UNIVERSITÀ DEGLI STUDI DI PALERMO

Dottorato di Ricerca in Ingegneria Informatica
Dipartimento di Ingegneria Chimica, Gestionale, Informatica, Meccanica
ING-INF/05

Multimodal Image Segmentation for Radiation Treatment Planning

IL DOTTORE
Ing. Alessandro Stefano

IL COORDINATORE
Prof. Salvatore Gaglio

IL TUTOR
Prof. Salvatore Vitabile

CICLO XXVI
ANNO CONSEGUIMENTO TITOLO 2016

Alla mia piccola Elisa.

Abstract

Image processing and analysis techniques find widespread use in clinical environments: information provided by medical imaging has a fundamental impact on patient health and quality of life. The aim of image processing and analysis methodologies is the diagnosis accuracy improvements and the assessment of treatment response by means of quantitative bio-markers in an efficient way. In addition, an early and efficacy prediction of therapy response in patients is a critical issue in the field of personalized medicine.

In the last few years, several criteria to assess therapy response in oncological patients have been proposed, ranging from anatomic to functional assessments. Changes in tumour size are not necessarily correlated with changes in tumour viability and outcome. In addition, morphological changes resulting from therapy occur slower than functional changes.

Nuclear medicine images show interesting characteristic information about the physiological properties of the oncological lesions. In addition, PET (Positron Emission Tomography) imaging allows for early prediction of therapy response after only a few days of treatment. Due to the nature of PET images (low spatial resolution, high noise and weak boundary), metabolic image processing is a critical task to enhance the current methodology performed by healthcare operators in clinical routine.

In this ph.D. Thesis, many issues in a real clinical environment have been addressed and a certain amount of improvements have been produced in the following areas:

- Automatic and real-time delineation of biological tumour volumes
- Automatic evaluation of therapy response in oncological patients
- Multi-modal image delineation of brain metastases in patients who underwent Gamma-Knife treatment.

The proposed contributions have produced scientific publications in indexed computer science and medical journals and conferences. In addition, some home-made tools have been developed to be used as components of Medical Decision Support Systems to enhance the current methodology performed by healthcare operators in clinical routine.

These studies are part of an on-going research collaboration in the area of computer engineering between my Department and colleagues at the:

- Dipartimento di Biopatologia e Biotecnologie Mediche e Forensi (DIBIMED), University of Palermo
- Cannizzaro Hospital of Catania
- Ospedale San Raffaele of Milano
- Ospedale Giglio of Cefalù
- Istituto di Bioimmagini e Fisiologia Molecolare (IBFM) Centro Nazionale delle Ricerche (CNR) of Cefalù

working closely to nuclear medicine and radiologist physicians, radiation therapists, and technical operators, and following the common denominator of automatic and operator independent methods applied to medical imaging issues.

Abbreviations

BTV: Biological Tumour Volume

CT: Computerized Tomography

DICOM: Digital Imaging and Communications in Medicine

FBP: Filtered Back-Projection

FDG: 8F-fluoro-2-deoxy-D-glucose

GTV: Gross Tumour Volume

HNC: Head and Neck Cancer

MET: 11C-labeled Methionine

MRI: Magnetic Resonance Imaging

NSCLC: Non-Small Cell Lung Cancer

PET: Positron Emission Tomography

PVE: Partial Volume Effect

PTV: Planning Treatment Volume

ROI: Region Of Interest

SUV: Standardized Uptake Value

TPS: Treatment Planning Systems

Contents

Chapter 1

Introduction

The medical imaging is a crucial field for diagnostic and treatment purposes. Several medical imaging techniques are available and their improvement is one of the most important goals in the health's research.

1. Medical Imaging

Medical imaging is characterized by great multitude of heterogeneous data enclosing several experts and expertise in different fields, such as physicians, engineers, biologists, and physics for disease staging and treatment purposes.

According to the source and the physical properties, the medical imaging techniques can be differentiated in two groups:

- anatomical/structural images, and
- functional images.

In particular, Positron Emission Tomography (PET), Computerized Tomography (CT), and Magnetic Resonance Imaging (MRI) are *in-vivo* tomographic techniques which allows to show axial, sagittal and coronal images without moving the patient. Furthermore, these techniques can be classified on the use of ionized or non-ionized radiation as well as external or internal sources.

