the easiest and highest resolution tomographic modality [5]. The junction of the two brings the best of both worlds for Nuclear Medicine (NM).

Nowadays, analog PET/CT with the radiopharmaceutical ^{18}F-FDG is widely used daily in clinics for assessing response to therapy in clinical trials and clinical practice for patients with tumors. Recent developments by Philips Healthcare created the first truly digital PET equipment, now available in Portugal only at Champalimaud Foundation, the Philips Vereos Digital PET/CT. This equipment has higher temporal and spatial resolution than any analog PET/CT commonly used in NM clinics. Therefore, additional precision and accuracy are provided by this equipment in malignance's evaluation. Besides the Philips Vereos, a Philips GEMINI TF16 (analog PET/CT) is also installed in NM of Champalimaud Foundation. Images from both these scanners are going to be used in this study.

The recent addition of the digital equipment to the clinical armamentarium of PET/CT leads to some concerns regarding inter-device variability. Reproducibility or robustness is an indispensable requirement for any quantitative measurement and imaging biomarker. Adequate reproducibility ensures the capability to produce identical results in the same patient when examined on different systems, which are crucial for the clinic management of subjects and for the use of ^{18}F-FDG PET/CT images within multicenter trials [4]. Thus, a reproducibility study on tumor imaging biomarkers (features) involving the two available equipment is imperative.

1.3 Objectives and dissertation plan

The primary objective of the dissertation is the reproducibility assessment of lesion features usually used in clinics or research, extracted from ^{18}F-FDG PET images (figure 1.1). The features to be tested will be computed based on the images acquired by an analog PET/CT -GEMINI- and by a digital PET/CT -VEREOS-. If there is a correlation but low agreement between the features extracted from the images of both equipment, conversion factors will be estimated.

The present dissertation has been structured to first cover the current state-of-the-art concerning PET/CT equipment, ^{18}F-FDG as a radiopharmaceutical daily used in nuclear medicine and methods to delineate lesion volume used in clinical studies or research. In chapter 2 the relevant theoretical underpinnings of PET and CT essential for the understanding of how the medical image is acquired are described. Afterward, in chapter 3, some common segmentation techniques and their possible advantages and disadvantages are described. In chapter 4, the materials and methods used for imaging processing and statistical analysis are presented. In the following chapter, the results are accessible, both for the segmentation and features extraction and comparison. Lastly, the discussion and final conclusions are addressed.

Patient

Philips Vereos Digital PET/CT

Philips GEMINI TF PET/CT

Whole Body PET/CT acquisition

Lesions identification

Automatic Segmentation

Features extraction

Features from VEREOS images

Statistical analysis

Features from GEMINI images
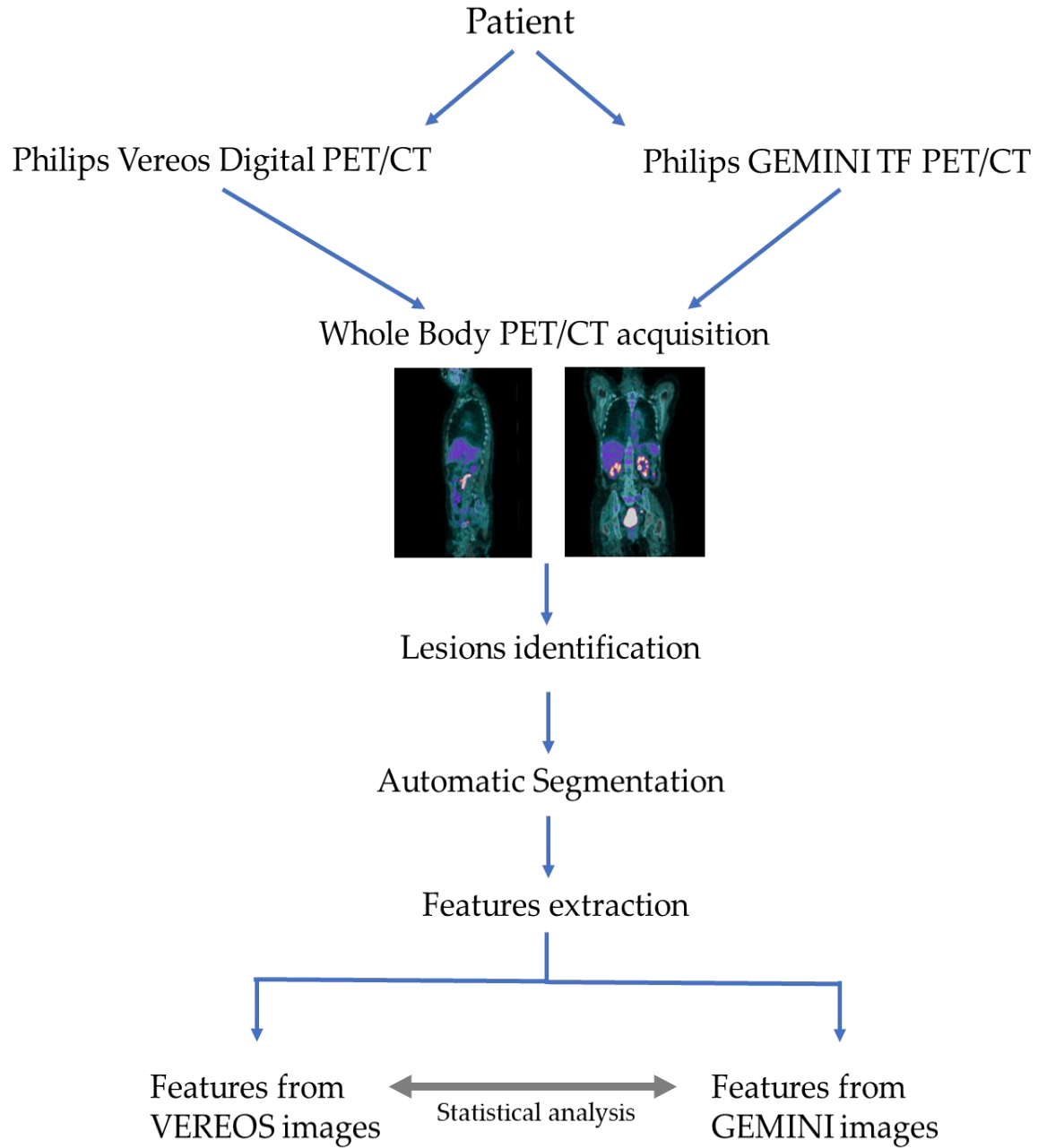
Figure 1.1: Flowchart representing the necessary steps for achieving the final goal: reproducibility assessment of lesions features usually used in clinics and research extracted from ^{18}F-FDG PET images. After images acquisition in both scanners, the lesions are identified by a physician and then automatically segmented. Therefore, the features are computed. Lastly, statistical analysis between the features is completed in order to find the reproducible features between equipments.

1.4 State-of-the-Art

NM emerged in the early 20th century due to the need for understanding the physiologic and biologic mechanisms of health and disease. An expansion of NM applications both in clinical studies and research has occurred since then. Advances in radiopharmaceutical development, instrumentation and computer processing led to the implementation of PET for clinical studies, and improved treatments with radiopharmaceuticals particularly in cancer patients.

Later, integrating data processing streams from PET and CT was considered and by 1998, the first PET/CT prototype was ready for use. The fusion of molecular/metabolic and anatomical/morphological information improved the diagnostic accuracy in the identification and characterization of malignancies, assessment of tumor stage, therapeutic response and tumor recurrence [6]. PET/CT combines two already excellent modalities [7], [8]. Its main advantages compared to the simple PET system are the possibility to apply image reconstruction methods that compensate for the photon attenuation in the tissues and anatomic referencing.

Nowadays, the majority of PET/CT equipment daily used have PET detectors based on photomultiplier tubes (PMT) with the disadvantages of limited photon-to-electron quantum conversion efficiency, bulkiness, and the relatively high cost of PMTs [9], [10]. To overcome the technical limitations, a novel digital photon counter (DPC) PET/CT was recently created and is now installed in a few clinics around the world. Its prototype system showed better image quality and diagnostic accuracy compared with the previously used analog PET/CT [11]. In the analog equipment, available around 2001, the analog output signal requires off-chip processing. The digital PET/CT works without analog signal processing because each single photon avalanche diode runs as a digital counter that delivers a direct count of the number of scintillation photons perceived.

In order to advance nuclear medicine research and multicenter studies, the European Association of Nuclear Medicine (EANM) launched EANM research Ltd (EARL) specifications [12]. These instructions were created considering analog devices, as GEMINI. Digital PET system has improved spatial and temporal resolution and originates activity concentration recovery coefficients in the NEMA tests above EARL specifications [13], [14]. So, using the standard manufacture reconstruction protocols, VEREOS scanner does not fulfill EARL requirements. In order to satisfy these specifications, VEREOS images have to be smoothed [14].

To quantify lesions and tumors uptake patterns is essential that the features used are reproducible between scanners. To our knowledge, at the start of this study, there were no reproducibility studies comparing the tumor features extracted from VEREOS and GEMINI devices with the radiopharmaceutical ^{18}F-FDG.

To evaluate the features reproducibility, medical image processing and analysis are crucial. Medical image analysis became important in the early 1990s, emerging from Artificial Intelligence and Computer Science. In the next years, since the expansion of imaging techniques, there was a need to delineate the regions of interest (ROI) of the human body images, in order to quantify the volume of tumors and study their structure. Consequently, studies developing segmentation algorithms were rapidly released, since the segmentation process is an essential step for tumor features extraction. Segmentation subdivides a digital image into sets of voxels, categorizing them in biological meaningful labels. Manual segmentation by an expert, slice by slice, is time-consuming and susceptible to the natural intra and inter-observer variability. So, it is not a good solution for large datasets [15] and automatic algorithms may be preferable. Numerous studies have used different segmentation methods to define tumor volume by automatic or semi-automatic segmentation techniques [16]–[18]. Several automatic algorithms have been developed, removing bias from the process. Appropriate segmentation methods may differ from imaging modalities, and accuracy and precision are of great importance. When designing an effective algorithm, the imaging modality, the structures to analyze, the influence of noise and partial volume effects should be considered [19]. For example, methods based on thresholding are commonly applied due to its simplicity and quickness [20]–[23]. Bayesian approaches have been recently developed and exhibit consistent performance and fewer errors when compared with other methods [24]. Answering to different problems in nuclear medicine, several algorithms for medical image segmentation were then created over the years. PET/CT segmentation has been applied in the lung [25], [26], prostate [27], esophagus [28], brain [29], heart [30], etc. Consequently, the accuracy and reproducibility of different algorithms on PET images segmentation must also be evaluated.

Quantitative analysis has shown to be more effective than visual assessment during the course of therapy, by distinguishing effective from ineffective treatment earlier [31]. The extraction of a large number of features from images has become popular under the denomination *radiomics* [15]. This kind of analysis may provide crucial information about tumor phenotypes. For example, tumoral heterogeneity may be useful for prognostic in several tumors [28], [32], [33], once malignant tumors are susceptible to be heterogeneous. Several research studies have considered feature extraction a crucial step to tumor diagnostic and prognostic. For example, Chan et al. found that tumor heterogeneity categorized by texture features was superior to traditional PET parameters for predicting outcomes in nasopharyngeal carcinoma [34]. Quantitative features may serve as image-based biomarkers with the potential to diagnosis and prognosis. For instance, metabolic tumor volume (MTV) and total lesion glycolysis (TLG) obtained from ^{18}F-FDG PET/CT are valuable for predicting treatment response in some tumors, as lung cancer [35]. So, after a

correct segmentation, features associated with tumor intensity and tumor geometry must be extracted from the images acquired in both scanners and compared using adequate statistics.

POSITRON EMISSION TOMOGRAPHY AND X-RAY COMPUTED TOMOGRAPHY

2.1 Positron emission tomography principles and concepts

PET is a biomedical imaging technique that makes available functional information about the human body and organs in a non-invasive quantitative assessment. PET bio-functional potentiality provides detection of some tumors that are not detected through other techniques. Another important application is neuroimaging. For example, PET can help to discriminate Alzheimer disease from other neurodegenerative diseases [36].

In PET images is possible to determine the amount of the radiopharmaceutical that is present in a specific ROI, since the counting rate of the respective reconstructed image is directly proportional to the radioactivity concentration. Common PET scanners are constituted by multiple rings of scintillation crystal detectors coupled to PMT, surrounding the patient.

Positrons are emitted via β^+ decay with continuous kinetic energy distribution. As represented in equation 2.1, the radioisotope fluorine-18 originates an oxygen atom, a positron (e^+) and an electron neutrino (ν).

$$^{18}_{9}F \rightarrow {}^{18}_{8}O + e^+ + \nu \tag{2.1}$$

Positrons have a very short time of life and distance range in condensed matter. They trans-
fer rapidly their energy undergoing ionizing events, collision interactions and excitation of mo-
lecular species and eventually thermalize before annihilation with an electron [37]. After positron
emission from ^{18}F radioisotope, the positron range in the human tissue is about 0.5 mm to 2 mm
[37]. When positrons have energy sufficiently low, and they meet an electron (positron antiparti-
cle), occurs annihilation, i.e., their mass is converted into radiation energy and two photons ap-
proximately collinear are emitted in opposite directions, as seen in figure 2.1(A). The non-collin-
earity is due to the conservation of linear momentum and energy. Consequentially, an angular
deviation in the biological tissue of ±0.25º is expected [38].
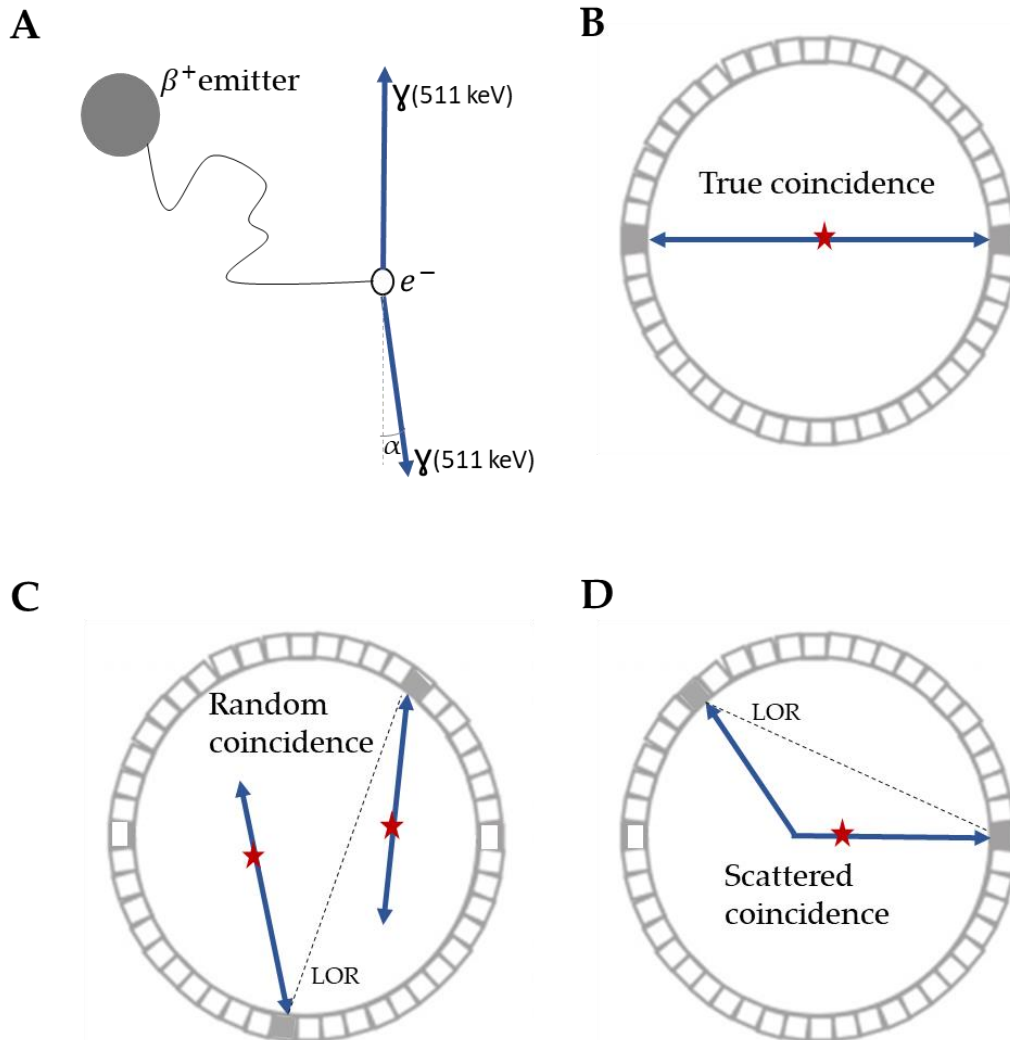
Figure 2.1: (A) β^+ particle slowing-down and its annihilation. (B) True coincidence. (C) Random
coincidence. (D) Scattered event.

The image acquisition inside the patient's body requires coincidence detection of both the 511 keV annihilation gamma rays within a resolving time. So, the PET detection system is founded on the two annihilation photons which are released in opposite directions. The fact that positrons travel a short distance before its annihilation causes a limitation of PET spatial resolution.

In Figure 2.1(B-D), true coincidence and two undesired events are represented. They can be perceived by the detectors through the coincidence technique. Scattered event coincidence can happen when one of both annihilation photons experiences a Compton interaction inside the body, which causes a change in direction and so the event is given by a wrong line of response (LOR). Random events are another possibility when two photons from different annihilation phenomena are detected as coincident. Random coincidence and Compton scattering lead to image blurring, loss in image contrast and degrade the accuracy of the quantitative analysis [39].

The spatial resolution of PET equipment is the minimum distance at which two radioactive sources can be placed so that they can be distinguishable in the image. The limited spatial resolution of the scintillation crystal is one of the factors that contribute to the degradation of spatial resolution, due to crystal dimensions. Another reason is the deviation of the two photons produced in the positron-electron annihilation from collinearity.

The time-of-flight (TOF) PET allows estimating the position of the annihilation events through the time difference of two interactions on the two detectors along a LOR. So, the position of annihilation is determined by the LOR and by the difference in time arrival of both photons on the two coincidence detectors. The TOF resolution depends on some factors, such as the scintillator material. The faster is the scintillation material, the shorter is the time resolution.

2.2 X-ray computed tomography principles and concepts

CT is one of the most used imaging methods in medicine. Nowadays, CT offers substantial anatomic information about the number, size, and locations of tumors. This technique uses radiation, or X-rays, joined with an electronic detector to record a pattern of densities. The X-ray beam passing from multiple projections rotates around the patient. An electronic detector array detects the X-rays that pass through the patient, measuring their attenuation. The emitter of X-rays rotates as the detector, diametrically on the opposite side. This data (attenuation and location) is then used to reconstruct a tomographic (3D) image. CT is based on the principle that the density of the tissue passed by the X-ray beam is calculated by the attenuation coefficient.

Depending on the type of tissue through which the X-rays beam pass, the attenuation tends to be different. The higher the attenuation on the tissue, the brighter the tissue appears in the images, and the opposite is equally valid.

2.3   Positron emission tomography-computed tomography

PET/CT imaging combines molecular-metabolic precision of PET with anatomic details of CT, constituting a reliable tool in diagnostic oncology, tumor staging, and treatment monitoring and planning. This equipment also allows acquisition of PET only and CT only images. CT may also be used to quantify tumor sizes and tumor involvement with vital organs and vasculature. PET image provides additional data on the biological or metabolic heterogeneity of tumors, as hypoxic regions within a tumor, apoptosis, and necrosis [40].