To support clinical decisions is mandatory to develop computer-assisted methods of quantitative biomarker extraction by medical images. To date, the computational power is mature: an effort is mandatory to address this power to analyze and process medical data in the best possible way. Fast and operator independent approaches are the crucial keywords in clinical environment to obtain an effective impact in the work of medical operators involved in diagnosis and treatment assessment.

Developing computer decision support systems that integrate the knowledge of the medical experts, and facilitate the diagnostic procedure are need for a better patient management. In particular, computer-assisted methods for segmenting regions of interests (ROIs) in medical images are increasingly important in assisting and automating specific radiotherapy tasks.

1.1.1 DICOM Protocol

The medical images produced by special diagnostic equipment, such as PET, CT, and MRI, are compliant with the DICOM (Digital Imaging and Communications in Medicine) standard.

DICOM standard was created by National Electrical Manufacturers Association (NEMA) and it enables users to retrieve images and related information from different medical imaging modalities with a standardized way allowing the communication between electronic devices and biomedical computing from different vendors facilitating the management of digital images. In this way, the introduction of new services that support the medical applications is facilitated: DICOM enables the integration of scanners, servers, workstations, printers, and network hardware from multiple manufacturers.

A DICOM file consists of a header (so called metadata) and image data. The header includes image related information such as image type, study, information on the mode of acquisition, resolution (height and width), colour, size of the voxel, number of stored bits, and patient's information.

In conclusion, the DICOM standard describes a directory structure that facilitates access to medical images and related information.

1.1.2 PET Imaging

PET is a non-invasive nuclear medical imaging technique based on the visualization of functional processes showing complementary information with respect to anatomical imaging. It provides an in vivo measure of the tumor biological processes (Spieth & Kasner 2002). Among several PET radiotracers derived from isotopes, 18F-fluoro-2-deoxy-D-glucose (FDG) is a glucose analogue widely used in the evaluation of several neoplastic pathologies as well as in radiotherapy planning. FDG uptake is increased in tissue with a high metabolic rate, such as tumor or inflammation regions. These areas appear as hot spots on PET images. FDG PET is able to identify the location of many primary tumors and metastases offering the opportunity to radically change patient treatment (i.e. from radiotherapy to chemotherapy) or the radiotherapy planning treatment volume (PTV) (Guido et al. 2009). For these reasons, the oncology field is the most used and developed for PET study: this technique has been recognized as an adequate staging and restaging tool in various cancer types. In addition, metabolic changes are often faster and more indicative of the effects of the therapy with respect to morphological changes (Wahl

et al. 2009).

To obtain a tomographic study, many acquisitions of the same object at different angles are executed; the object is reconstructed in three dimensions using complex mathematical algorithms such as the Filtered Back-Projection algorithm (FBP) or iterative methods. After applying reconstruction algorithms, a digital image that represents the uptake distribution in the tissues is obtained.

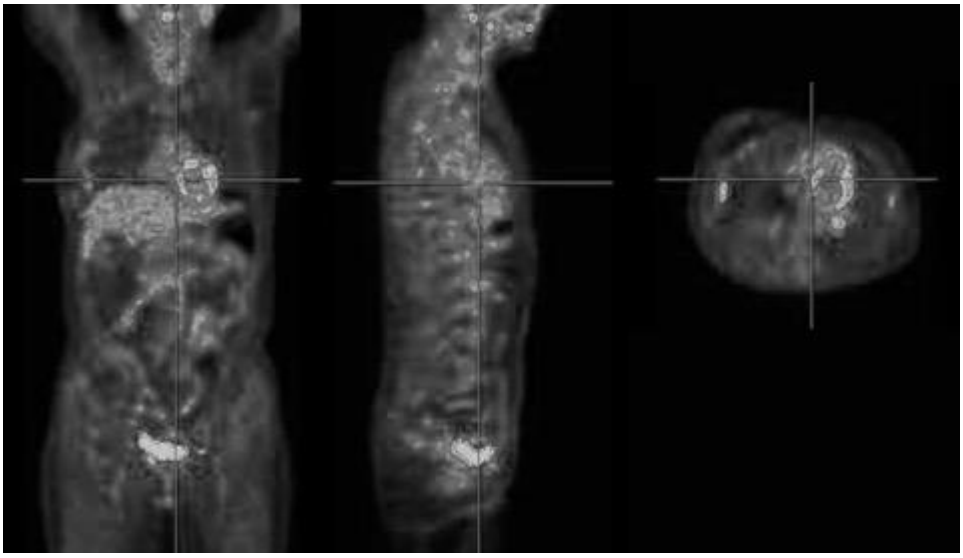

Figure 1 An example of PET study (coronal, sagittal, and axial imaging planes).

1.1.3 CT Imaging

CT is a diagnostic imaging technique which uses ionizing radiation. CT studies are acquired by rotating an X-ray source around the patient. X-ray sensors are positioned on the opposite side, and the patient's bed is continuously moved to obtain a helical acquisition mode. X-ray beam intensity exponentially decreases in extent to the mass attenuation coefficient: the beam will be more attenuated by tissues with a high atomic number. Vice versa, if the beam crosses a low density tissue, the attenuation will be less. In this way, high-density tissues appear clear (maximum attenuation) and lower-density tissue appear dark (minimum attenuation).

After reconstruction, the CT images reflect the attenuation of each voxel according to the Hounsfield scale, where air has a radio-density of -1000 Hounsfield Units (HU) and distilled water a radio-density of 0 HU. CT provides high resolution

morphological images of the body, with an excellent contrast between air, adipose tissue, soft tissue and bone. However, the soft tissue contrast is poor if compared to other imaging modalities such as MRI. Contrast agents may additionally be used to highlight some regions, such as the gastrointestinal tract or the blood vessels.

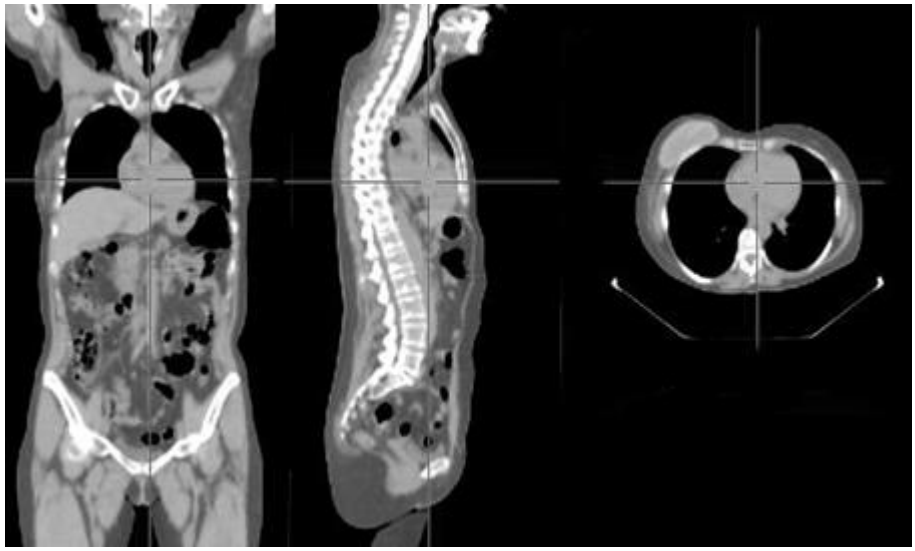

Figure 2 An example of CT acquisition (coronal, sagittal, and axial imaging planes).

1.1.4 MR Imaging

MRI is based on the different relaxation times of tissues after being subjected to an electromagnetic stimulus:

- *T1-weighted images*: the time T1, or longitudinal relaxation time is a measure of the time for which the protons to return to the initial equilibrium conditions, through the transfer of energy to the surrounding microenvironment (lattice), in order to obtain a T1-weighted SE sequence, using a short relaxation time (TR) associated with a short echo time (TE). On T1-weighted images, the cerebrospinal fluid is dark while the fat is brilliant.
- *T2-weighted images*: the time T2, or transverse relaxation time, is a measure of the time taken by the spin of protons to get out of sync. This progressive desynchronization will void the transverse magnetization. A sequence to get a T2-weighted sequence will have a long TR associated with a long TE. Liquids or at least very hydrated tissues, appear bright white in T2-weighted

images.

The relaxation time of a tissue is related with its water content.

MRI provide excellent soft tissue contrast. For this reason, MRI is widely applied in the diagnosis and treatment of neurological, cardiovascular, musculoskeletal, liver and gastrointestinal diseases. Moreover, contrast can be further enhanced through the injection of a contrast enhancement agent.