Several PET/CT equipment with similar characteristics exist and are used in the clinical environment, mostly analog PET/CT. Digital PET/CT is very recent and is equipped with DPC, an innovative scintillation detector. Analog PET/CT uses conventional photomultipliers coupled with crystal scintillators. Digital PET/CT inserts an array of silicon photomultipliers (SiPM) instead of the conventional analog photomultipliers. These digital detectors present high-performance without the need for noise-sensitive off-chip analog to digital conversion [41]–[43]. The recent Philips Vereos Digital PET/CT equipment has better TOF resolution than the analog one, which also contributes to improve the spatial resolution. Thus, digital PET/CT affords better image quality, diagnostic confidence, as well as accuracy, compared with analog PET/CT [11].

IMAGE SEGMENTATION

Segmentation process divides an image into non-overlapping regions. Typically, a correct segmentation must group pixels in the same region if they have a similar gray level, color, texture, brightness or contrast [44]. This process aims to study anatomical or functional structure, identify ROI helping in the location of the tumor, lesion, and other abnormalities, measure tissue volume and help in treatment planning. Assuming that the domain of the medical image is given by Ω, the segmentation process should discover the sets $S_i \subset \Omega$ ($i \leq k$, where k is the number of classes). Consequently, the sets that make up the segmentation must verify equation 3.1,

$$\Omega = \bigcup_{i=1}^{k} S_i \tag{3.1}$$

where $S_i \cap S_j = \varnothing$ for $i \neq j$ [45]. A segmentation process identifies S_i sets that correspond to dissimilar anatomical or functional structures or ROI in the medical image. Segmentation can be applied through a manual, semi-automatic or automatic process.

The labeling process gives a meaningful classification to every category individually. Basically, it maps the numerical index i of S_i to an anatomical or functional term. Reliable algorithms are mandatory for the delineation of a ROI. According to algorithms principal methodologies, there are three general classifications: based on thresholds, gradients, clustering, and deformable models, for instance [19].

3.1 Algorithms based on thresholds

Threshold-based segmentation methods ensure that the tumor or organs with interest are distinguished based on image intensity. This method is one of the simplest and fastest, founded on the

13

supposition that the regions of interest have different grey levels than the surrounding regions. The threshold is an intensity value that splits the desired classes. So, a binary segmentation divides intensities into "background" (classified as 0) for the pixels with intensities less than the threshold (for instance), and "foreground" (classified as 1) for the pixels with intensities greater than or equal to the threshold. The following equation 3.2 describes the method for a binary segmentation of a 2D image,

$$g(x,y) = \begin{cases} \text{background if } f(x,y) < T \\ \text{foreground if } f(x,y) \geq T \end{cases} \tag{3.2}$$

where $f(x,y)$ is the pixel intensity in the (x,y) position and T is the defined threshold value. For multiclass segmentation, several thresholds must be defined. Threshold values may be found, for instance, in the histogram images, with the diverse peaks and valleys that enable the division of the images into different regions. In figure 3.1 three obvious thresholds at the valleys of the histogram are shown, which may be used to divide the image into two or three classes.
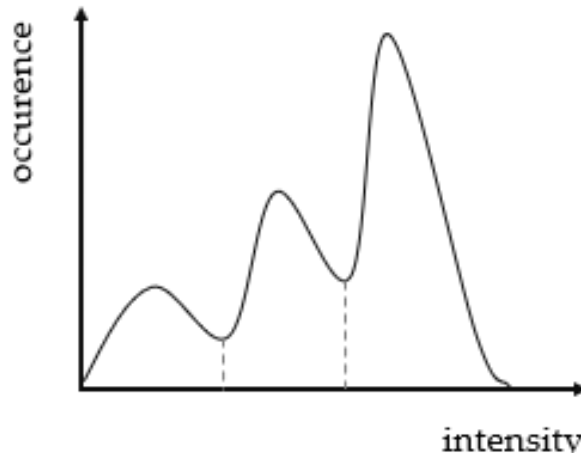
Figure 3.1: Histogram showing three apparent classes.

Thresholds used in these algorithms can be designated manually or automatically. The first way requires a trial experiment and *a priori* information to discover the correct values. A number of groups have used fixed threshold values derived from phantom validation work (generally between 50% and 80% of maximum local activity concentration value) [46], [47]. For example, Paulino et al [48] used a threshold of 50% in head and neck tumors. Automatic selection collects the image information to find the adaptative threshold values automatically. One of the techniques that have been developed is the Otsu's method [49], in which an initial guess of the threshold values is made, and then maximizes de separation between different threshold classes.

3.2   Algorithms based on gradients

Gradient based segmentation algorithms start by the computation of the gradient image. Then, different segmentation techniques can be used on these images, namely the threshold based. In the borders of the organs and tumors, an accentuated transition of the grey level (or uptake in the PET images) is commonly observed. Thus, threshold techniques on gradient images may allow detecting the edges of the structures of interest. This leads to a common classification of the segmentation algorithms in edge-based or region-based [19].

In edge-based algorithms, threshold values are associated with the edge information, and consequentially pixels are classified as edge or non-edge taking into account the filter output (a gradient filter). Pixels that are not divided by an edge are assigned to the same category. Laplacian edge detection belongs to this type of algorithms [50]. This method applies the second spatial derivative information of the voxel's intensity. Canny and Sobel edge detection are often applied [19], [51]. The first makes use of the gradient magnitude, discover the possible edge pixels and overpowers them with non-maximal suppression and hysteresis thresholding. Sobel filters identify and extract borders with gradient operators.

3.3   Algorithms based regions homogeneity

Region-based algorithms are based on the principle of homogeneity i.e., operate by grouping neighbors' pixels that have similar image features of interest (as intensity, for instance) and splitting them from the ones with dissimilar values. A boundary is formed from the differences between these two regions. Region growing algorithms are typical examples of this type, and integrates the use of seed points, manually selected by an operator or automatically set using another algorithm. Then, the neighboring pixels of the seeds with similar properties, within the threshold, are merged together. Seeded region growing process aims to enlarge a seed region through the integration of unallocated neighbor pixels with the minimum intensity difference comparatively to the seeded region. In order to remove the dependence on original seeds defined by the operator, *a priori* knowledge and statistics can be combined into the algorithms [52], [53]. The norms to decide the pixel connectivity implemented to determine neighbors depend on the algorithm used.

3.4   Algorithms based on clustering techniques

Clustering algorithms try to group the pixels into classes based on an adequate rule. Unsupervised methods do not use training data, they rely only on the image to be segmented [54]. Frequently,

these methods do not consider the spatial information, and therefore they are sensitive to noise and intensity inhomogeneities.

Three clustering algorithms will be explained in this section: K-means, the fuzzy C-means, and Bayesian based method.

K-means algorithms or also called CM algorithms [19] assign the unlabeled pixels to the nearest clusters, in which the measure of "distance" is the pixel intensity. Usually, it labels the pixels so that the within-cluster variance is minimized. The process is iterative, i.e., after a round, a new one starts, and the pixels may be reclassified until no improvement is achieved. Every time a pixel is reclassified, the centroid of each class (the average of the intensity of the pixels belonging to that cluster) needs to be updated. K is the number of clusters in which the image was divided.

The fuzzy C-means (FCM) algorithm permits soft segmentation based on fuzzy set theory. It generalizes k-means algorithm. As k-means, the FCM technique groups similar data in the same clusters based on the minimization of the within-cluster variance [55]. It computes the degree of belonging to each class, and do not classify the pixel into a static cluster as K-means.

Bayesian automatic algorithms allow noise modeling. Thus, when compared with other segmentation algorithms, they are less sensitive to noise due to their statistical modeling. In the image segmentation process, these algorithms provide an unsupervised estimation of the essential parameters which limit the classes selection in the image by the user.

The Bayesian method is an unsupervised classifier based on Bayes probability formula, represented in equation 3.3,

$$P(C_k|x) = \frac{P(x|C_k)\,P(C_k)}{P(x)} \quad , P(x) \neq 0 \tag{3.3}$$

where $P(C_k|x)$ is the posterior probability of the gray value x belong to the class C_k, $P(x|C_k)$ is the likelihood of x inside the class C_k, $P(C_k)$ is the prior probability of class C_k and $P(x)$ is probability of x. Each x is classified as belonging to class C_k if $P(C_k|x)$ is highest between all other classes. The classification result is independent of $P(x)$, because this value is constant for each x.

An example of an algorithm based on the Bayesian method is the fuzzy locally adaptive Bayesian (FLAB) segmentation. This approach has revealed better performance, principally for smaller objects, when compared with the threshold methods and FCM algorithms [24]. While the

Bayesian part of the algorithm measures the uncertainty of the classification, assuming that each voxel is identified but the observed data is noisy, the fuzzy part measures the imprecision of each voxel classification, with the hypothesis being that the respective voxel may coexist in two homogeneous (or "hard") classes [24]. In this method, instead of the standard implementation with a finite number of hard classes, a finite number of fuzzy levels in combination with two hard classes is implemented [56].

3.5   Algorithms based on deformable models

Deformable models have been extensively used in medical image segmentation. These algorithms define the boundaries of the ROI by applying closed parametric curves, or surfaces that deform under the influence of external and internal forces [57]. A closed curve or surface near the chosen boundary should be introduced in an image. Then, the external forces resulted from the image drive the curve to the desired contour of the region. The internal forces effort to keep the deformation smooth. Figure 3.2(A) demonstrates an example of applying an active contour in the bladder, initialized as a circle. In figure 3.2(B), the active contour is permitted to deform to the inner boundary of the bladder.
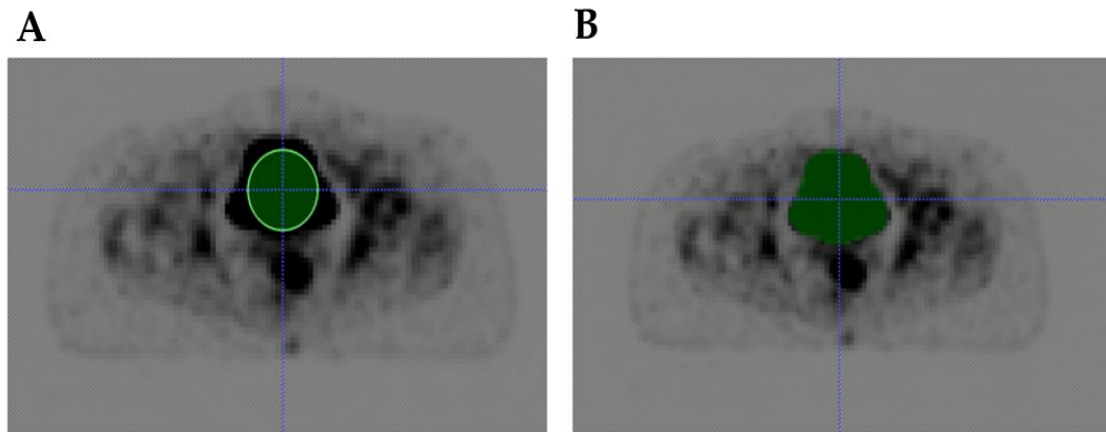
Figure 3.2: Extraction of the inner wall of the bladder from ^{18}F-FDG PET image using deformable models. Example created using ITK-SNAP software.

4

MATERIALS AND METHODS

4.1 Dataset

The dataset used in this study is composed by 53 oncological patients who underwent double whole-body PET/CT, a scan on the Philips Vereos Digital and the other on Philips GEMINI TF (Philips Medical Systems). Only patients with lesions were included. Patients characteristics are registered in table 4.1.

The study was realized at NM department of the Champalimaud Clinical Centre (CCC) between December 2018 and February 2019. All patients gave written informed consent. To ensure confidentiality, patients' data were de-identified. This observational and cross-sectional research study was approved by the Ethics Committee of the CCC.

Table 4.1: Patients characteristics included in the final dataset.

Characteristic	
Sex (M/F)	24 / 29
Age (years)*	65 ± 10
Height (cm)*	165 ± 10
Weight (kg)*	73 ± 16
Body mass index (kg/m^2)*	26.7 ± 4.6

*Data presented as mean ± standard deviation.

4.2   Image acquisition and reconstruction

For the dual imaging protocol, all the patients included underwent a single intravenous injection of 245 ± 55 MBq of ^{18}F-FDG (in average 3.4 MBq/kg).

The first acquisition started approximately 70 ± 15 min post-injection and the second scan approximately 40 ± 22 min after the beginning of the first. On average, the VEREOS time acquisition post-injection was 63 min and 104 min, respectively for the first and second scan. In GEMINI, the acquisition time post-injection was for the first scan 74 min and for the second 113 min. Twenty-six patients performed the first image acquisition on the digital PET/CT device, and twenty-seven on the analog device.

The time difference between the acquisitions was the shortest possible based on the constraints of the clinical department as throughput requirements and external workflow factors. The time difference presented includes the time period of the first scan, patient exit from the equipment and preparation for the next acquisition.

The data were acquired with an acquisition time of 70 sec/bed position on both scanners. The total acquisition time depends on the size of each subject and the effective axial field of view (AFOV) of the scanners, so it was different from patient to patient and scanner to scanner. Then, image reconstruction was applied with default/equivalent clinical parameters, which means that none of the scanners were used at its maximum capability. Table 4.2 summarizes the main parameters of GEMINI and VEREOS images reconstruction. PET images were reconstructed with a 3D ordered subset expectation maximization (OSEM) algorithm, with 3 iterations and 33 subsets on GEMINI and with 3 iterations and 15 subsets on VEREOS. In both cases, an isotropic voxel size of 4 mm was set. The existence of a relaxation factor of 0.7 in the reconstruction of the GEMINI images (in VEREOS is 1.0) compensates for the difference in the number of subsets by controlling image smoothing.

Three datasets of images from the VEREOS were built: the original (as obtained from the standard reconstruction algorithm) and another two after smoothing the original VEREOS images with two different Gaussian filters. Although the smoothing represents a downgrade of the images, it is necessary in order to VEREOS images meet the requirements necessary to satisfy the EARL standards. To do so, it is recommended by Philips Healthcare to use a 3D Gaussian smoothing filter with a full width at half maximum (FWHM) of 5 mm in the images from VEREOS [58]. However, Koopman *et al* [14] found that to satisfy EARL standards applying the same reconstruction parameters as we used and a Gaussian smoothing filter with 2 to 4 mm FWHM is the better option. Thus, we built two new datasets of smoothed images from the VEREOS after computing two Gaussian filters, one with 3 mm and the other with 5 mm FWHM. Consequently, the lesion features extracted from the GEMINI images were compared with the lesion features

extracted from the three datasets of VEREOS images. The standard reconstruction used to obtain the GEMINI images already fulfill EARL specifications.

Table 4.2: VEREOS and GEMINI reconstruction parameters.

Parameter	GEMINI	VEREOS
PET		
Volume of each voxel	4×4×4 mm^3	4×4×4 mm^3
3D algorithm	OSEM	OSEM
Iterations	3	3
Subsets	33	15
Relax factor	0.7	1.0
Post-reconstruction filter (FWHM)	None	None, 3 mm and 5 mm
CT		
Attenuation Correction	Yes	Yes

4.3 Images preparation

After all the images were de-identified, the respective DICOM (Digital Imaging and Communications in Medicine) files were converted to the NIFTI (Neuroimaging Informatics Technology Initiative) format using an in-house built program, created in C++ language. This was applied for all CT and PET images of each patient. Besides that, in the PET images, the program uses the DICOM information to calculate the standardized uptake value (SUV) conversion factors and, afterward, it converts the image values from tissue radioactivity concentration to SUV (robust PET quantifier). The SUV is used to calculate quantitative measurements of tumor uptake since it has enhanced or replaced qualitative interpretation [59].

The SUV is calculated simply as a ratio of the tissue concentration (in Bq/ml) at the time of the acquisition, C_{PET}, and the injected activity (in Bq) divided by the body weight (in g), as it is represented in equation 4.1.

$$SUV = \frac{C_{PET}}{\frac{Injected\ activity}{Weight}} \tag{4.1}$$

4.4 Lesions identification

For lesions identification, the images were prepared in 3D Slicer 4.10.0 software platform (*https://www.slicer.org*) [60] for three-dimensional visualization, as it is shown in figure 4.1. Using this software, the identification and annotation of all lesions, by an experienced nuclear medicine physician, was accomplished. The annotation of each lesion was done first on the VEREOS data, in which the nuclear medicine physician was unaware of the image source.
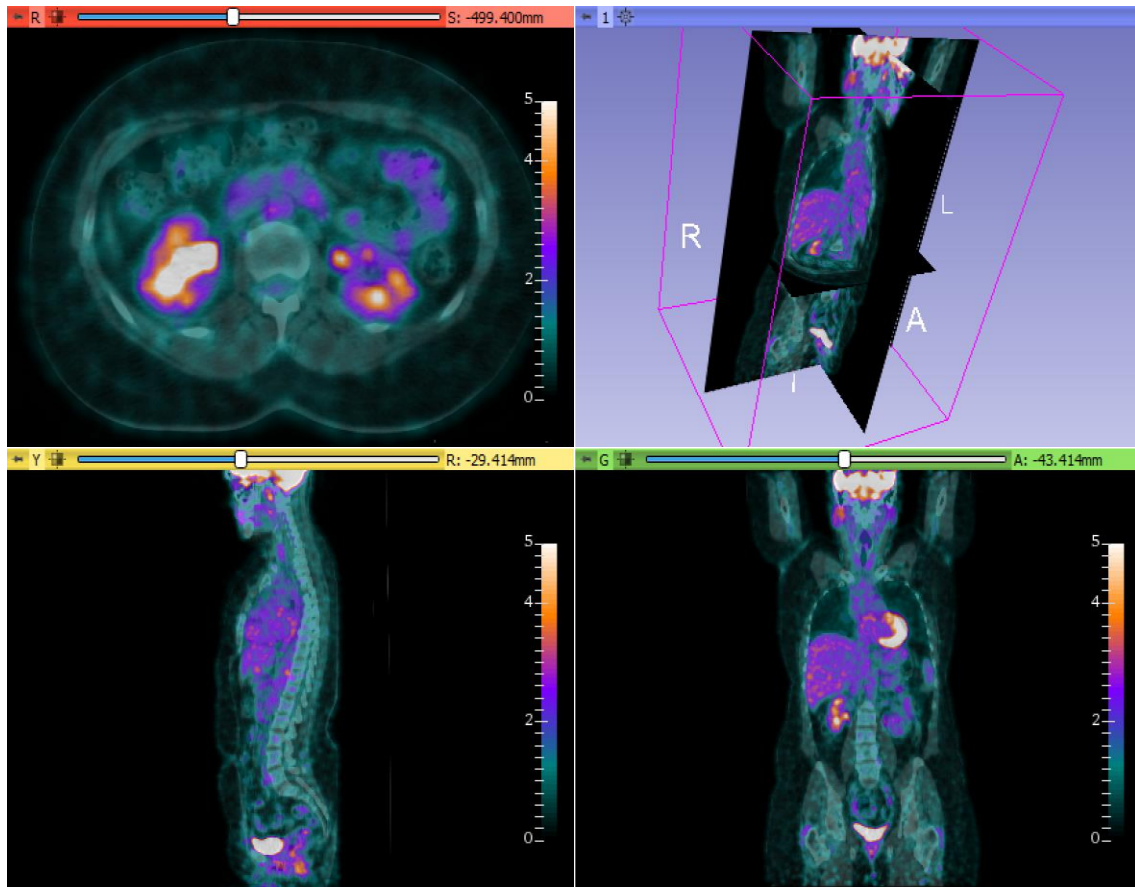
Figure 4.1: Representative axial, coronal and sagittal slices of a PET/CT image of the same patient in the software 3D Slicer. The color map used is similar to the one used in the Nuclear Medicine Department daily.

The annotation of the lesion to be used in the study was done by delineating the lesion ensuring that the surrounding background was included, as it is represented in figure 4.2 for a liver and uterus lesion (example not included in the dataset). For each subject, a mask was created, and for each lesion identified, a ROI surrounding the lesion was drawn with a different label number. After all lesions being annotated, the mask created was saved in the NIFTI format as the original PET.
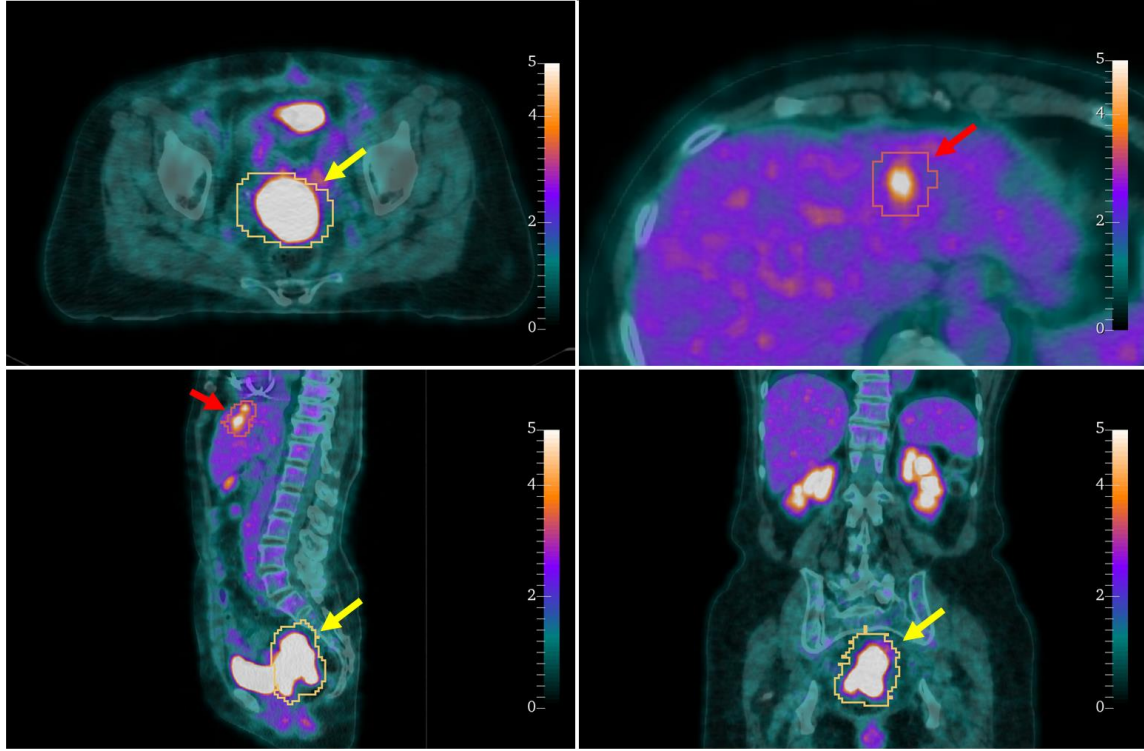
Figure 4.2: Delineation process of two lesions in one subject. A mask was created, and for the two lesions, located in liver and uterus, it was drawn a ROI, containing the surrounding background. Each ROI was labeled with a different number. The uterus lesion is contained in the label number 1 (yellow) and the one located in liver in the label number 2 (red).

When the identification and delineation of the lesions in the images acquired in the VE-REOS were finished, all annotations were transposed to GEMINI data. If a lesion was identified in VEREOS image but it was not visible in the respective GEMINI image, the lesion was not included in the study (exclusion criteria).

4.5 Automatic segmentation

Manual segmentation of all lesions present in the images by the nuclear medicine physician would be the best scenario, but this task is impracticable due to the number of hours the physician would have to be available. Consequently, an automatic segmentation program was implemented to make this task feasible.

From the segmentation methods presented in Chapter 3, it was implemented one based on thresholds, and two based on the clustering techniques: K-means and Bayesian segmentation. The

techniques for the segmentation were implemented in C++ language using a free open-source library insight toolkit (ITK) (*https://itk.org*) [61], [62].

The automatic segmentation is applied to the PET images, in the local of each ROI previously delineated by the physician, and not in the entire whole-body image. For each patient, each lesion identified by the physician is segmented independently from the other possible lesion that the patient may have.

4.5.1 Threshold

As the algorithms based on thresholds are commonly used in medical images [20]–[23], this method was implemented. In this study, it was considered as belonging to the lesion the voxels with an SUV value higher than the chosen threshold. In this case, two thresholds were tested: 50% and 80% of maximum SUV presented in the region delineated by the physician [46], [47]. Thus, two different segmentations were performed using this technique.

4.5.2 K-Means

In this method, the voxels are assigned to the nearest cluster depending on the distance between the voxel intensity and the cluster centroid. When implementing the K-means method, an initialization is necessary (initial centroids of each class). In our implementation, the initial centroids were calculated based on the percentiles of the SUV distribution in the delineated region. Equation 4.2 shows how the percentiles are calculated, in which k is the number of clusters (classes) selected by the user, and i the incrementing values in the interval $[0, k-1]$,

$$centroid_i = SUV \ with \ percentil \left(\frac{1}{2k} + i\frac{1}{k} \right) \tag{4.2}$$

After specifying the initialization, the algorithm uses the K-means statistical classifier in order to attribute a label to every voxel in the ROI to segment. The output of the K-mean classifier is a label map with pixel values indicating the classes they belong to. Pixels with an intensity equal to 0 belong to the first class, with intensity 1 belong to the second class and so on.

Although only two classes need to be defined (lesion and not lesion), in our implemented algorithm more classes can be specified. In these cases, after the initial multiclass segmentation, groups of classes need to be merged in order to obtain just two classes. The optimal criteria for the initial number of classes and the classes that are going to be merged (if necessary) were

stablished experimentally and are defined in section 4.5.4. This strategy allows the fine tuning of the segmentation.

4.5.3 Bayesian

The Bayesian classifier predicts membership probabilities for each class, i.e., the probability of a given voxel belongs to a particular class. Then, the class with the highest probability is considered as the most likely class.

Regarding the initialization of this method, it was applied a filter to generate a Gaussian mixture model for the image to enter in the Bayesian Classifier. The filter produces a membership image that indicates the membership of each voxel to each class by creating Gaussian density functions centered around a number of intensity values. This number of intensity values is obtained by running k-means on the ROI.

As for the K-means, the output of the Bayesian classifier is a label map with pixel values indicating the classes they belong to. The number of classes in which the Bayesian classifier divides the voxels can be defined by the user. As for the K-means, an optimized automatic strategy was defined to choose the optimal number of initial classes and the classes that are going to be merged, if necessary (see next section).

4.5.4 Segmentation optimization

The lesions to be segmented are very heterogeneous. There are lesions with hundreds of voxels and lesions with less than 10 voxels. Some of them are located in organs with high normal uptake (as is the liver, for instance) and others are in organs with very low normal uptake (as are the lungs). Others are next to regions with very high uptake, as is the bladder. In order to try to improve the segmentation, we designed a strategy to make it more adaptable. It was developed simultaneously to the implementation of the K-means and Bayesian segmentation methods.

Based on our experience (training) in this type of data, dividing voxels of each region delineated by the physician in 2 or 3 classes gives satisfactory segmentation results in the ^{18}F-FDG PET images. There are lesions where a simple binary classification (two classes) is very good. In other lesions is better to segment initially in three classes and then merge the two lower classes or merge the two higher classes. Thus, an automatic criterion needed to be defined.

So, instead of the user indicate the number of initial classes in which the region delineated by the physician must be divided and the number of classes to merge at the end of the process, the program divides the sample automatically in 3 classes, according to the segmentation process

chosen (K-means or Bayesian). Thereafter, a coefficient (equation 4.3) is calculated, and then, depending on their index, the classes to merge will be the two lower or the two higher, or the program runs again the segmentation method but now defining only 2 initial classes. The coefficient is calculated by,

$$coef = (m_3 - m_1)/(m_3 + m_1) \qquad (4.3)$$

where m_3 is the mean SUV of the voxels in class 3 (with higher values) and m_1 the mean SUV of the voxels in the class with lower values (class 1).

The empirical rule created based on prior knowledge define the number of classes ($n_{classes}$) to be initially created and the criterion for the merging (if necessary). The rule is defined as follows, depending on the coefficient index ($coef \in [0,1]$),

$$\begin{cases} coef < 0.92, \ then \ n_{classes} = 3 \ and \ merge \ the \ two \ lower \ classes \\ 0.92 < coef < 0.94, \ then \ repeat \ the \ segmentation \ with \ n_{classes} = 2 \qquad (4.4) \\ coef > 0.94, \ then \ n_{classes} = 3 \ and \ merge \ the \ two \ higher \ classes \end{cases}$$

4.6   Automated segmentation validation

In order to validate the automatic segmentation, it was verified the superposition between manual segmentation of a set of lesions and the results of the implemented segmentation methods. To ensure that possible differences in the comparison of the features extracted from both medical images acquired in VEREOS and GEMINI are not due to the automatic segmentation, this validation was done in images from both equipment. This validation process involves the steps represented in figure 4.3.
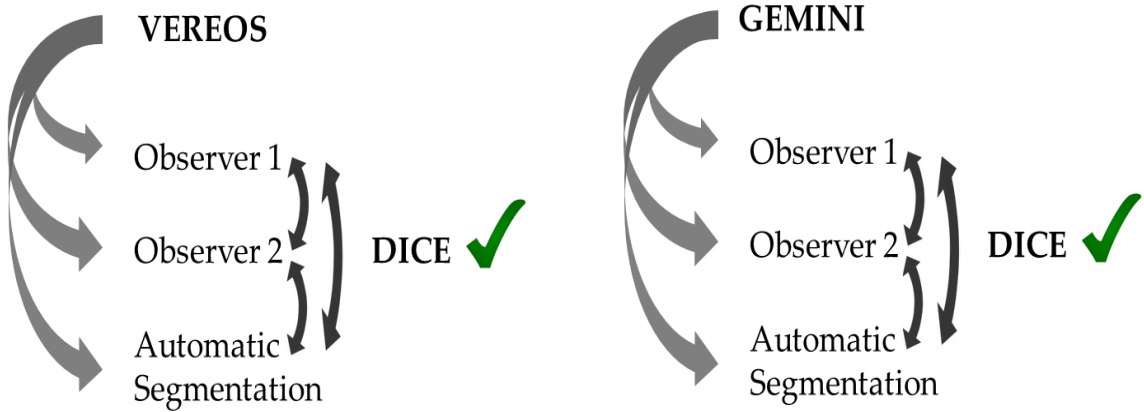
Figure 4.3: Schematic representation of the validation segmentation process.

A manual and independent segmentation was accomplished by two different observers in the images acquired from VEREOS (standard reconstruction) and GEMINI. This segmentation was performed in 50 consecutive lesions previously identified by an experienced nuclear medicine physician (Chapter 4.3) in the first 15 PET/CT images included in the dataset. So, each observer did 100 manual segmentations, 50 in VEREOS images and the correspondent 50 lesions in GEMINI images. This was done independently for both equipment (unaware of the segmentations performed in the images from the other equipment). The manual segmentation was performed on lesions previously identified by the physician, using ITK-SNAP 3.8.0 (*www.itksnap.org*) [63] and 3D Slicer 4.10.0 (*https://www.slicer.org*) [60] software.

To measure the segmentation agreement between observers and the automatic segmentation, the Dice coefficient was calculated. This coefficient is a widely accepted evaluation metric, used as a measure to assess the spatial overlap between two segmentations [64]–[66]. Being A and B two regions, the Dice Coefficient is given by:

$$Dice(A, B) = 2 \times \#(A \cap B)/(\#A + \#B) \qquad (4.5)$$

where $\cap$ is the intersection and # is an operator that represents the number of elements of the set (figure 4.4). The Dice coefficient can be defined as the overlap proportion.
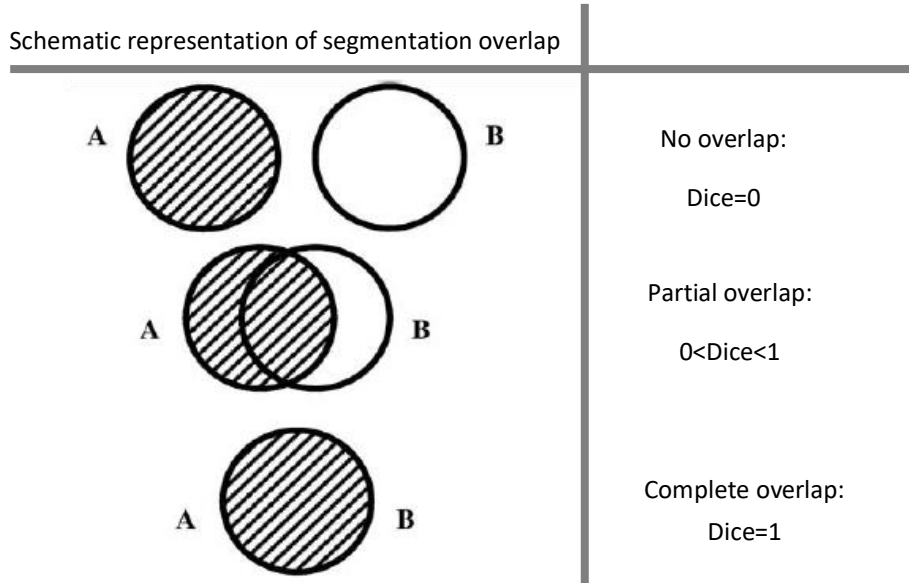
Figure 4.4: Schematic representation of segmentation overlap and Dice coefficient. Adapted from [66].

The manual segmentations were compared with the automatic segmentation obtained by applying the segmentation methods already implemented. The same 100 lesions (50 lesions on the VEREOS images and the corresponding 50 on the GEMINI images) that were manually segmented were also segmented automatically using the following methods:

- Threshold segmentation with 50% of the maximum SUV;
- Threshold segmentation with 80% of the maximum SUV;
- K-means segmentation with $n_{classes} = 2$;
- K-means segmentation with $n_{classes} = 3$ & merge the two lower classes;
- K-means segmentation with optimized parameters for the number of classes and merging process;
- Bayesian segmentation with $n_{classes} = 2$;
- Bayesian segmentation with $n_{classes} = 3$ & merge the two lower classes;
- Bayesian segmentation with optimized parameters for the number of classes and merging process.

Each of these segmentation results was compared with the manual segmentation. The automatic segmentation method that is going to be applied to the full dataset is the one with the higher agreement (higher Dice) with the manual segmentation. In order to understand if Dice obtained between manual and automatic segmentations are significantly different from each other, the Friedman's test was calculated using IBM SPSS (significance level $\alpha=0.05$). The

comparison groups are: Dice between observer 1 and observer 2; Dice between observer 1 and automatic segmentation and Dice between observer 2 and automatic segmentation.

4.7 Features extraction

Image features describing tumor characteristics were extracted in an automated way after tumor segmentation. Twenty-five features were defined. They can be broadly divided into two groups: tumor intensity and tumor geometry. The first group quantifies tumor intensity by first-order statistics, i.e. calculated directly from the histogram of all tumor voxel intensity values. The second group involves features based on the shape and size of the tumor. All algorithms to be used for feature extraction were implemented in C++ language. The definition of the features is mostly based on the paper by Aerts et al. [67].

4.7.1 Features based on first order statistics

First order statistics (FOS) define the voxel intensities distribution within a medical image through usually used and basic metrics. In table 4.3, FOS extracted are clarified, considering A lesion voxel intensities, N the number of voxels and H the first order histogram with N_l discrete intensity levels.

The standard deviation, variance and mean absolute deviation are measures of how much the uptake diverges from the mean, i.e., measures of histogram dispersion. The peak skewness is a measure of histogram asymmetry around the mean, and kurtosis measures histogram sharpness. Entropy and uniformity allow measuring histogram randomness. The SUV_{peak} is an average of voxels intensities within a fixed volume comprising the highest pixel value and it has shown to be a reliable parameter for ^{18}F-FDG PET/CT quantification [68].

4.7.2 Geometric features

Features related to the three-dimensional size and shape of the tumor region are comprised in this group. Geometric parameters are explained in table 4.4.

Surface area (SA) and MTV provide information on the size of the tumor. SA is achieved by summing the areas of voxel faces that are in the boundary between the lesion and its background. The two compactness methods, sphericity, spherical disproportion, and surface to volume ratio are measures of how spherical, rounded or elongated the shape of the lesion is. TLG is the product of the lesion volume and its mean uptake. TLG is among the extracted features since it is a biomarker considered valuable to help indicating the optimal therapeutic strategy [69].

Let in the following definitions (table 4.4), N denotes the number of voxels in the lesion, V_v the volume of each voxel, and A the set of the lesion voxel intensities, as considered to FOS.

Table 4.3: Implementation of the first order statistics.

First Order Statistics	Methods[*]	
Energy	$$energy = \sum_{i}^{N} A(i)^2$$	
Entropy	$$entropy = \sum_{i}^{N_l} H(i) \log_2 H(i)$$	
Kurtosis	$$kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(A(i)-\bar{A})^4}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(A(i)-\bar{A})^2}\right)^2}$$	, $\bar{A}$ is the mean of A.
Maximum (SUV$_{max}$)	The maximum intensity value of A.	
Mean (SUV$_{mean}$)	$$mean = \frac{1}{N}\sum_{i}^{N} A(i)$$	
Mean Absolute Deviation (MAD)	The absolute deviations of the mean of all voxel intensities.	
Median	The 50th percentile of A.	
Minimum	The minimum intensity value of A.	
Range	The range of intensity value of A.	
Root Mean Square (RMS)	$$RMS = \sqrt{\frac{\sum_{i}^{N} A(i)^2}{N}}$$	
Skewness	$$skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(A(i)-\bar{A})^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(A(i)-\bar{A})^2}\right)^3}$$	, $\bar{A}$ is the mean of A.
Standard Deviation (SD)	$$SD = \left(\frac{1}{N-1}\sum_{i=1}^{N}(A(i)-\bar{A})^2\right)^{1/2}$$	, $\bar{A}$ is the mean of A.
Uniformity	$$uniformity = \sum_{i=1}^{N_l} H(i)^2$$	
Variance	$$variance = \frac{1}{N-1}\sum_{i=1}^{N}(A(i)-\bar{A})^2$$	, $\bar{A}$ is the mean of A.

Coefficient of Variation (CV)	$CV = \dfrac{SD}{mean}$	
SUV$_{peak(3x3x3)}$	Maximum average SUV within 3x3x3 voxels volume	
SUV$_{peak(2x2x2)}$	Maximum average SUV within 2x2x2 voxels volume	

[*]A is the set of lesion voxel intensities, N is the number of voxels and H the first order histogram with N_l discrete intensity levels.

Table 4.4: Implementation of the geometric features.

Geometric features	Methods[*]	
Surface Area (SA)	Sum of faces area that are in the border between lesion and background	
Metabolic Tumor Volume (MTV)	$MTV = N \times V_v$	
Compactness 1	$compactness1 = \dfrac{MTV}{\sqrt{\pi} \times SA^{\frac{2}{3}}}$	
Compactness 2	$compactness2 = 36\pi \times \dfrac{MTV^2}{SA^3}$	
Spherical disproportion	$spherical_{dis} = \dfrac{SA}{4\pi \times R^2}$	, R is the radius of a sphere with the same volume as the tumor.
Sphericity	$sphericity = \dfrac{\pi^{\frac{1}{3}} \times (6 \times MTV)^{\frac{2}{3}}}{SA}$	
Surface to volume ratio (SVR)	$SVR = \dfrac{SA}{MTV}$	
Total Lesion Glycolysis[**] (TLG)	$TLG = V \times \displaystyle\sum_i^N A(i)$	

[*]N is the number of voxels in the lesion, V_v the volume of each voxel, and A the set of the lesion voxel intensities.
[**]TLG does not belong exclusively to geometric features since it relies on first order statistics too.