1.2 Multimodal Medical Imaging

PET/CT or PET/MR imaging merges anatomical and functional information. This combination is clinically advantageous but it suffers from a number of practical problems. It is both expensive and time consuming for patients to undergo separate imaging studies. In addition, patient images must be aligned, or registered, before successful integration due to the patient position change between different imaging sessions. To overcome these issues, the development of hybrid systems has been carried out. In this way, multimodality medical imaging requires a single imaging session. This provides anatomical images that are inherently registered with the functional data.

To date, the availability of multimodality PET/CT (or PET/MRI) systems, providing functional PET images integrated with morphological CT (or MRI) images, allow physicians to accurately localize and characterize oncological lesions for a correct interpretation in most cases. PET images have a low spatial resolution and anatomical imaging techniques are still needed to localize and characterize abnormal regions. The diagnostic accuracy of the combined systems has proven superior to single techniques thanks to their complementary features (Schwartz et al. 2005). The high contrast of PET images and high spatial resolution of CT or MRI images are then fused in multimodal images (see figure 3).

In the case of non-hybrid systems, the multimodality registration problem has been partly solved by recent developments in image processing. The majority of methods are automatic and do not require external markers. The clinical interpretation of registered anatomical and functional information is then performed by displaying the images either simply side-by-side or as a single fused image.

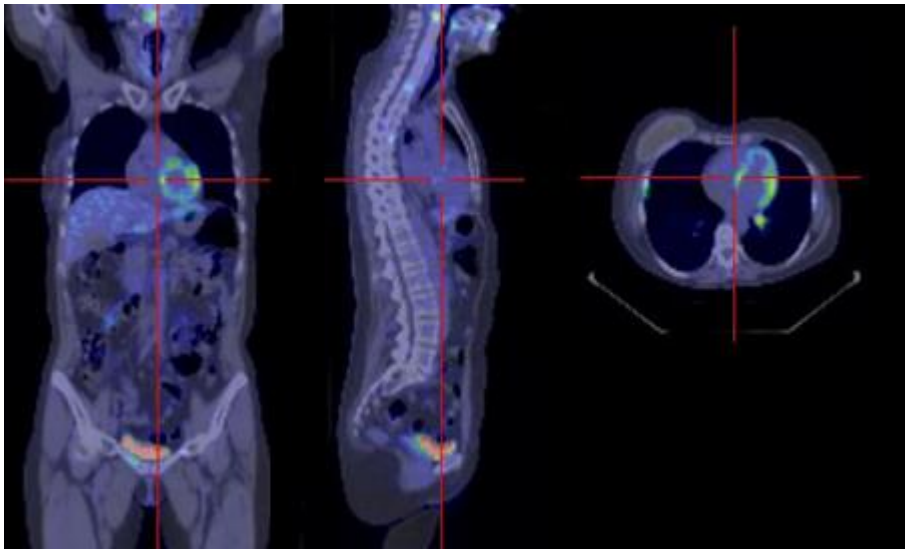

Figure 3 An example of PET/CT study (coronal, sagittal, and axial imaging planes).

1.3 Radiation Therapy Planning

Radiotherapy aims to deliver the necessary therapeutic dose of ionizing radiation to the oncological lesions minimizing irradiation to normal tissues. In this way, radiation is delivered to target to damage and to kill the cancerous cells stopping them from regenerating. Since both cancer and healthy tissues are affected by radiation, any treatment plan should be designed in such a way that the radiation dose delivered to the tumor is high enough to destroy the cancer cells avoiding the delivery of excessive doses of radiation to surrounding normal tissue. Precise tumour volume delineation is a very critical step in order to ensure safe and effective radiation therapies.

The Radiation Therapy Planning (RTP) is to implement the treatment strategy assigning the desired dose to the various ROI. Recent advances in radiotherapy, such as intensity modulated radiation therapy, have improved the dose painting of RTP. These techniques enable a precise delivery of a high dose to the target maintaining a low radiation dose to nearby critical organs. However, the hardware precision in the delivering radiation dose is far greater than the software precision in the target volume delineation: accurate target volume definition is essential for escalating the radiation dose without increasing normal tissue injury.