4.8 Statistical analysis

As a final step for the reproducibility study, a statistical analysis involving the 25 features extracted from the lesions identified in the PET/CT images from 53 oncological patients was completed. To find out the features that are reproducible between both PET/CT devices (VEREOS and GEMINI), a comparison analysis was processed using IBM SPSS version 20 (*https://www.ibm.com*), software commonly used for statistical analysis in this type of studies [70]–[73]. The features analysis were performed between the images from GEMINI and VEREOS reconstructed with default clinical parameters and between the images from both devices reconstructed to accomplish EARL specifications (GEMINI reconstructed with default parameters and VEREOS reconstructed with an additional Gaussian post-smoothing filter of 3mm and 5mm).

Agreement between the features extracted from the image of both equipment was measure first using the intraclass correlation coefficient (ICC) for absolute agreement. The respective *p*-values were also computed. Then, linear correlation coefficients between the correspondent features were also calculated and depending on their values, conversion factors were estimated using linear regression.

For all statistical tests done, a significance level of 5% was defined.

RESULTS

5.1 Detected lesions

The following results are representative of the step explained in Chapter 4.4. In this stage, 287 lesions were identified by an experienced nuclear medicine physician first on VEREOS images. Then, all the annotations were transposed to the GEMINI images by an observer. In 4 of the 287 cases, the lesions already identified in VEREOS images were not easily identifiable in the corresponding GEMINI images. For that reason, these cases were reviewed by the nuclear medicine physician that corroborated the uncertainty transposing the identification to GEMINI image. So, these lesions were excluded from the dataset, which resulted in 283 lesions to analyze. An image showing one of these questionable cases can be consulted on appendix A (figure A.1).

The main anatomical localization of the lesions included bones, lungs, and liver. The localization distribution of the 283 lesions involved in the study are exhibited in the following table (table 5.1).

Table 5.1: Anatomical localization of all lesions included in the study.

Lesions location (N=283)	Number of lesions
Bones	106
Soft tissue	
Head and neck	
Lymph nodes	16
Thorax	
Lymph nodes	14

	Lung	80
	Breast	10
Abdomen		
	Lymph nodes	3
	Liver	34
	Pancreas	3
	Stomach	1
Pelvis		
	Lymph nodes	9
	Uterus	1
	Rectum	5
	Kidneys	1

In figure 5.1 lesions delineation examples are demonstrated. On the left column, is presented the delineation done by the physician in the VEREOS images and on the right the delineation transposed to the GEMINI images. In this patient, 9 lesions were identified. Therefore, a mask containing 9 labels (one label for each lesion) was drawn, although neither all lesions are visible in the slices shown. Five of those nine lesions are illustrated in the figure (three in the liver, one in the lungs and another in the pelvic region).
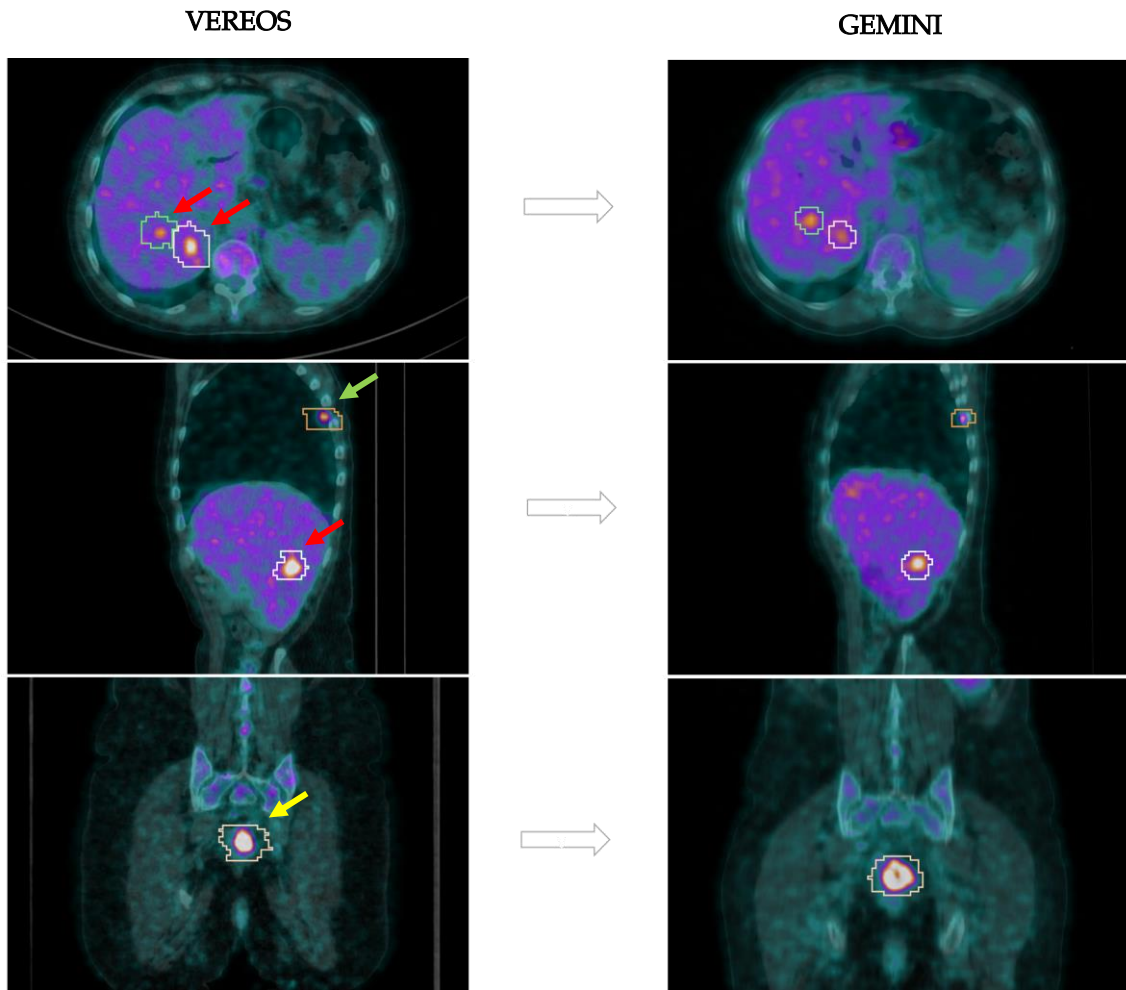
Figure 5.1: Mask annotation of five lesions in VEREOS and GEMINI images. The identification and annotation in the mask were made first on VEREOS by a nuclear medicine physician. Then, the regions drawn were transposed to the corresponding GEMINI image. These three lesions in the liver (marked by the arrows in red), one in the pelvic region (arrow in yellow), and one in the lungs (in green) were included in the dataset.

5.2   Lesions segmentation

The purpose of the automatic segmentation algorithm was to segment each of the lesions identified by the physician. For each patient, each lesion is segmented independently from the other possible lesions the patient may have.

This section is divided into the computation results from the automatic segmentation and validation with the manual segmentation.

The PET images acquired from GEMINI and VEREOS have an isotropic voxel size of 4 mm, and so the real images are not as smooth as the standard visualization images presented to

the physician (see an example in figure 5.2). All the images presented so far were smooth, but this option may induce a bias in the observer, especially when segmentations are being evaluated. Thus, in this section, the segmentation results are shown with pixelated images to better understand the real segmentation results.
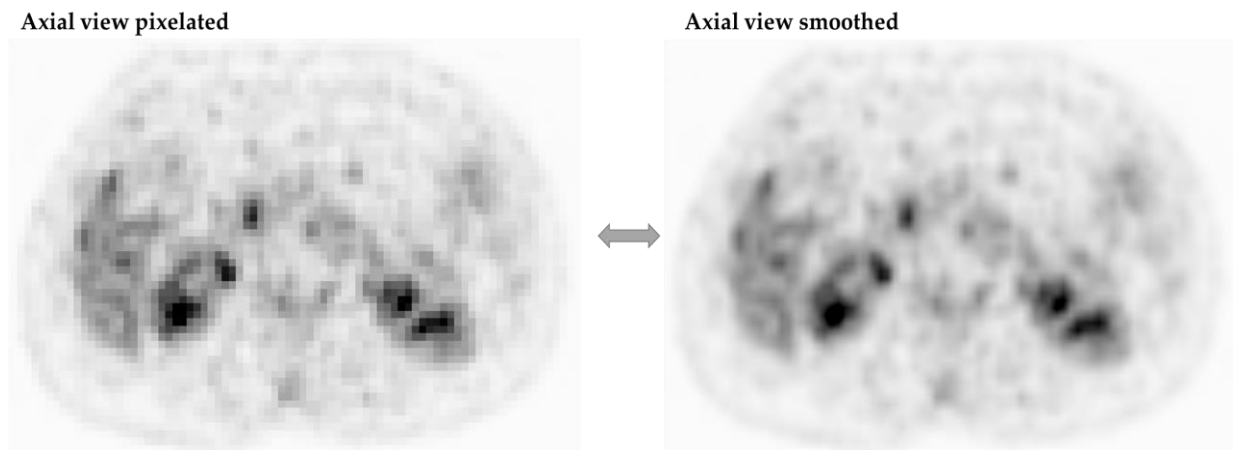
Figure 5.2: Axial view of a PET image pixelated on the left and smooth on the right.

5.2.1 Computed segmentation

Threshold segmentation was firstly implemented and applied to the dataset of the identified lesions. A threshold of 50% and 80% of the maximum SUV were set. Figure 5.3 shows three lesions delineated by this method. Note that these lesions have different sizes, locations, and surrounding background.
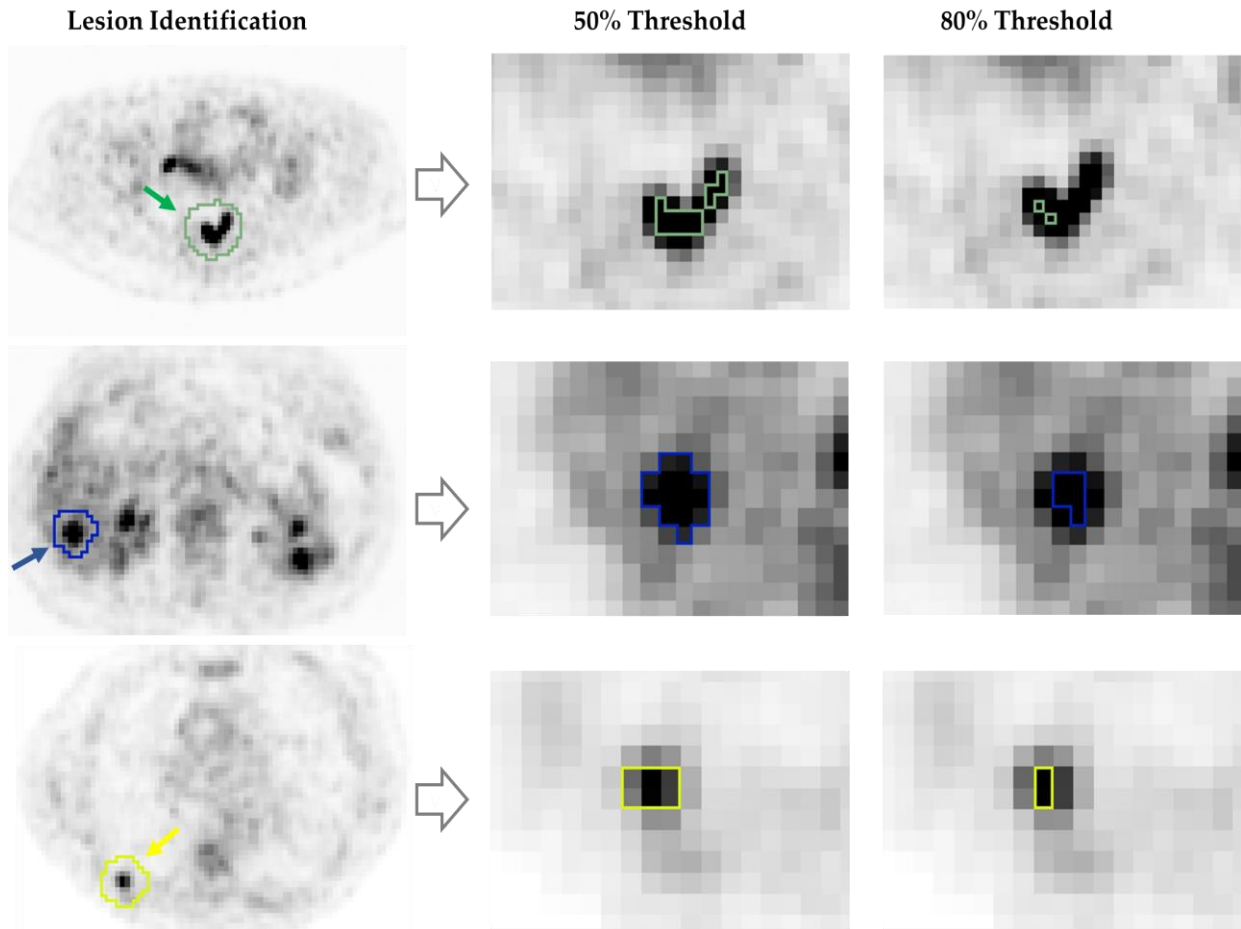
Figure 5.3: Results of threshold segmentation of 50% and 80% applied in three different lesions. The lesion identified by a green arrow is located in the pelvic region. In blue, a lesion in the liver and in yellow, a lesion in the lungs. The results of the segmentation are in the images on the two right columns, in which the lesions are zoomed.

K-means segmentation was also evaluated. Different configurations were evaluated (see sections 4.5.2 and 4.5.4). Three segmentations examples are shown in figure 5.4 for the same exactly lesions represented in the previous images (figure 5.3).
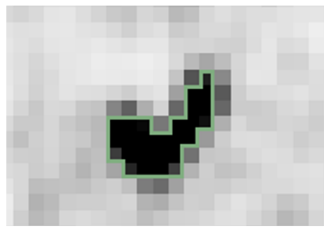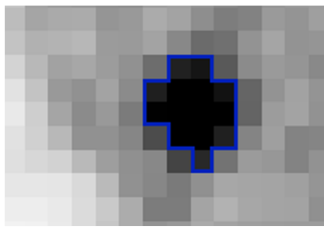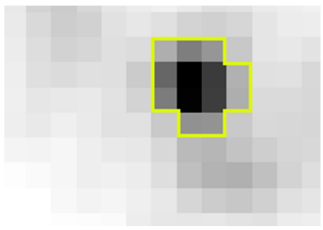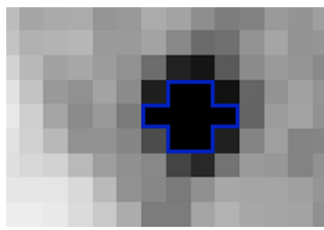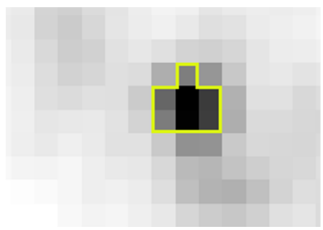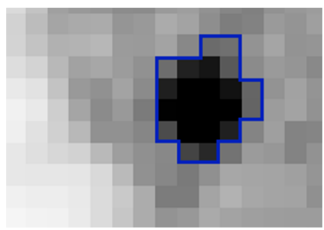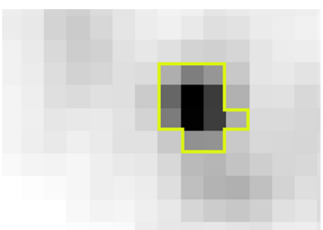
	Pelvic Region	Liver	Lung
K-means $n_{classes} = 2$			
K-means $n_{classes} = 3$ & merge the two lower classes			
K-means optimized			

Figure 5.4: Results of K-means segmentation in three lesions using different configurations. Apparently, the auto-adjusted configuration (K-means optimized) originated better segmentations than the two fixed configurations.

The equivalent was executed using the Bayesian segmentation. The results are presented in the following figure 5.5.
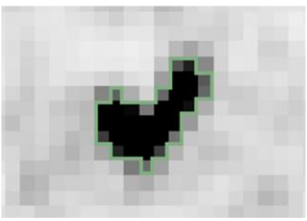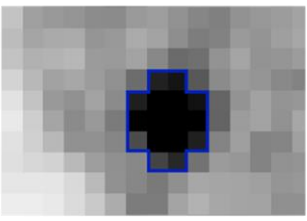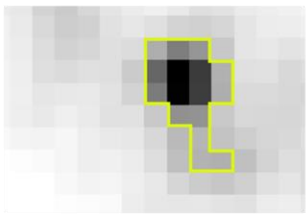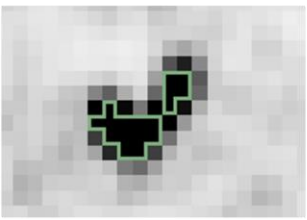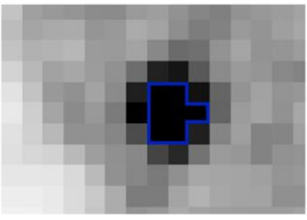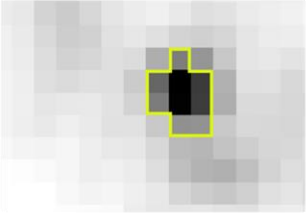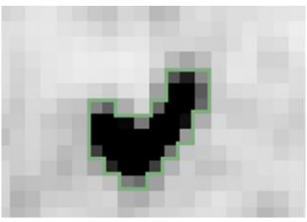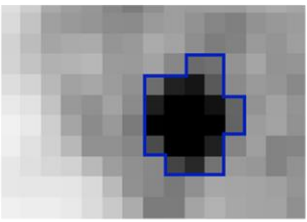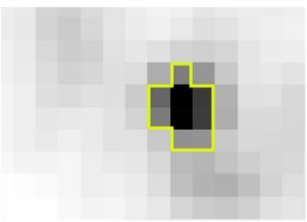
	Pelvic Region	Liver	Lung
Bayesian $n_{classes} = 2$			
Bayesian $n_{classes} = 3$ & merge the two lower classes			
Bayesian optimized			

Figure 5.5: Results of Bayesian segmentation in three lesions using different configurations. Auto-adjusted configuration (Bayesian optimized) apparently originated better results than the fixed configurations.

5.2.2 Validation

Manual segmentation is a time-consuming process. However, with the objective of validating the results from the automatic segmentation, two observers delineated the same 50 lesions. Lesions were segmented twice by each observer, one in the images from the GEMINI and the other in the images from the VEREOS. An example of the outcomes of these segmentations is shown in figure 5.6.
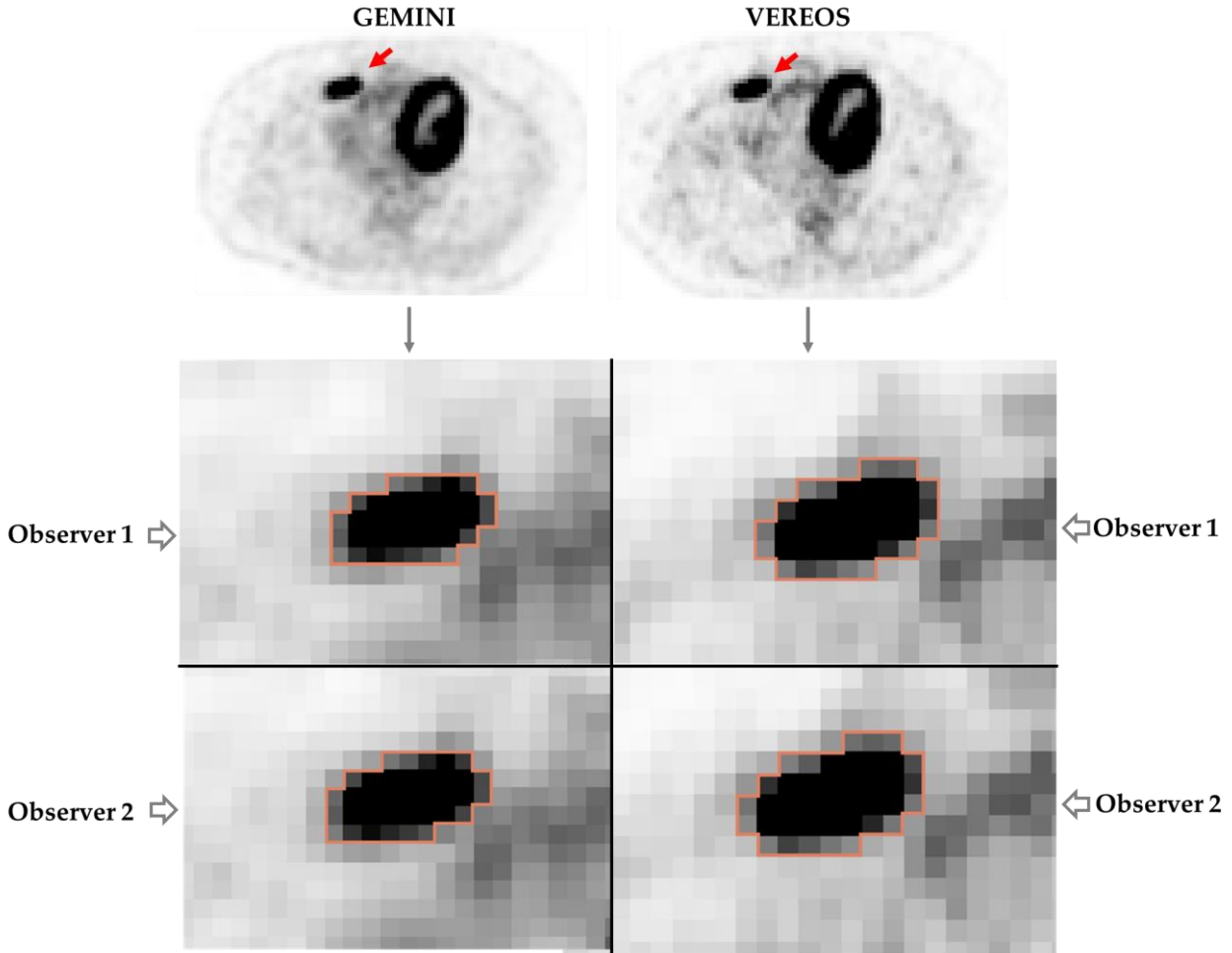
Figure 5.6: Manual segmentation by two observers in the same lesions in images acquired by VEREOS and GEMINI scanners. The lesions to delineate manually are identified by a red arrow in the original PET/CT images. Each manual segmentation is represented by a rosy contour.

After all the lesions of the validation set were manually segmented, a validation of the automatic segmentation with the manual was completed. It was calculated the Dice coefficient between the segmentations. This coefficient allows us to evaluate the overlap between two segmentations.

In each of the following tables (table 5.2 – table 5.9), the median and interquartile range (IQR) of Dice coefficients between the manual and automatic segmentation are presented. Besides that, the Dice resulted from the superposition among the lesions segmented by observers was also obtained, and it is repeated in each table. The nonparametric test (Friedman's test) evaluates if exists significant difference among the three Dice distributions. If there is significant difference in the Dice coefficient, the pairwise comparison between the groups is presented. The results presented consider all manual segmentations done in both scanners (100 segmentations each one).

➢ Manual segmentation vs threshold method (50%)

Table 5.2: Dice coefficient between manual segmentation and automatic segmentation using 50% of the maximum SUV.

		Observer 1 vs Observer 2 (I)	Observer 1 vs Threshold 50% (II)	Observer 2 vs Threshold 50% (III)	*p-value* (Friedman's test)
Dice	Median	0.84	0.70	0.66	< 0.001*
	IQR	0.22	0.34	0.29	

*For pairwise comparison only the samples II and III showed no statistical significance ($p = 0.112$).

➢ Manual segmentation vs threshold method (80%)

Table 5.3: Dice coefficient between manual segmentation and automatic segmentation using 80% of the maximum SUV.

		Observer 1 vs Observer 2 (I)	Observer 1 vs Threshold 80% (II)	Observer 2 vs Threshold 80% (III)	*p-value* (Friedman's test)
Dice	Median	0.84	0.17	0.18	< 0.001*
	IQR	0.22	0.26	0.24	

*For pairwise comparison only the samples II and III showed no statistical significance ($p = 0.860$).

➢ Manual segmentation vs K-means (2 initial classes)

Table 5.4: Dice coefficient between manual segmentation and automatic segmentation using the K-means method for $n_{classes} = 2$.

		Observer 1 vs Observer 2 (I)	Observer 1 vs K-means (II)	Observer 2 vs K-means (III)	*p-value* (Friedman's test)
Dice	Median	0.84	0.74	0.76	0.028*
	IQR	0.22	0.23	0.21	

*For pairwise comparison, the samples II, III and I, III showed no statistical significance (respectively, $p = 0.377$ and $p = 0.083$).

> ➢ Manual segmentation vs K-means (3 initial classes and then merge the two lower classes)

Table 5.5: Dice coefficient between manual segmentation and automatic segmentation using the K-means method with $n_{classes} = 3$ and then merge the two lower classes.

		Observer 1 vs Observer 2 (I)	Observer 1 vs K-means (II)	Observer 2 vs K-means (III)	p-value (Friedman's test)
Dice	Median	0.84	0.66	0.64	< 0.001*
	IQR	0.22	0.39	0.33	

*For pairwise comparison only the samples II and III showed no statistical significance ($p = 0.750$).

> ➢ Manual segmentation vs K-means Optimized

Table 5.6: Dice coefficient between manual segmentation and automatic segmentation using optimized K-means method.

		Observer 1 vs Observer 2 (I)	Observer 1 vs K-means (II)	Observer 2 vs K-means (III)	p-value (Friedman's test)
Dice	Median	0.84	0.80	0.79	0.002*
	IQR	0.22	0.21	0.21	

*For pairwise comparison only the samples II and III showed no statistical significance ($p = 1.000$).

> ➢ Manual segmentation vs Bayesian (2 initial classes)

Table 5.7: Dice coefficient between manual segmentation and automatic segmentation using Bayesian method for $n_{classes} = 2$.

	Observer 1 vs Observer 2 (I)	Observer 1 vs Bayesian (II)	Observer 2 vs Bayesian (III)	p-value (Friedman's test)

Dice	Median	0.84	0.76	0.79	0.007*
	IQR	0.22	0.21	0.18	

*For pairwise comparison, the samples II, III and I, III showed no statistical significance (respectively, $p = 0.191$ and $p = 0.066$).

➢ Manual segmentation vs Bayesian (3 initial classes and then merge the two lower classes)

Table 5.8: Dice coefficient between manual segmentation and automatic segmentation using Bayesian method for $n_{classes} = 3$ and then merge the two lower classes.

		Observer 1 vs Observer 2 (I)	Observer 1 vs Bayesian (II)	Observer 2 vs Bayesian (III)	*p-value* *(Friedman's test)*
Dice	Median	0.84	0.55	0.53	< 0.001*
	IQR	0.22	0.40	0.38	

*For pairwise comparison only the samples II and III showed no statistical significance ($p = 0.525$).

➢ Manual segmentation vs Bayesian Optimized

Table 5.9: Dice coefficient between manual segmentation and optimized Bayesian method.

		Observer 1 vs Observer 2 (I)	Observer 1 vs Bayesian (II)	Observer 2 vs Bayesian (III)	*p-value* *(Friedman's test)*
Dice	Median	0.84	0.83	0.83	0.052
	IQR	0.22	0.26	0.25	

The segmentations of the two observers reached a median Dice coefficient of 0.84. The optimized K-means and Bayesian methods reached a median Dice coefficient between them and the manual segmentation of 0.80 and 0.83, respectively. Tables 5.2 to 5.9 show that the Bayesian

optimized segmentation method is the one that has no statistically significant differences comparatively to the manual segmentations (Friedman's test, $p = 0.052$)

To evaluate data dispersion among the three groups (Dice obtained between observers and Dice between observers and Bayesian optimized segmentation), figure 5.7 presents the boxplot for descriptive statistics. The median (Q_2) is similar among the three boxplots, as seen in the previous table 5.9. The minimum value in the data "Observer1 vs Bayes Opt." is an outlier that was reviewed. In this case, the program segmented less than the observer in VEREOS image. In this lesion, the ROI drawn in the identification step did not catch enough background for a correct segmentation. Contrariwise, the same lesion in GEMINI image was segmented properly by the program (Dice = 0.89 between observer and automatic segmentation).
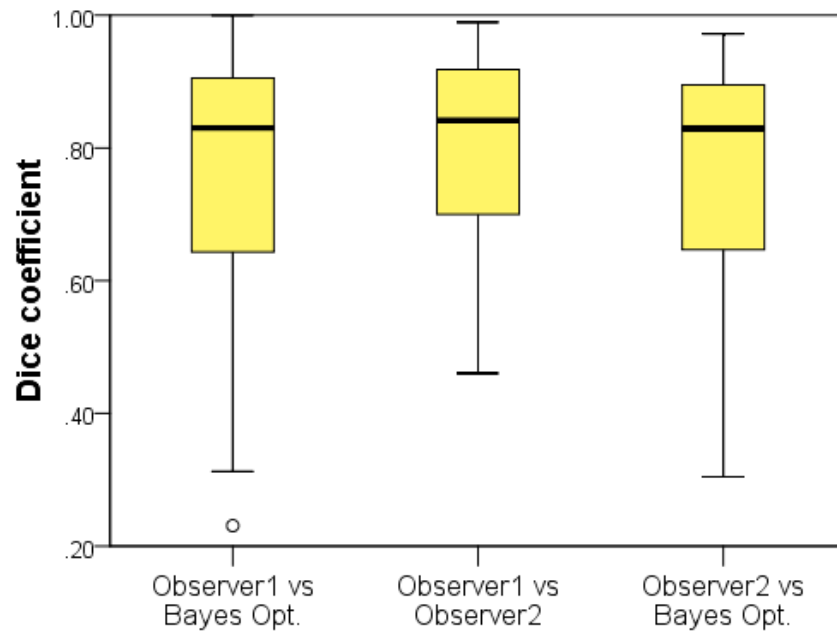
Figure 5.7: Boxplots representing Dice coefficient dispersion: Observer 1 vs Observer 2; Observer 1 vs Bayesian optimized; Observer 2 vs Bayesian optimized.

The boxplot analysis corresponding to the data presented in the tables 5.2 to 5.8 can be accessed on appendix A (figures A.2 – A.5).

5.3 Features extraction

Finally, to reach our goal, tumor biomarkers were extracted from the segmented lesions. Taking into consideration the previous results presented from segmentation attempts, the optimized Bayesian method was selected to delineate all 283 lesions included in the final dataset.

As explained in Chapter 4.7, 25 image features describing tumor characteristics were taken out. The extracted features can be divided into two groups: tumor intensity (first-order statistics) and tumor geometry (features based on shape and size). The agreement level between the features is accessed by the ICC.

5.3.1 Standard reconstruction

Firstly, the outcomes of the 25 features pulled out from the images of VEREOS and GEMINI reconstructed with default clinical parameters are presented in table 5.10. The median and IQR are exhibited for each feature extracted from the 283 lesions. These measures are shown for the lesions identified in GEMINI and VEREOS images. Correlation coefficient and ICC were computed between features extracted from images acquired in GEMINI and the corresponding ones from VEREOS scanner.

Table 5.10: Absolute agreement assessment of lesion features extracted from ^{18}F-FDG PET images acquired in GEMINI and VEREOS scanners with standard reconstruction. (IQR = Interquartile range; ICC = Intraclass correlation coefficient).

	FEATURES	GEMINI (standard reconstruction)		VEREOS (standard reconstruction)		Pearson Correlation	ICC**
		Median	IQR	Median	IQR		
	Energy	538.6	1869	695.3	2182	0.93	**0.93**
	Entropy	1.864	1.012	2.172	0.854	0.90	**0.85**
	Kurtosis	2.892	0.944	3.007	1.263	0.60	**0.57**
	Maximum (SUV$_{max}$)	4.863	3.839	5.916	3.894	0.88	**0.83**
	Mean (SUV$_{mean}$)	3.316	1.475	3.793	1.495	0.85	**0.80**

FIRST ORDER STATIS-TICS	Mean Absolute Deviation	0.435	0.594	0.637	0.702	0.89	**0.85**
	Median	3.199	1.332	3.574	1.443	0.83	**0.79**
	Minimum	2.515	0.826	2.718	0.961	0.66	**0.64**
	Range	2.169	3.192	3.281	3.816	0.88	**0.84**
	Root Mean Square	3.384	1.631	3.910	1.547	0.86	**0.81**
	Skewness	0.864	0.413	0.916	0.499	0.44	**0.43**
	Standard Deviation	0.545	0.730	0.782	0.850	0.88	**0.83**
	Uniformity	0.175	0.161	0.134	0.110	0.79	**0.71**
	Variance	0.297	1.034	0.612	1.626	0.85	**0.83**
	Coefficient of Variation	0.169	0.145	0.219	0.173	0.82	**0.76**
	$SUV_{peak(3x3x3)}$	3.480	2.661	3.864	2.610	0.93	**0.92**
	$SUV_{peak(2x2x2)}$	4.145	3.168	4.862	3.400	0.91	**0.88**
GEOMET-RIC FEATURES	Surface Area	1728	2592	1760	2976	0.98	**0.97**
	Metabolic Tumor Volume (MTV)	3200	6848	3200	7680	0.98	**0.98**
	Compactness1	12.31	10.76	11.99	13.06	0.97	**0.97**
	Compactness2	0.235	0.124	0.229	0.140	0.75	**0.74**
	Spherical dispro-portion	1.621	0.316	1.635	0.377	0.94	**0.92**
	Sphericity	0.617	0.114	0.612	0.131	0.85	**0.85**
	Surface to volume ratio	0.545	0.206	0.556	0.252	0.83	**0.82**
	Total Lesion Gly-colysis (TLG)[*]	10514	26855	11426	29493	0.97	**0.97**

* TLG does not belong exclusively to geometric features since it relies on first order statistics too.
** A *p-value* < 0.001 was achieved for all calculated ICC and Pearson correlation coefficient.

As an additional analysis, the variability of SUV_{max} (most frequently used feature in the clinic) was assessed between acquisitions when the first acquisition was on VEREOS and the second on GEMINI (88 lesions) and vice-versa (195 lesions).

For the first case (figure 5.8), in which the first image acquisition was performed in VEREOS scanner and the second in GEMINI scanner, there were no statistically significant differences between the SUV_{max} of lesions (Wilcoxon test, *p-value* = 0.755). When the first acquisition was done first in GEMINI and then in VEREOS (figure 5.9), the SUV_{max} differences were statistically significant (Wilcoxon test, *p-value* < 0.001).

Figure 5.8: Bland and Altman plot representing the variability of SUV_{max} between first acquisition in VEREOS and second acquisition in GEMINI. Both VEREOS and GEMINI images were reconstructed with standard protocol. In the x-axis is the mean SUV_{max} between acquisitions and in the y-axis is the SUV_{max} difference between second acquisition (GEMINI) and first acquisition (VEREOS). N=88 lesions.

Figure 5.9: Bland and Altman plot representing the variability of SUV_{max} between first acquisition in GEMINI and second acquisition in VEREOS. Both GEMINI and VEREOS images were reconstructed with standard protocol. In the x-axis is the mean SUV_{max} between acquisitions and in the y-axis is the SUV_{max} difference between second acquisition (VEREOS) and first acquisition (GEMINI). N=195 lesions.

5.3.2 Reconstruction for EARL specifications

Next, with the aim of fulfil EARL accreditation specifications, it was performed a comparison between the features extracted from the GEMINI images reconstructed with standard parameters (the same results previously presented) and the features extracted from VEREOS images after a post-reconstruction Gaussian filter of 3 mm (Table 5.11), as recommended by Koopman *et al* [14]. The same statistics accessible in the previous table are now in table 5.11.

Table 5.11: Absolute agreement assessment of lesion features extracted from ^{18}F-FDG PET images acquired in GEMINI with standard reconstruction and for VEREOS with an additional post-reconstruction Gaussian filter of 3 mm (IQR = Interquartile range; ICC = Intraclass correlation coefficient).

	FEATURES	GEMINI (standard reconstruction)		VEREOS (with Gaussian filter of 3 mm)		Pearson Correlation	ICC**
		Median	IQR	Median	IQR		
FIRST ORDER STATIS-TICS	Energy	538.6	1869	657.3	2112	0.93	**0.93**
	Entropy	1.864	1.012	2.027	0.900	0.90	**0.88**
	Kurtosis	2.892	0.944	2.936	1.098	0.58	**0.57**
	Maximum (SUV_{max})	4.863	3.839	5.513	3.794	0.89	**0.87**
	Mean (SUV_{mean})	3.316	1.475	3.714	1.503	0.85	**0.82**
	Mean Absolute Deviation	0.435	0.594	0.553	0.610	0.90	**0.88**
	Median	3.199	1.332	3.517	1.465	0.83	**0.80**
	Minimum	2.515	0.826	2.723	1.015	0.66	**0.64**
	Range	2.169	3.192	2.858	3.432	0.89	**0.87**
	Root Mean Square	3.384	1.631	3.806	1.558	0.86	**0.84**
	Skewness	0.864	0.413	0.903	0.495	0.41	**0.41**
	Standard Deviation	0.545	0.730	0.701	0.770	0.90	**0.88**
	Uniformity	0.175	0.161	0.147	0.129	0.80	**0.76**
	Variance	0.297	1.034	0.491	1.300	0.87	**0.86**
	Coefficient of Variation	0.169	0.145	0.203	0.152	0.82	**0.80**
	$SUV_{peak(3x3x3)}$	3.480	2.661	3.775	2.542	0.93	**0.93**
	$SUV_{peak(2x2x2)}$	4.145	3.168	4.583	3.193	0.91	**0.90**

	Surface Area	1728	2592	1728	2912	0.98	**0.97**
	Metabolic Tumor Volume (MTV)	3200	6848	3072	7552	0.98	**0.98**
GEOMET-	Compactness1	12.31	10.76	11.97	12.56	0.97	**0.97**
RIC	Compactness2	0.235	0.124	0.233	0.129	0.76	**0.76**
FEATURES	Spherical dispro-portion	1.621	0.316	1.626	0.325	0.94	**0.94**
	Sphericity	0.617	0.114	0.615	0.117	0.86	**0.86**
	Surface to volume ratio	0.545	0.206	0.567	0.243	0.84	**0.82**
	Total Lesion Gly-colysis (TLG)*	10514	26855	10653	30243	0.97	**0.97**

* TLG does not belong exclusively to geometric features since it relies on first order statistics too.
** A *p-value* < 0.001 was achieved for all calculated ICC and Pearson correlation coefficient.