CT and MRI are considered to be the standard for target volume delineation in many cancer district. Radiotherapy results based on morphological target volume

delineation are disappointing due to the radio resistance of the tumor and/or inadequate dose to target due to missed lesion. CT or MRI imaging may not show the viable extension of tumors and not localize isolated positive lymph nodes (Lauve et al. 2004). To improve these results, PET has been introduced in the radiotherapy field to assist the radiation oncologist in RTP: within the Gross Tumor Volume (GTV), defined on anatomical images, it is possible to define target volumes based on functional area (BTV – Biological Target Volume) and to apply a strategy that will deliver radiation to these regions. Though the use of PET images has been shown to improve target volume definition by reducing intra and inter-observer variability compared to CT images only (Ciernik et al. 2003).

1.4 Thesis Outline

The thesis is structured as follows:

- Chapter 2 describes the state-of-the-art of segmentation approaches in medical imaging.
- Chapter 3 describe an innovative graph based approach for BTV delineation.
- Chapter 4 presents clinical applications to assess the treatment response in oncological patients.
- Chapter 5 describes a delineation approach for Gamma Knife treatment purpose: the multimodal segmentation is described.
- Chapter 6 shows the experimental results.
- Discussion and conclusion about this thesis, and some possible direction for future research.

1.5 Research Contributions

1. **A. Stefano**, Gallivanone F, Messa C, Gilardi MC, Castiglioni I. Metabolic Impact of Partial Volume Correction of [18F] FDG PET-CT oncological studies on the assessment of tumor response to treatment. QJ Nucl Med Mol Imag (2014) 58(4):413-23.
I.F. 1.92
Quartile in Category: Q2 (Radiology, Nuclear Medicine & Medical Imaging)

2. **A. Stefano**, Vitabile S, Russo G, Ippolito M, Sardina D, Sabini MG, Gallivanone

F, Castiglioni I, and Gilardi MC. A graph-based method for biological target volume segmentation. *Oral Communication* at the VIII Congresso Nazionale AIFM (Associazione Italiana Fisica Medica), Torino 2013.

3. **A. Stefano**, Vitabile S, Russo G, Ippolito M, Sardina D, Sabini MG, Gallivanone F, Castiglioni I, and Gilardi MC. A Graph-Based Method for PET Image Segmentation in Radiotherapy Planning: A Pilot Study. A. Petrosino (Ed.): ICIAP 2013, Part II, LNCS 8157, pp. 711–720, 2013.
I.F.: 0.402
Quartile in Category: Q4 (Computer Science, Theory & Methods)

4. **A. Stefano**, S. Vitabile, D. D'Urso, G. Russo, M. Ippolito, M. Sabini, O. Gambino, R. Pirrone, E. Ardizzone, and M. Gilardi, Head and Neck Cancer Tumor Volume Segmentation in PET images Using an Enhanced Random Walk Algorithm, *Under Review at* Medical & Biological Engineering & Computing.
I.F.: 1.726
Quartile in Category: Q2 (Biomedical Engineering)

5. **A. Stefano**, S. Vitabile, G. Russo, D. D'Urso, M. Ippolito, F. Marletta, M.G. Sabini, I.V. Patti, S. Pittera, D. Sardina, O. Gambino, R. Pirrone, E. Ardizzone, and M. Gilardi, Biological Target Volume Segmentation for Radiotherapy Treatment Planning, *accepted to* IX Congresso Nazionale AIFM (Associazione Italiana Fisica Medica), Perugia 2016.

6. **A. Stefano**, Russo G, Ippolito M, Cosentino S, Murè G, Baldari S, Sabini MG, Sardina D, Valastro L, Bordonaro R, Messa C, Gilardi MC and Soto Parra H. Evaluation of erlotinib treatment response in non-small lung cancer using metabolic and anatomic criteria. QJ Nucl Med Mol Imag, *EPUB ahead of print*.
I.F. 1.92
Quartile in Category: Q2 (Radiology, Nuclear Medicine & Medical Imaging)

7. M. Ippolito, **A. Stefano**, G. Russo, S. Gieri, S. Cosentino, G. Murè, S. Baldari, M. G. Sabini, F. Fraggetta, S. Vitabile, M. C. Gilardi, and G. Banna. Evaluation of Platinum-based Therapy Response in Non-Small Cell Lung Cancer, Eur J Nucl Med Mol Imaging (2015) 42: S714-S715.
I.F.: 5.217
Quartile in Category: Q1 (Radiology, Nuclear Medicine & Medical Imaging)