Again, the SUV$_{max}$ variability between acquisitions was assessed. The following figures 5.10 and 5.11 show the results. The SUV$_{max}$ was statistically higher in the second acquisition comparatively to the first acquisition (Wilcoxon test, *p-value* = 0.030 when the first acquisition was in VEREOS and *p-value* < 0.001 when the GEMINI was first).
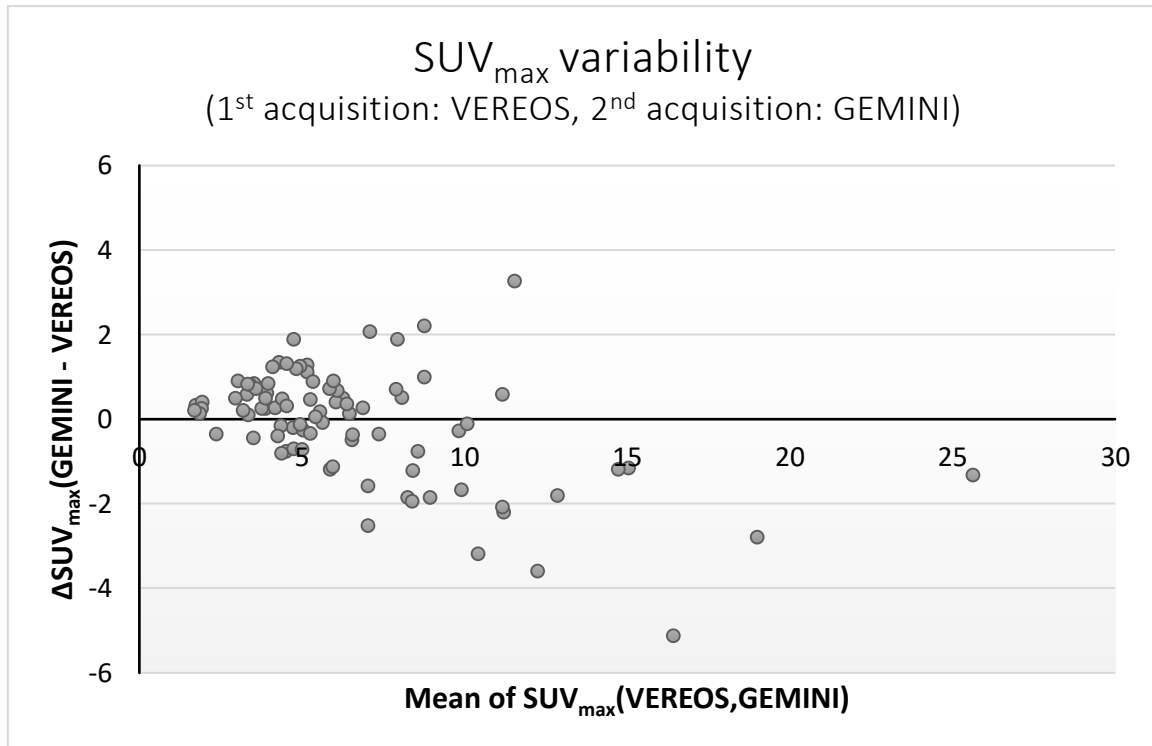
Figure 5.10: Bland and Altman plot representing the variability of SUV_{max} between first acquisition in VEREOS and second acquisition in GEMINI. GEMINI images were reconstructed with standard reconstruction and VEREOS images with a Gaussian post-reconstruction filter of 3 mm. In the x-axis is the mean SUV_{max} between acquisitions and in the y-axis is the SUV_{max} difference between second acquisition (GEMINI) and first acquisition (VEREOS). N=88 lesions.
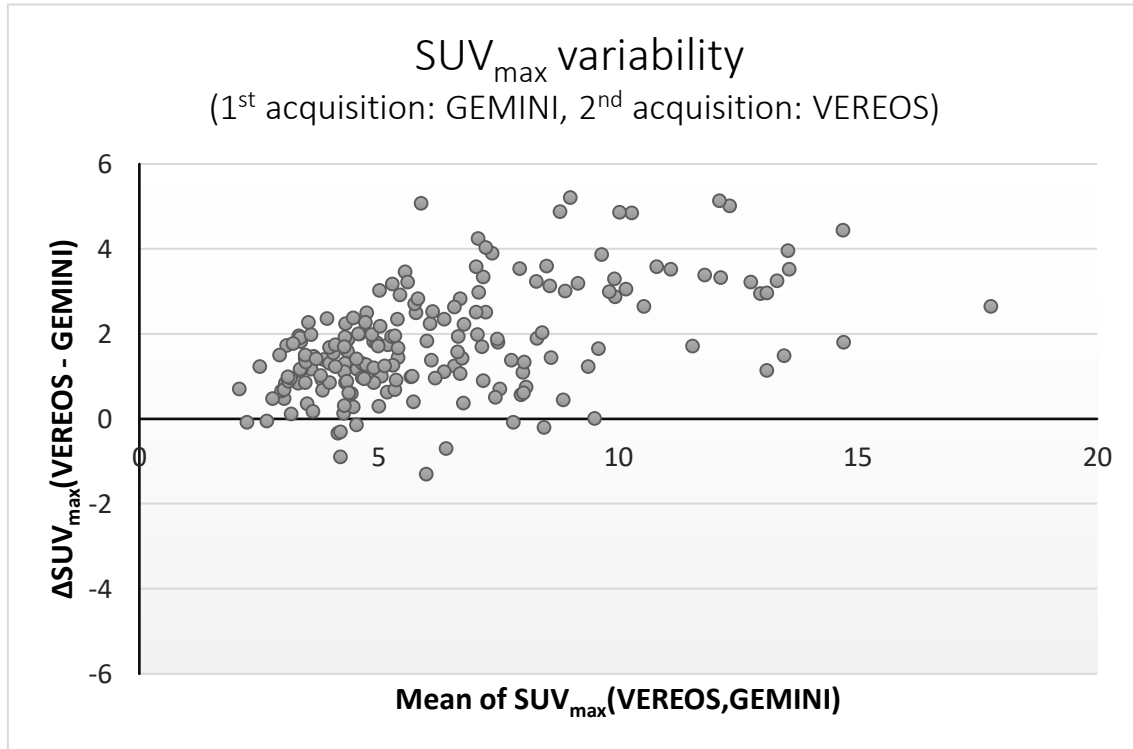

Figure 5.11: Bland and Altman plot representing the variability of SUV_{max} between first acquisition in GEMINI and second acquisition in VEREOS. GEMINI images were reconstructed with standard reconstruction and VEREOS images with a Gaussian post-reconstruction filter of 3 mm. In the x-axis is the mean SUV_{max} between acquisitions and in the y-axis is the SUV_{max} difference between second acquisition (VEREOS) and first acquisition (GEMINI). N=195 lesions.

51

Also, with the purpose of fulfill EARL accreditation specifications, it was performed the same comparison as before, but now with a post-reconstruction Gaussian filter with FWHM of 5 mm in VEREOS images, as recommend by Philips Healthcare to satisfy EARL specifications. Once again, the same statistics done in the previous tables are now in the table 5.12.

Table 5.12: Absolute agreement assessment of lesion features extracted from ^{18}F-FDG PET images acquired in GEMINI with standard reconstruction and for VEREOS with an additional post-reconstruction Gaussian filter of 5 mm (IQR = Interquartile range; ICC = Intraclass correlation coefficient).

	FEATURES	GEMINI (standard reconstruction)		VEREOS (with Gaussian filter of 5 mm)		Pearson Correlation	ICC**
		Median	IQR	Median	IQR		
FIRST ORDER STATIS-TICS	Energy	538.6	1869	580.1	1868	0.93	**0.93**
	Entropy	1.864	1.012	1.874	0.960	0.91	**0.91**
	Kurtosis	2.892	0.944	2.870	1.064	0.58	**0.58**
	Maximum (SUV$_{max}$)	4.863	3.839	4.978	3.375	0.91	**0.90**
	Mean (SUV$_{mean}$)	3.316	1.475	3.565	1.467	0.87	**0.86**
	Mean Absolute Deviation	0.435	0.594	0.456	0.541	0.91	**0.91**
	Median	3.199	1.332	3.393	1.404	0.85	**0.84**
	Minimum	2.515	0.826	2.663	0.917	0.71	**0.70**
	Range	2.169	3.192	2.354	2.908	0.91	**0.90**
	Root Mean Square	3.384	1.631	3.620	1.501	0.88	**0.88**
	Skewness	0.864	0.413	0.848	0.446	0.46	**0.46**
	Standard Deviation	0.545	0.730	0.580	0.649	0.91	**0.90**
	Uniformity	0.175	0.161	0.173	0.154	0.83	**0.82**

	Variance	0.297	1.034	0.336	0.921	0.88	**0.87**
	Coefficient of Variation	0.169	0.145	0.180	0.139	0.85	**0.85**
	$SUV_{peak(3x3x3)}$	3.480	2.661	3.621	2.474	0.93	**0.93**
	$SUV_{peak(2x2x2)}$	4.145	3.168	4.282	2.911	0.92	**0.92**
GEOMETRIC FEATURES	Surface Area	1728	2592	1696	2784	0.97	**0.97**
	Metabolic Tumor Volume (MTV)	3200	6848	3136	7296	0.97	**0.98**
	Compactness1	12.31	10.76	12.34	11.58	0.97	**0.97**
	Compactness2	0.235	0.124	0.247	0.127	0.77	**0.77**
	Spherical disproportion	1.621	0.316	1.593	0.302	0.94	**0.94**
	Sphericity	0.617	0.114	0.628	0.112	0.87	**0.86**
	Surface to volume ratio	0.545	0.206	0.553	0.222	0.83	**0.82**
	Total Lesion Glycolysis (TLG)*	10514	26855	9947	28411	0.97	**0.97**

* TLG does not belong exclusively to geometric features since it relies on first order statistics too.
** A *p-value* < 0.001 was achieved for all calculated ICC and Pearson correlation coefficient.

Once again, the variability of SUV_{max} between acquisitions was assessed. The results are in figures 5.12 and 5.13. In both cases, the SUV_{max} was statistically higher in the second acquisition comparatively to the first acquisition (Wilcoxon test, *p-value* < 0.001).
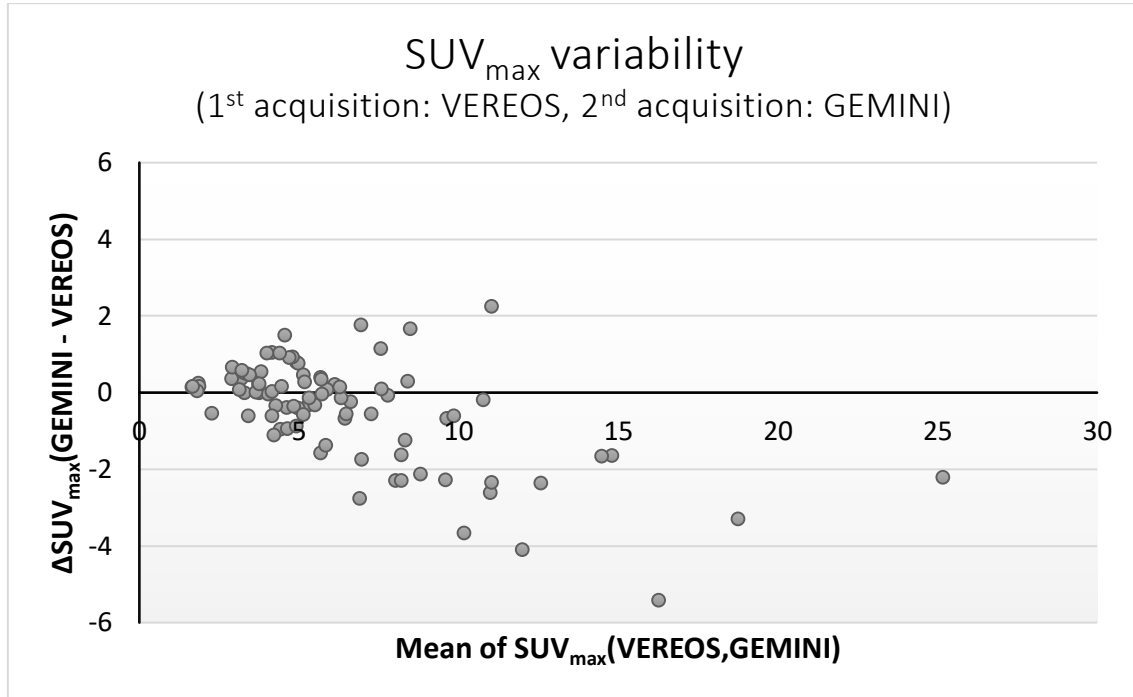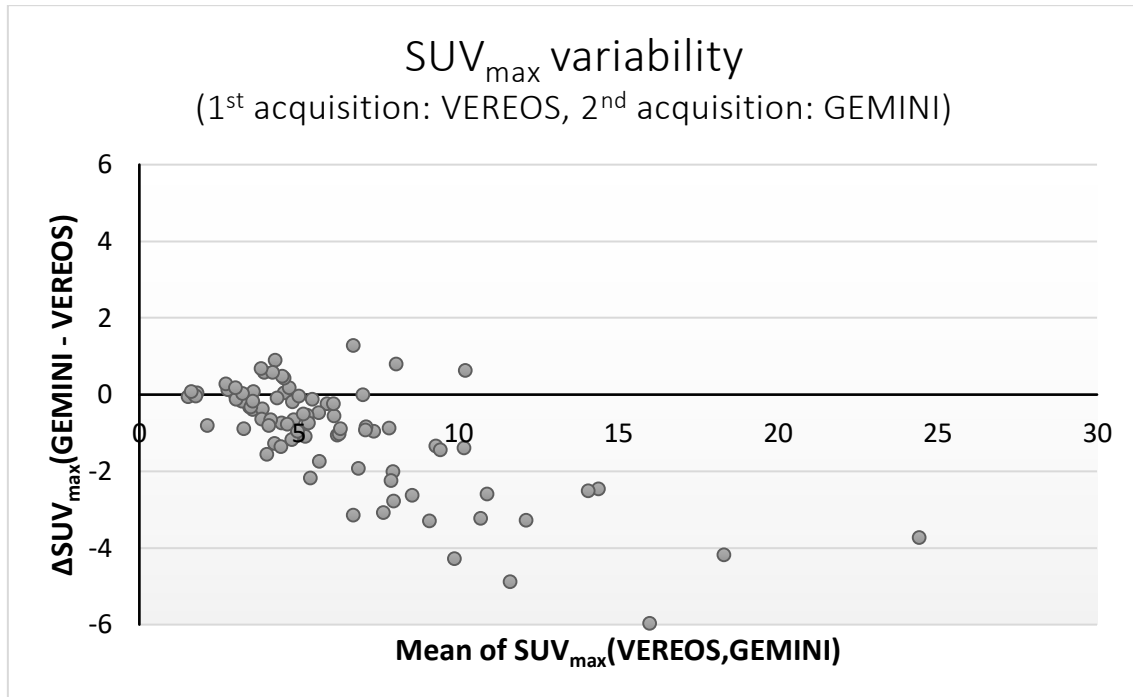
Figure 5.12: Bland and Altman plot representing the variability of SUV_{max} between first acquisition in VEREOS and second acquisition in GEMINI. Both reconstructed images satisfied EARL specifications according to Philips recommendations. In the x-axis is the mean SUV_{max} between acquisitions and in the y-axis is the SUV_{max} difference between second acquisition (GEMINI) and first acquisition (VEREOS). N=88 lesions.
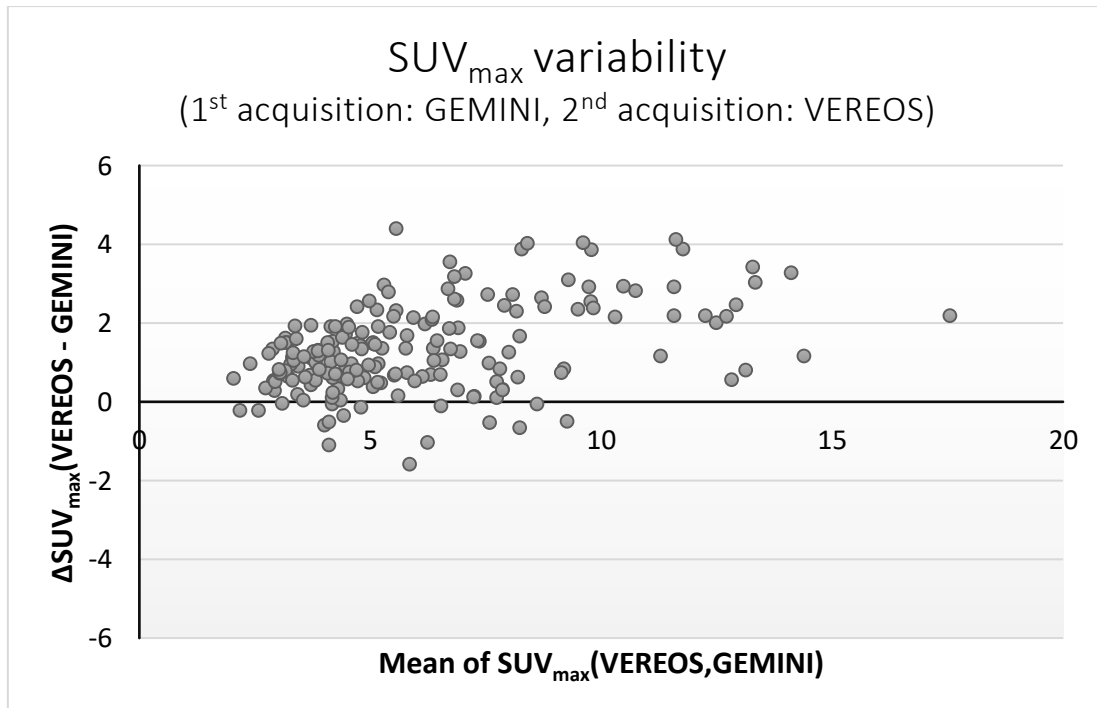

Figure 5.13: Bland and Altman plot representing the variability of SUV_{max} between first acquisition in GEMINI and second acquisition in VEREOS. Both reconstructed images satisfied EARL specifications according to Philips recommendations. In the x-axis is the mean SUV_{max} between acquisitions and in the y-axis is the SUV_{max} difference between second acquisition (VEREOS) and first acquisition (GEMINI). N=195 lesions.
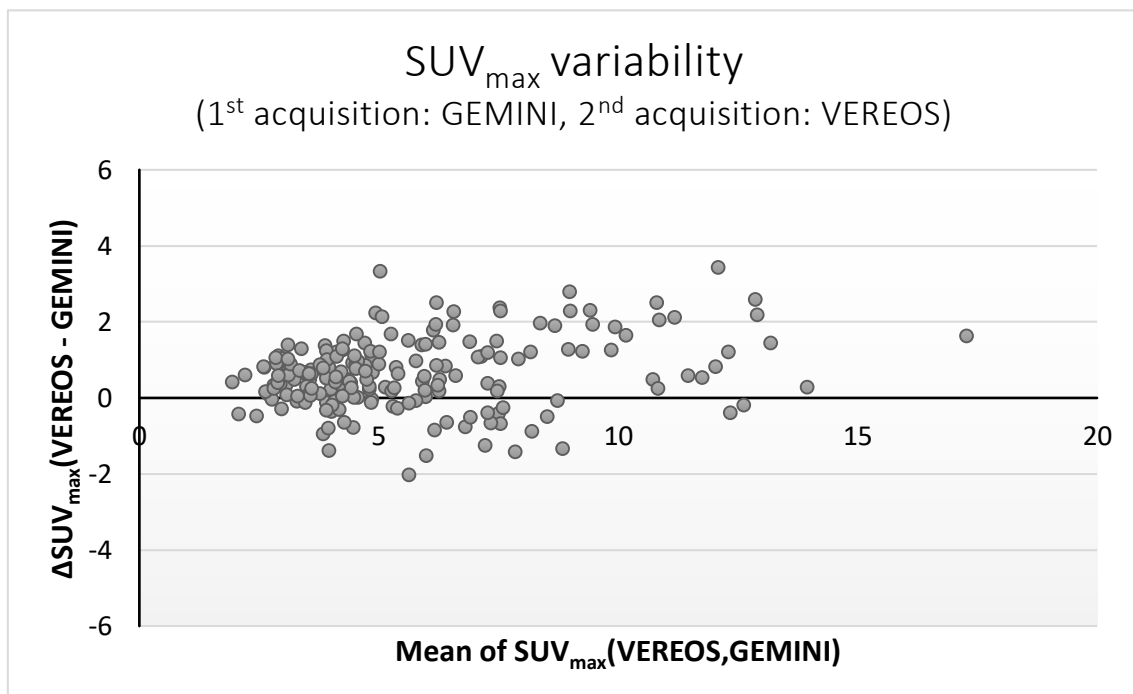
54

DISCUSSION AND CONCLUSIONS

In the present study, ^{18}F-FDG PET/CT images were acquired and reconstructed with default clinical protocols, for both VEREOS and GEMINI, as recommended by the manufacturer. The images acquired in GEMINI followed the instructions provided by EANM research Ltd (EARL) procedures [12]. VEREOS standard reconstructions protocols do not fulfill EARL specifications unless a Gaussian post-reconstruction smoothing filter is applied. A PET with DPC technology shows recovery coefficients above EARL requirements [14]. In this study is being compared two protocols defined by default (VEREOS: 4x4x4 mm^3, OSEM 3 iterations and 15 subsets, relaxation factor 1.0; GEMINI: 4x4x4 mm^3, OSEM 3 iterations and 33 subsets, relaxation factor 0.7). Two additional analyses were also done with VEREOS reconstruction protocols that fulfill the EARL standards.

In the initial stage of this work, several attempts were made to transpose the identification of the lesions done by the nuclear physician in one scan to the scan acquired in the other equipment. This was done using image registration algorithms that included linear and deformable registration strategies [74]. The idea was to find the optimal geometric transformation that aligns both patient's images and then apply that transformation to the lesion identified by the physician. The results of this technique were not encouraging for this study, mainly in small lesions. For this reason, a manual transposition of the annotations completed in VEREOS to GEMINI images had to be executed.

As manual segmentation is not feasible in large datasets due to the time needed to accomplish the task, several automatic segmentation procedures were tried. The results achieved for the threshold method with a cut-off of 50% and 80% were not adequate for this kind of study. For K-means and Bayesian segmentation methods were created an optimized process. In order to

validate the automatic segmentation, two observers segmented manually several lesions. If the agreement between two observers is similar to the agreement of automatic segmentation with manual segmentation, then automatic segmentation can be computed with reliability. Dice coefficient between observer's segmentation was 0.84 (median) and 0.83 (median) between manual and automatic segmentation (Bayesian optimized) without statistically significant differences between them. Based on these segmentation results, the Bayesian optimized method was applied to all lesions in the images acquired in both scanners, VEREOS and GEMINI.

Finally, lesion features usually used in clinics and research based on first order statistics and geometry were extracted (25 features). The most important is to access the agreement between these features extracted from the images of both equipment (GEMINI and VEREOS). ICC represents the absolute agreement between features. In literature is suggested that ICC values less than 0.5 represent poor reliability, values between 0.5 and 0.75 are representative of moderate reliability, values between 0.75 and 0.90 indicate good reliability and values greater than 0.90 show excellent reliability [75]–[77]. Under such conditions, and with images reconstructed with standard protocol, a good or excellent agreement (ICC > 0.75) was demonstrated for most of the features, both in the first order statistics and those related to geometry. Energy, $SUV_{peak2x2x2}$, surface area, MTV, compacteness1, spherical disproportion, and TLG revealed an excellent agreement (ICC > 0.90). Skewness showed to be an unreliable parameter (ICC ≈ 0.43), even though it exhibited a statistically significant ICC ($p < 0.001$), as all the remaining features. Kurtosis, minimum SUV, uniformity and compactness were the ones that demonstrated moderate reliability (0.50 ≤ ICC ≤ 0.75). Particularly, SUV_{max}, the most frequently used feature in the clinics, showed high agreement (ICC ≈ 0.83). On average, this feature was statistically higher in VEREOS than in the GEMINI (95% confidence interval: 0.96-1.35).

VEREOS images reconstructed with an additional post-reconstruction Gaussian smoothing filter in order to fulfill EARL requirements were also compared with GEMINI images reconstructed with default parameters. According to the achievements presented in [14], to meet EARL standards a Gaussian filter with a kernel width of 3 mm was applied. Yet, considering Philips guidelines, a Gaussian filter with FWHM of 5 mm should be applied to fulfill EARL specifications. Then, both filters were used. Naturally, the ICC achieved between the features were equal or superior using the Gaussian post-reconstruction filter than using the standard reconstruction, especially for a FWHM of 5 mm (ICC > 0.80 for most of the features). Nevertheless, kurtosis, minimum SUV and skewness showed an ICC ≤ 0.70, even with these reconstruction parameters. Very likely, the reason for that is not related to differences between equipment, but the lack of reliability of these lesional features. So, these features are not adequate for lesions characterization in ^{18}F-FDG PET images.

Results have demonstrated high reliability among the lesional features, especially when the reconstruction protocols fulfill the EARL standards. Thus, conversion factors between the features extracted from the images acquired in both scanners are not necessary.

As SUV_{max} is the most used feature by the nuclear medicine physicians in routine practice or clinical trials, a particular analysis of the variability of this measure between acquisitions in VEREOS and GEMINI was done. Using the VEREOS standard reconstruction, the overall lesions SUV_{max} in VEREOS images was significantly higher than in GEMINI, because of the superior recovery coefficients in VEREOS images [14]. But, when the VEREOS images were reconstructed accordingly the EARL standards, the difference vanished.

The lesions SUV_{max} extracted from the second scan was statistically higher than the lesion SUV_{max} extracted from the first scan, when VEREOS images were reconstructed with EARL standards, independently of the order of the scans (figure 5.10 – 5.13). This is due to the effect of time in ^{18}F-FDG lesions uptake.

To summarize, there is a high agreement between most of the first order and geometric features extracted from the lesions, especially when the reconstruction protocols applied to both equipment fulfill the EARL standards. Thus, it is feasible to extrapolate that, under controlled acquisition and reconstruction parameters, these features can be used in multi-scanner studies (including longitudinal patient evaluation), using VEREOS and GEMINI PET/CT devices.

6.1 Limitations

This study faced some limitations that could lead to differences between the features extracted from both devices, GEMINI and VEREOS. These limitations should be considered and are summarized below:

A) According to the European Association of Nuclear Medicine (EANM) guidelines, the suggested whole-body PET/CT uptake time (min post-injection of ^{18}F-FDG) is 60 min, with an acceptable range of 55-75 min [4], [31]. The effect of the time interval on SUV in healthy organs is relatively small [78], but in lesions, there is a significant improvement in target-to-background ratio when acquiring the PET image in an approximate 90 min window [79]. On average, the interval between ^{18}F-FDG patient's administration and the start of scanning was 70 min, and the time between acquisitions was 40 min (mean). All data were acquired in a clinical environment. Consequently, there was a notable variation caused by dual imaging studies and external workflow factors. Summing up, the time difference between acquisitions should have been smaller in some acquisitions, but it was the shortest possible achieved in the clinic. In addition, in few acquisitions the difference time exceeded the acceptable.

B) Even though the lesions segmentation algorithm performed similarly to human observers, we may assume that other segmentation methods could lead to different results.

C) Although the patients were laid down approximately in the same position in both scans, the shape of the body changes from one scan to another, and consequently the shape of the lesions.

D) The same lesions were comprised in the ROI drawn by the physician in GEMINI and VEREOS images. Naturally, the surrounding background region included is different. Depending on the lesion location, if the background in the ROI is not enough, the automatic segmentation may be compromised. An example of this issue is the outlier obtained in the boxplot of figure 5.7.

6.2 Future Work

For future perspectives, there are several projects that can be conducted:

- Using images from VEREOS and GEMINI, identical projects should be studied using different reconstruction protocols:

> **1.** In the future, when EARL2 specifications are totally implemented, images from VE-REOS should be reconstructed based on the new EARL guidelines. A reproducibility study between EARL2 VEREOS reconstructions and default GEMINI images reconstruction can be concluded.
>
> **2.** GEMINI images can be reconstructed to obtain sharper images, closer to VEREOS definitions. Yet, it is not possible to obtain images equal to the ones from VEREOS due to the capability limitations of the analog device.
>
> **3.** VEREOS allows whole-body reconstruction with voxels of 2 x 2 x 2 mm^3. It would be interesting to evaluate how different the image features will be reconstructing images from VEREOS with 2 mm and GEMINI with 4 mm.

- Instead of lesions, it can be measured the uptake distribution of healthy organs.
- Add texture analysis. This is only feasible for large lesions (at least 64 voxels [80]).
- Compare the number of lesions identifiable in VEREOS and GEMINI. Novel techniques implemented in VEREOS improve small lesions detectability [81]. Thus, in order to test if there are differences in the number of lesions detectable in VEREOS and GEMINI images, a study in which various nuclear medicine physicians identify lesions independently in VEREOS and GEMINI will be important.
- Development of an automatic identification of the lesions. This task is a huge challenge because, sometimes, the lesion uptake is slightly higher than the surrounding region, the

lesion is next to an organ with high uptake or may exist several foci with high uptake that are no lesions.

BIBLIOGRAPHY

[1] M. S. Hofman and R. J. Hicks, "How We Read Oncologic FDG PET/CT," *Cancer Imaging*, vol. 16, no. 1, p. 35, 2016, doi:10.1186/s40644-016-0091-3.

[2] M. G. Vander, L. C. Cantley, and C. B. Thompson, "Understanding the Warburg Effect : The Metabolic Requirements of Cell Proliferation," *Science (80-.).*, vol. 324, no. 5930, pp. 1029–1033, 2009, doi:10.1126/science.1160809.

[3] A. Zhu, D. Lee, and H. Shim, "Metabolic Positron Emission Tomography Imaging in Cancer Detection and Therapy Response," *Semin. Oncol.*, vol. 38, no. 1, pp. 55–69, 2011, doi:10.1053/j.seminoncol.2010.11.012.

[4] R. Boellaard *et al.*, "FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 42, no. 2, pp. 328–354, 2015, doi:10.1007/s00259-014-2961-x.

[5] L. K. Griffeth, "Use of Pet/Ct Scanning in Cancer Patients: Technical and Practical Considerations," *Baylor Univ. Med. Cent. Proc.*, vol. 18, no. 4, pp. 321–330, 2005, doi:10.1080/08998280.2005.11928089.

[6] M. D. Seemann, "Whole-Body PET/MRI: The Future in Oncological Imaging," *Technol. Cancer Res. Treat.*, vol. 4, no. 5, pp. 577–582, 2005, doi:10.1177/153303460500400512.

[7] S. J. Rosenbaum, T. Lind, G. Antoch, and A. Bockisch, "False-Positive FDG PET Uptake−the Role of PET/CT," *Eur. Radiol.*, vol. 16, no. 5, pp. 1054–1065, 2006, doi:10.1007/s00330-005-0088-y.

[8] B. A. Siegel and F. Dehdashti, "Oncologic PET/CT: current status and controversies," *Eur. Radiol. Suppl.*, vol. 15, no. S4, pp. d127–d132, 2005, doi:10.1007/s10406-005-0116-7.

[9] E. Roncali and S. R. Cherry, "Application of Silicon Photomultipliers to Positron Emission Tomography," *Ann. Biomed. Eng.*, vol. 39, no. 4, pp. 1358–1377, 2011, doi:10.1007/s10439-011-0266-9.

[10] R. Mirzoyan, M. Laatiaoui, and M. Teshima, "Very high quantum

efficiency PMTs with bialkali photo-cathode," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 567, no. 1, pp. 230–232, 2006, doi:10.1016/j.nima.2006.05.094.

[11] N. C. Nguyen *et al.*, "Image Quality and Diagnostic Performance of a Digital PET Prototype in Patients with Oncologic Diseases: Initial Experience and Comparison with Analog PET," *J. Nucl. Med.*, vol. 56, no. 9, pp. 1378–1385, 2015, doi:10.2967/jnumed.114.148338.

[12] R. Boellaard *et al.*, "FDG PET and PET/CT : EANM procedure guidelines for tumour PET imaging : version 1.0," *Eur J Nucl Med Mol Imaging*, vol. 37, pp. 181–200, 2010, doi:10.1007/s00259-009-1297-4.

[13] *NEMA NU 2-2012. Performance Measurements of Positron Emission Tomographs*. National Electrical Manufacturers Association, 2012.

[14] D. Koopman *et al.*, "Digital PET compliance to EARL accreditation specifications," *EJNMMI Phys.*, vol. 4, no. 9, pp. 1–6, 2017, doi:10.1186/s40658-017-0176-5.

[15] M. Hatt, F. Tixier, L. Pierce, P. E. Kinahan, C. C. Le Rest, and D. Visvikis, "Characterization of PET/CT images using texture analysis: the past, the present… any future?," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 44, no. 1, pp. 151–165, 2017, doi:10.1007/s00259-016-3427-0.

[16] I. N. Bankman, T. Nizialek, I. Simon, O. B. Gatewood, I. N. Weinberg, and W. R. Brody, "Segmentation algorithms for detecting microcalcifications in mammograms," *IEEE Trans. Inf. Technol. Biomed.*, vol. 1, no. 2, pp. 141–149, 1997, doi:10.1109/4233.640656.

[17] G. V Saradhi *et al.*, "A Framework for Automated Tumor Detection in Thoracic FDG PET Images Using Texture-Based Features," in *2009 IEEE International Symposium on Biomedical Imaging*, 2009, pp. 97–100.

[18] Z. Zhang, W. V. Stoecker, and R. H. Moss, "Border detection on digitized skin tumor images," *IEEE Trans. Med. Imaging*, vol. 19, no. 11, pp. 1128–1143, 2000, doi:10.1109/42.896789.

[19] Z. Ma, J. M. R. S. Tavares, R. N. Jorge, and T. Mascarenhas, "A review of algorithms for medical image segmentation and their applications to the female pelvic cavity," *Comput. Methods Biomech. Biomed. Engin.*, vol. 13, no. 2, pp. 235–246, 2010, doi:10.1080/10255840903131878.

[20] W. Jentzen, L. Freudenberg, E. G. Eising, M. Heinze, W. Brandau, and A. Bockisch, "Segmentation of PET volumes by iterative image thresholding." *J. Nucl. Med.*, vol. 48, no. 1, pp. 108–14, 2007.

[21] F. Gallivanone, F. Fazio, L. Presotto, M. C. Gilardi, C. Canevari, and I. Castiglioni, "Adaptive threshold method based on PET measured lesion-to-background ratio for the estimation of Metabolic Target Volume from

18F-FDG PET images," in *2013 IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2013, pp. 1–7, doi:10.1109/NSSMIC.2013.6829383.

[22] Y. E. Erdi *et al.*, "Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding," *Cancer*, vol. 80, no. 12 Suppl, pp. 2505–9, Dec. 1997.

[23] U. Ilhan and A. Ilhan, "Brain tumor segmentation based on a new threshold approach," *Procedia Comput. Sci.*, vol. 120, pp. 580–587, 2017, doi:10.1016/j.procs.2017.11.282.

[24] M. Hatt, C. Cheze, A. Turzo, C. Roux, D. Visvikis, and I. T. Oncology, "A Fuzzy Locally Adaptive Bayesian Segmentation Approach for Volume Determination in PET," *IEEE Trans. Med. Imaging*, vol. 28, no. 6, pp. 881–893, 2009.

[25] S. Mercieca, J. Belderbos, J. van Loon, K. Gilhuijs, P. Julyan, and M. van Herk, "Comparison of SUVmax and SUVpeak based segmentation to determine primary lung tumour volume on FDG PET-CT correlated with pathology data," *Radiother. Oncol.*, vol. 129, no. 2, pp. 227–233, 2018, doi:10.1016/j.radonc.2018.06.028.

[26] W. Ju, D. Xiang, B. Zhang, L. Wang, I. Kopriva, and X. Chen, "Random Walk and Graph Cut for Co-Segmentation of Lung Tumor on PET-CT Images," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5854–5867, 2015, doi:10.1109/TIP.2015.2488902.

[27] M. Piert *et al.*, "Accuracy of tumor segmentation from multi-parametric prostate MRI and 18F-choline PET/CT for focal prostate cancer therapy applications," *EJNMMI Res.*, vol. 8, no. 1, p. 23, 2018, doi:10.1186/s13550-018-0377-5.

[28] K. G. Foley *et al.*, "Development and validation of a prognostic model incorporating texture analysis derived from standardised segmentation of PET in patients with oesophageal cancer," *Eur. Radiol.*, vol. 28, no. 1, pp. 428–436, 2017, doi:10.1007/s00330-017-4973-y.

[29] B. Abualhaj *et al.*, "Comparison of five cluster validity indices performance in brain [18 F]FET-PET image segmentation using k -means," *Med. Phys.*, vol. 44, no. 1, pp. 209–220, 2017, doi:10.1002/mp.12025.

[30] A. Juslin and J. Tohka, "Unsupervised Segmentation of Cardiac PET Transmission Images for Automatic Heart Volume Extraction," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 1077–1080, doi:10.1109/IEMBS.2006.259416.

[31] R. L. Wahl, H. Jacene, Y. Kasamon, and M. A. Lodge, "From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors," *J. Nucl. Med.*, vol. 50, no. 5, pp. 122–150, 2009,

doi:10.2967/jnumed.108.057307.

[32] B. Huang, T. Chan, D. L.-W. Kwong, W. K. S. Chan, and P.-L. Khong, "Nasopharyngeal Carcinoma: Investigation of Intratumoral Heterogeneity With FDG PET/CT," *Am. J. Roentgenol.*, vol. 199, no. 1, pp. 169–174, 2012, doi:10.2214/AJR.11.7336.

[33] N.-M. Cheng *et al.*, "Textural Features of Pretreatment 18F-FDG PET/CT Images: Prognostic Significance in Patients with Advanced T-Stage Oropharyngeal Squamous Cell Carcinoma," *J. Nucl. Med.*, vol. 54, no. 10, pp. 1703–1709, 2013, doi:10.2967/jnumed.112.119289.

[34] S.-C. Chan *et al.*, "Tumor heterogeneity measured on F-18 fluorodeoxyglucose positron emission tomography/computed tomography combined with plasma Epstein-Barr Virus load predicts prognosis in patients with primary nasopharyngeal carcinoma," *Laryngoscope*, vol. 127, no. 1, 2016, doi:10.1002/lary.26172.

[35] T. Kitao *et al.*, "Reproducibility and uptake time dependency of volume-based parameters on FDG-PET for lung cancer," *BMC Cancer*, vol. 16, no. 1, p. 576, Dec. 2016, doi:10.1186/s12885-016-2624-3.

[36] R. Ossenkoppele *et al.*, "Discriminative Accuracy of [18 F]flortaucipir Positron Emission Tomography for Alzheimer Disease vs Other Neurodegenerative Disorders," *JAMA*, vol. 320, no. 11, p. 1151, 2018, doi:10.1001/jama.2018.12917.

[37] B. N. Ganguly, N. N. Mondal, M. Nandy, and F. Roesch, "Some physical aspects of positron annihilation tomography: A critical review," *J. Radioanal. Nucl. Chem.*, vol. 279, no. 2, pp. 685–698, Feb. 2009, doi:10.1007/s10967-007-7256-2.

[38] J. L. Humm, A. Rosenfeld, and A. Del Guerra, "From PET detectors to PET scanners," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 30, no. 11, pp. 1574–1597, Nov. 2003, doi:10.1007/s00259-003-1266-2.

[39] H.-H. Lin, K.-S. Chuang, C.-C. Lu, Y.-C. Ni, and M.-L. Jan, "Use of beam stoppers to correct random and scatter coincidence in PET: A Monte Carlo simulation," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 711, pp. 27–37, 2013, doi:10.1016/j.nima.2012.12.124.

[40] B. Furlow, "PET-CT Cancer Imaging," vol. 90, no. 2, 2018.