8. **A. Stefano**, N. Porcino, G. Banna, G. Russo, V. Mocciaro, G. Anile, S. Gieri, S. Cosentino, G. Murè, S. Baldari, M.G. Sabini, D. Sardina, F. Fraggetta, S. Vitabile, M.C. Gilardi, and M. Ippolito. Metabolic Response Assessment in Non-Small Cell Lung Cancer Patients after Platinum-Based Therapy: A Preliminary Analysis. Current Medical Imaging Reviews, vol. 11, n. 4, pp. 218-227, 2015.
I.F.: 0.730
Quartile in Category: Q4 (Radiology, Nuclear Medicine & Medical Imaging)

9. **A. Stefano**, S. Vitabile, G. Russo, M. Ippolito, F. Marletta, C. D'Arrigo, D. D'Urso, M. Sabini, O. Gambino, R. Pirrone, E. Ardizzone, and M.C. Gilardi, An Automatic Method for Metabolic Evaluation of Gamma Knife Treatments, A. Petrosino (Ed.): ICIAP 2015, Part II, LNCS 9279, pp. 579–589, 2015.
I.F.: 0.402
Quartile in Category: Q4 (Computer Science, Theory & Methods)

10. **A. Stefano**, G. Russo, S. Vitabile, F. Marletta, C. D'Arrigo, and M. Ippolito. Biological Target Volume Segmentation for Gamma Knife Treatment, *accepted to* International Conference on Clinical PET-CT and Molecular Imaging (IPET) 2015.

11. **A. Stefano**, S. Vitabile, V. Mocciaro, G. Russo, M. G. Sabini, O. Gambino, R. Pirrone, E. Ardizzone, M. Gilardi, and M. Ippolito. An operator independent method for lesion segmentation to evaluate metabolic response in MET-PET studies, *accepted to* World Molecular Imaging Congress (WMIC) 2015.

12. **A. Stefano**, S. Vitabile, G. Russo, F. Marletta, C. D'Arrigo, D. D'Urso, O. Gambino, R. Pirrone, E. Ardizzone, M. Gilardi, and M. Ippolito. An Automatic Method for PET Delineation of Cerebral Tumors, Eur J Nucl Med Mol Imaging (2015) 42: S710-S711.
I.F.: 5.217
Quartile in Category: Q1 (Radiology, Nuclear Medicine & Medical Imaging)

13. **A. Stefano**, S. Vitabile, G. Russo, M. Ippolito, F. Marletta, C. D'Arrigo, D. D'Urso, M. Sabini, O. Gambino, R. Pirrone, E. Ardizzone, and M. Gilardi, A Fully Automatic Method for Biological Target Volume Segmentation of Brain Metastases, International Journal of Imaging Systems and Technology, *in press*.
I.F.: 1.301
Quartile in Category: Q2 (Engineering, Electrical & Electronic)

14. Ippolito M, Cosentino S, Militello C, **A. Stefano**, Rundo L, Russo G, D'Urso D, Pisciotta P, Vitabile S, Sabini MG, Valastro LM, Pittera S, Patti V, D'Arrigo C, Marletta F, and Gilardi MC. Using anatomic and metabolic imaging in Gamma Knife treatments, *accepted to* World Molecular Imaging Congress (WMIC) 2015.

15. P. Pisciotta, C.Militello, L. Rundo, **A. Stefano**, G. Russo, S. Vitabile, M.G. Sabini, C. D'Arrigo, F. Marletta, D. D'Urso, M. Ippolito, M. Midiri, and M.C. Gilardi. Using anatomic and metabolic imaging in sterotactic radio neuro-surgery treatments, *accepted to* IX Congresso Nazionale AIFM (Associazione Italiana Fisica Medica), Perugia 2016.

16. L. Rundo, **A. Stefano**, C. Militello, G. Russo, M. G. Sabini, C. D'Arrigo, F. Marletta, M. Ippolito, G. Mauri, S. Vitabile, and M. C. Gilardi. A Fully Automatic Approach for Multimodal PET and MR Image Segmentation in Gamma Knife Treatment Planning, *submitted to* Computer Methods and Programs in Biomedicine.
I.F.: 1.897
Quartile in Category: Q1 (Computer Science, Theory & Methods)

Chapter 2

Review of the State of the Art

2.1 Medical Image Segmentation

Image segmentation is a fundamental and critical task in numerous biomedical imaging applications. It corresponds to extract a target out of an image identifying edges and areas of similar features.

Various delineation methods have been proposed in the literature. Each segmentation method has a different approach to find the optimal segmentation: depending on the application, appropriate segmentation algorithm must be chosen. It is impossible to make a categorical statement; different segmentation methods have different goals, and each method may be useful in different imaging applications.

In this study, the aim is to segment functional and anatomical regions for RTP. In the next paragraphs, various methods for BTV and multimodality image delineation are discussed.

2.1.1 PET Segmentation

Most of the segmentation techniques on medical images have been applied on CT, MR or ultrasound images and not on PET images since PET has received increasing attention only in recent years in order to incorporate metabolic information in RTP and to evaluate early treatment response in oncological patients.

To date, clinical analyses of PET studies are performed on software platforms with predefined window-level setting. This setting can be liberally modified by the nuclear physician, altering the visual appearance of PET lesions. The manual segmentation approach is then dependent on the experience of the nuclear physician limiting the measurement accuracy. In addition, the manual segmentation is time-consuming and impaired by inter and intra-observer variability.

Currently available tumor delineation methods in PET imaging are based on a fixed threshold of the maximum tracer uptake value in the lesion. Other methods

based on fuzzy c-means (FCM) (Zaidi et al. 2002), region growing (Li et al. 2008) or watershed segmentation (Geets et al. 2007) have been suggested, but few validation studies are available and there is no consensus for a proper BTV delineation method with no clear guidelines on how to incorporate metabolic data into target delineation (Schinagl et al. 2007). Moreover, PET delineation approaches can be categorized on the basis of anatomical sites or radio-tracers used and, according to a comprehensive review of segmentation algorithms in PET imaging (Zaidi & El Naqa 2010), some segmentation methodology categories can be identified.

Image threshold methods are the most widely used due to their intuitive basis and simple implementation. Usually, each PET image voxel is initially converted into its corresponding body-weight Standardized Uptake Value (SUV). The SUV is a widely used PET semi-quantitative parameter, calculated as a ratio of tissue radioactivity concentration and FDG injected dose at the time of injection divided by body weight. Subsequently, a fixed, or adaptive, or iterative threshold (Ford et al. 2006; Drever et al. 2006; Drever et al. 2007; Kao et al. 2010) based on phantom simulations and on scanner hardware features is calculated. An absolute 2.5 SUV value, independent of background and lesion volume, was proposed by Paulino et al. (Paulino et al. 2005) to discriminate benign from malignant lesions and to delineate the BTV. This approach is very controversial and not accepted by the whole scientific community. However, some fixed threshold methods based on analytical equations extracted from realistic phantom experiments can have a better accuracy with respect to more complex adaptive or iterative approaches. These techniques need to be calibrated for each PET scanner and should not be used without optimization. Generally, image threshold methods are inter- and intra-operator independent but strongly scanner dependent, requiring a significant effort in the calibration data to identify the equation parameters. In addition, they are too sensitive to image noise and heterogeneity and partial volume effect (PVE) due to the limited spatial resolution of the PET scanner (Zaidi & El Naqa 2010).

Region growing is a recursive algorithm that groups pixels or sub-regions into larger regions. The grouping is based on predefined criteria, such as grey level threshold values. A region growing usually starts off with one seed point and then recursively add neighbouring pixels that fulfil the criteria to the region. Different connectivities can be used, in 2-D region growing the choice is usually between four and eight-connectivity. The algorithm ends by itself when no more pixels fulfil the criteria on a pixel level (Li et al. 2008).

Supervised and unsupervised learning methods such as artificial neural networks (ANN), support vector machines (SVM), k-means algorithms, FCM algorithms are efficient but only in large lesions of simple shape. These approaches discriminate

target from background based on a set of image features. However, supervised algorithms (ANN and SVM) have limited application in PET imaging, unlike in the MRI or CT fields, due to high heterogeneity that makes the recognition of stable features in the training set difficult. On the other hand, clustering methods such as FCM or Fuzzy Locally Adaptive Bayesian are optimal for the fuzzy nature of the PET lesion edge where a finite number of levels, rather than only two (foreground and background), is used to label voxels within the target area (Zaidi et al. 2002; Hatt et al. 2009). The main issue is the correct identification of the cluster number; Zaidi et al. (Zaidi et al. 2002) started the algorithm with an oversized number of clusters followed by a merging process to group clusters with similar properties to reach a number of clusters in agreement with uniform attenuation areas based on anatomical delineation. However, learning methods require high computational steps, especially in supervised methods, and they are sensitive to image heterogeneity and to study protocol variability, as for example scanner characteristics, PET radiotracers, radiotracer injected dose and interval between radiotracer injection and examination start.