[41] T. Frach, G. Prescher, C. Degenhardt, R. De Gruyter, A. Schmitz, and R. Ballizany, "The Digital Silicon Photomultiplier – Principle of Operation and Intrinsic Detector Performance," in *2009 IEEE Nuclear Science Symposium Conference Record*, 2009, pp. 1959–1965.

[42] C. Degenhardt *et al.*, "Performance evaluation of a prototype Positron

Emission Tomography scanner using Digital Photon Counters (DPC)," in *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record*, 2012, pp. 2820–2824, doi:10.1109/NSSMIC.2012.6551643.

[43]   C. Dege *et al.*, "The digital Silicon Photomultiplier — A novel sensor for the detection of scintillation light," in *2009 IEEE Nuclear Science Symposium Conference Record*, 2009, pp. 2383–2386, doi:10.1109/NSSMIC.2009.5402190.

[44]   N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognit.*, vol. 26, no. 9, pp. 1277–1294, 1993, doi:10.1016/0031-3203(93)90135-J.

[45]   D. L. Pham, C. Xu, and J. L. Prince, "Current methods in Medical Image Segmentation," *Annu. Rev. Biomed. Eng.*, 2000.

[46]   P. H. Jarritt, K. J. Carson, A. R. Hounsell, and D. Visvikis, "The role of PET/CT scanning in radiotherapy planning," *Br. J. Radiol.*, vol. 79, pp. S27–S35, 2006, doi:10.1259/bjr/35628509.

[47]   N. C. Krak, R. Boellaard, O. S. Hoekstra, J. W. R. Twisk, C. J. Hoekstra, and A. A. Lammertsma, "Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 32, no. 3, pp. 294–301, 2005, doi:10.1007/s00259-004-1566-1.

[48]   A. C. Paulino, M. Koshy, R. Howell, D. Schuster, and L. W. Davis, "Comparison of CT- and FDG-PET-defined gross tumor volume in intensity-modulated radiotherapy for head-and-neck cancer," *Int. J. Radiat. Oncol.*, vol. 61, no. 5, pp. 1385–1392, 2005, doi:10.1016/j.ijrobp.2004.08.037.

[49]   N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66, 1979, doi:10.1109/TSMC.1979.4310076.

[50]   L. S. Davis, "A survey of edge detection techniques," *Comput. Graph. Image Process.*, vol. 4, no. 3, pp. 248–270, 1975, doi:10.1016/0146-664X(75)90012-X.

[51]   J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 6, pp. 679–698, 1986, doi:10.1109/TPAMI.1986.4767851.

[52]   R. Pohle and K. D. Toennies, "Segmentation of medical images using adaptive region growing," in *Medical Imaging 2001: Image Processing.*, 2001, pp. 1337–1346, doi:10.1117/12.431013.

[53]   J. Dehmeshki, X. Ye, and J. Costello, "Shape based region growing using derivatives of 3D medical images: application to semiautomated detection of pulmonary nodules.," *2003 Int. Conf. Image Process.*, vol. 1, pp. 1085–8, 2003.

[54]   A. Norouzi *et al.*, "Medical Image Segmentation Methods, Algorithms, and Applications," *IETE Tech. Rev.*, vol. 31, no. 3, pp. 199–213, 2014, doi:10.1080/02564602.2014.906861.

[55]   P. A. N. Su, T. Chen, W. Xu, X. Shao, H. Wang, and Y. Zhao, "Dominant-set-based Consensus for Fuzzy C-means Clustering Ensemble," in *2018 International Conference on Machine Learning and Cybernetics (ICMLC)*, 2018, pp. 85–90.

[56]   H. Caillol, W. Pieczynski, and A. Hillion, "Estimation of Fuzzy Gaussian Mixture and Unsupervised Statistical Image Segmentation," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 425–440, 1997.

[57]   T. Mcinerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Med. Image Anal.*, vol. 1, no. 2, pp. 91–108, 1996, doi:10.1016/s1361-8415(96)80007-7.

[58]   Philips, "Technical Reference Guide - Vereos PET/CT." 2018.

[59]   B. Bai, J. Bading, and P. S. Conti, "Tumor Quantification in Clinical Positron Emission Tomography," *Theranostics*, vol. 3, no. 10, 2013, doi:10.7150/thno.5629.

[60]   A. Fedorov *et al.*, "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magn. Reson. Imaging*, vol. 30, no. 9, pp. 1323–1341, 2012, doi:10.1016/j.mri.2012.05.001.

[61]   H. J. Johnson, M. M. Mccormick, and L. Ib, *The ITK Software Guide Book 1: Introduction and Development Guidelines. Fourth Edition*. 2005.

[62]   T. S. Yoo *et al.*, "Engineering and Algorithm Design for an Image Processing API : A Technical Report on ITK - the Insight Toolkit," *Stud. Health Technol. Inform.*, vol. 85, pp. 586–592, 2002.

[63]   P. A. Yushkevich *et al.*, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage*, vol. 31, pp. 1116–1128, 2006, doi:10.1016/j.neuroimage.2006.01.015.

[64]   A. S. Ribeiro, E. R. Kops, H. Herzog, and P. Almeida, "Hybrid approach for attenuation correction in PET/MR scanners," *Nucl. Inst. Methods Phys. Res. A*, vol. 734, pp. 166–170, 2014, doi:10.1016/j.nima.2013.09.034.

[65]   G. Wagenknecht *et al.*, "Attenuation Correction in MR-BrainPET with Segmented T1-weighted MR images of the Patient's Head - A Comparative Study with CT," in *2011 IEEE Nuclear Science Symposium Conference Record*, 2011, pp. 2261–2266, doi:10.1109/nssmic.2011.6153858.

[66]   K. H. Zou *et al.*, "Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index1," *Radiol. Alliance Heal. Serv. Res.*, vol. 11,

no. 2, pp. 178–189, 2004, doi:10.1016/S1076-6332(03)00671-8.

[67] H. J. W. L. Aerts *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nat. Commun.*, vol. 5, p. 4006, 2014, doi:10.1038/ncomms5006.

[68] A. Sher *et al.*, "For avid glucose tumors , the SUV peak is the most reliable parameter for [18F]FDG-PET/CT quantification , regardless of acquisition time," *EJNMMI Res.*, pp. 4–9, 2016, doi:10.1186/s13550-016-0177-8.

[69] E. Woff, A. Hendlisz, L. Ameye, C. Garcia, and T. Kamoun, "Metabolic Active Tumor Volume and Total Lesion Glycolysis by F- FDG PET / CT Validated as Prognostic Imaging Biomarkers in Chemorefractory Metastatic Colorectal Cancer," *J. Nucl. Med.*, 2018, doi:10.2967/jnumed.118.210161.

[70] G. Kuhnert *et al.*, "Impact of PET/CT image reconstruction methods and liver uptake normalization strategies on quantitative image analysis," *Eur J Nucl Med Mol Imaging*, vol. 43, no. 2, pp. 249–258, 2015, doi:10.1007/s00259-015-3165-8.

[71] E. M. Blanchet *et al.*, "18F-FDG PET/CT as a predictor of hereditary head and neck paragangliomas," *Eur. J. Clin. Invest.*, vol. 44, pp. 325–332, 2014, doi:10.1111/eci.12239.

[72] N. G. Mikhaeel *et al.*, "Combination of baseline metabolic tumour volume and early response on PET/CT improves progression-free survival prediction in DLBCL," *Eur J Nucl Med Mol Imaging*, vol. 43, no. 7, pp. 1209–1219, 2016, doi:10.1007/s00259-016-3315-7.

[73] A. Archier *et al.*, "18F-DOPA PET/CT in the diagnosis and localization of persistent medullary thyroid carcinoma," *Eur J Nucl Med Mol Imaging*, vol. 43, no. 6, pp. 1027–1033, 2015, doi:10.1007/s00259-015-3227-y.

[74] F. P. M. Oliveira and J. M. R. S. Tavares, "Medical image registration : a review," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 17, no. 2. pp. 73–93, 2014, doi:10.1080/10255842.2012.670855.

[75] T. K. Koo and M. Y. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research," *J. Chiropr. Med.*, vol. 15, no. 2, pp. 155–163, 2015, doi:10.1016/j.jcm.2016.02.012.

[76] J. P. Eir, "Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM," *J. Strength Cond. Res.*, vol. 19, no. 1, pp. 231–240, 2005, doi:10.1519/00124278-200502000-00038.

[77] S. Qin, L. Nelson, L. Mcleod, S. Eremenco, and S. Joel, "Assessing test–retest reliability of patient-reported outcome measures using intraclass correlation coefficients : recommendations for selecting and documenting the analytical formula," *Qual. Life Res.*, vol. 28, pp. 1029–1033, 2019,

doi:10.1007/s11136-018-2076-0.

[78] R. Wang, H. Chen, and C. Fan, "Impacts of time interval on 18F-FDG uptake for PET/CT in normal organs," *Medicine (Baltimore).*, vol. 97, no. 45, p. e13122, 2018.

[79] Z. Al-Faham, P. Jolepalem, J. Rydberg, and C.-Y. O. D. Wong, "Optimizing 18F-FDG uptake time prior to imaging improves the accuracy of PET/CT in liver lesions," *J. Nucl. Med. Technol.*, vol. 44, no. 2, pp. 70–72, 2016, doi:10.2967/jnmt.115.169953.

[80] F. Orlhac, C. Nioche, and I. Buvat, *Texture — User Guide*. 2019.

[81] C. S. Van Der Vos, D. Koopman, S. Rijnsdorp, A. T. M. Willemsen, and E. P. Visser, "Quantification , improvement , and harmonization of small lesion detection with state-of-the-art PET," *Eur J Nucl Med Mol Imaging*, vol. 44, pp. 4–16, 2017, doi:10.1007/s00259-017-3727-z.

APPENDIX

In the lesion's identification step, the lesions were detected in VEREOS images and a ROI surrounding the lesions was drawn. When transposing that information to GEMINI images, a few lesions were not identifiable. For that reason, and after being reviewed by the nuclear physician, these lesions were excluded from the dataset. The image that follows (figure A.1) shows one of these questionable cases.
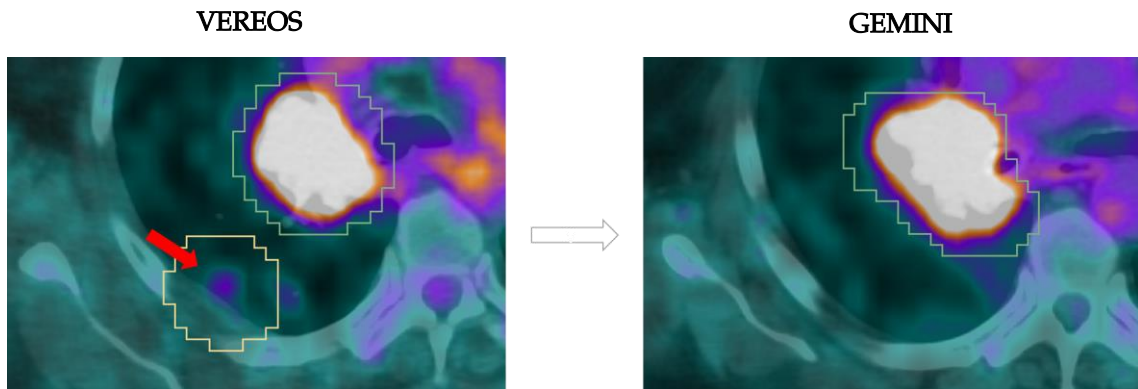

Figure A.1: Lesion identified in an image acquired from VEREOS scanner that is not visible in the corresponding GEMINI image. The lung lesion is identified by a red arrow in the VEREOS mask. In the corresponding GEMINI image, the same lesion is not visible, and so it is not included in the study.

Figures A.2 to A.5 represent the boxplot graphs to better understand data dispersion among the groups: Dice between observers' segmentation; Dice between observer 1 and automatic segmentation; Dice between observer 2 and automatic segmentation. It is presented the results for different automatic segmentation methods used (Threshold, K-means, and Bayesian approach).
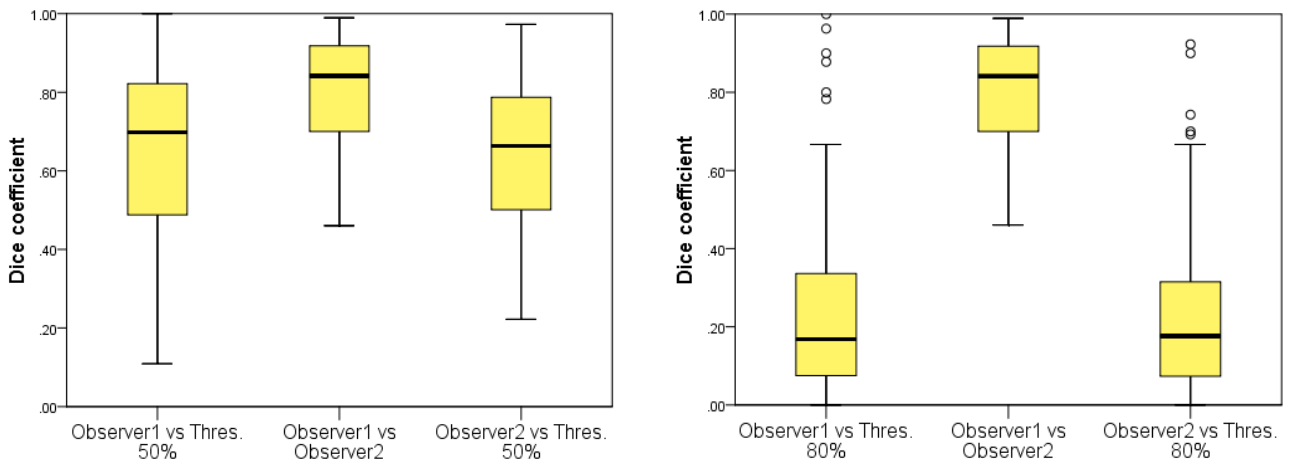
Figure A.2: Boxplot representing Dice coefficient dispersion between manual segmentation and segmentation using threshold method (50% on the left graph and 80% on the right graph).
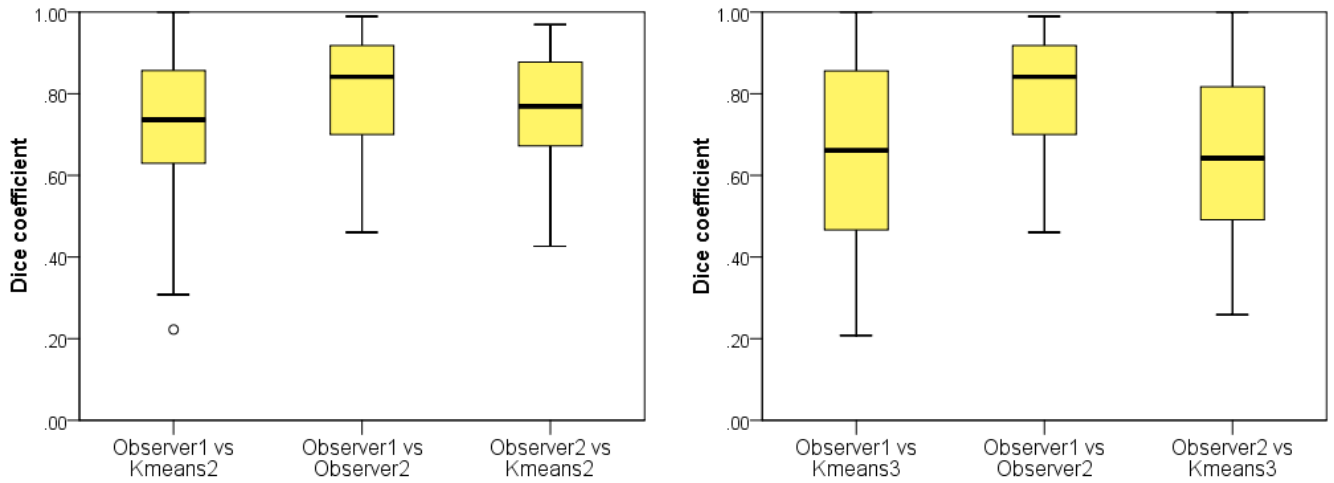

Figure A.3: Boxplot representing Dice coefficient dispersion between manual segmentation and segmentation using K-means method with $n_{classes} = 2$ (left graph) and $n_{classes} = 3$ & merge two lower classes (right graph).
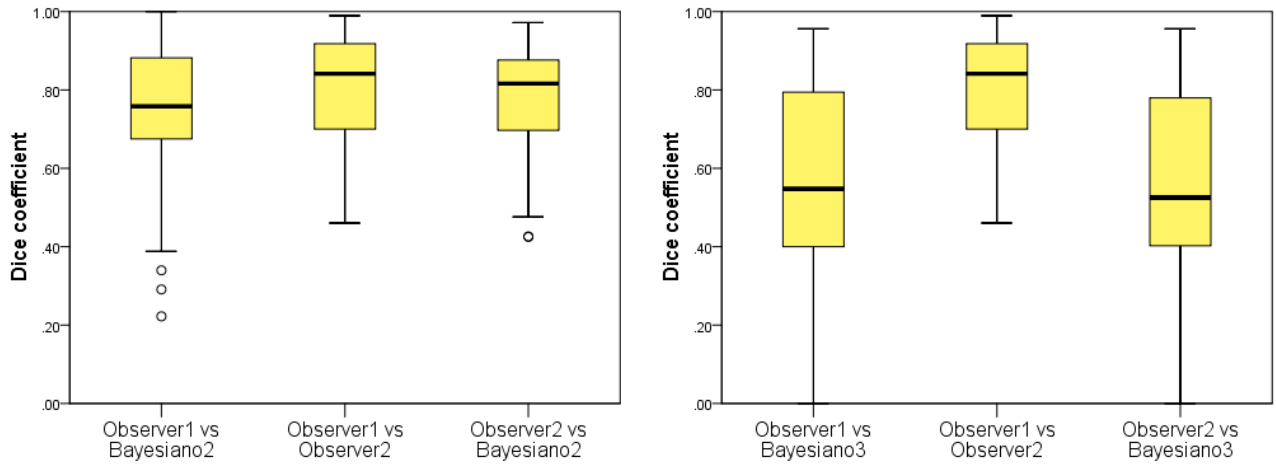
Figure A.4: Boxplot representing Dice coefficient dispersion between manual segmentation and segmentation using Bayesian method with $n_{classes} = 2$ (left graph) and $n_{classes} = 3$ & merge two lower classes (right graph).
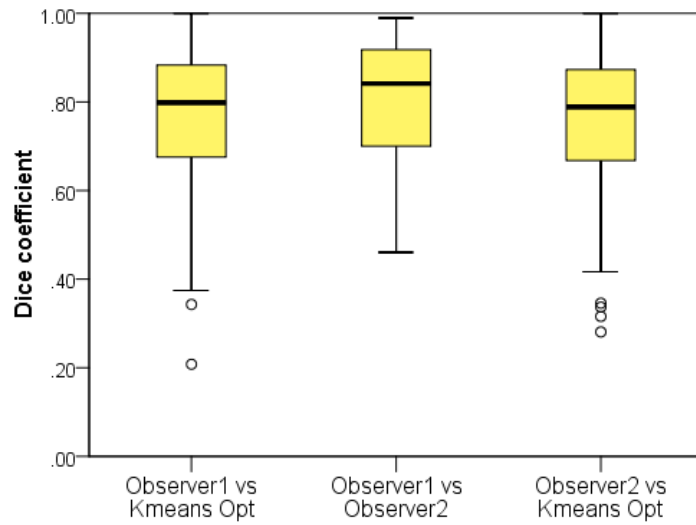

Figure A.5: Boxplot representing Dice coefficient dispersion between manual segmentation and optimized segmentation for K-means method.