Variational approaches are based on gradient differences between target and background regions attempting to exploit information provided by intensity variation (Wanet et al. 2011). They require smooth regions with reliable boundary information but PET images are characterized by low contrast. In the study proposed by El Naqa et al. (El Naqa et al. 2007), deformable active contour models are computed as directional gradients by means of energy minimization. These approaches are mathematically efficient but require an initialization and are sensitive to image noise and subject to numerical fluctuation.

2.1.2 Multimodal Segmentation

In clinical routine, it is highly desirable to have both functional and structural quantifiable information so the disease can be both identified and localized, potentially resulting in an earlier diagnosis and more effective treatment plan. A concise but complete review of the state of the art regarding multimodal co-segmentation approaches is reported in the following.

Several studies approach PET/CT tumor identification and characterization in radiation therapy scenarios. In (Yu et al. 2009), co-registered FDG-PET/CT were used for the textural characterization of head and neck cancer (HNC) for radiotherapy treatment planning. After a manual segmentation on co-registered PET/CT images (performed by an experienced radiation oncologist), useful textural

features were selected for distinguishing tumor from normal tissue in HNC subjects. Both k-nearest neighbours (KNNs) and decision tree (DT)-based KNN classifiers were employed to discriminate images of cancerous and healthy tissues. Han et al. (Han et al. 2011) presented a Markov random field (MRF)-based co-segmentation of the PET/CT image pair with a regularized term that penalizes the segmentation difference between PET and CT. This graph-based method utilizes the strength of PET and CT modalities for target delineation in a group of 16 patients with HNC. Background and foreground seed voxels must be always manually identified by the user. A similar approach is reported in (Song et al. 2013), where the segmentation is seen as a minimization problem of a MRF model, which encodes the information from both modalities. This optimization is solved using a graph-cut based method, by constructing two sub-graphs for PET and CT segmentation, respectively. The algorithm was validated in robust delineation of lung tumors on 23 PET/CT datasets and two HNC subjects. A further MRF-based systematic solution for the automated co-segmentation of brain PET/CT images into grey matter, white matter and CSF regions is exposed in (Xia et al. 2012). A PET/CT image pair and its segmentation result are modelled as a MRF triplet, and segmentation is eventually achieved by solving a maximum a posteriori (MAP) problem using the expectation maximization (EM) algorithm with simulated annealing. The overall MRF-MAP model was tested both on both simulated and real patient PET/CT data.

The authors of (Potesil et al. 2007) proposed a method for automated delineation of tumor boundaries in whole-body PET/CT by jointly using information from both PET and diagnostic CT images. After an initial robust hot-spot detection and segmentation performed in PET, a model for tumor appearance and shape in corresponding CT structures is learned by weighted non-parametric density estimate. This voxel-based CT classification is then probabilistically integrated with PET classification using the joint likelihood ratio test technique to derive the final segmentation. The algorithm was tested on patient studies with lung and liver tumors identifiable in both the PET and CT images acquired using the same scanner.

Yezzi et al. (Yezzi 2001) introduced a geometric variational framework that uses active contours to simultaneously segment and register features from multiple images. The key aspect of this approach is that multiple images may be segmented by evolving a single contour as well as the mappings of that contour into each image during feature-based realignment steps. The results of three experiments on MRI/CT images of the head and the spine are reported. Also the authors of (El Naqa et al. 2007) developed variational methods based on multivalued level set deformable models for simultaneous 2D or 3D segmentation of multimodality images consisting of combinations of co-registered PET, CT, or MRI datasets. In particular, only three

patients are considering: a non-small cell lung cancer case with PET/CT, a cervix cancer case with PET/CT, and a prostate patient case with CT/MRI. In addition, CT, PET, and MRI phantom data were used for quantitative validation of the proposed multimodality segmentation approach.

An automatic algorithm for the co-segmentation of HNC based on PET/MRI data was proposed in (Leibfarth et al. 2015), in order to standardize tumor volume delineation. For both imaging modalities tumor probability maps were derived, assigning each voxel a probability of being cancerous according to its signal intensity. A combination of these maps was subsequently segmented using a threshold level set algorithm. The algorithm processes both the anatomical T2-weighted MRI and FDG-PET data concerning 10 HNC patient datasets acquired in a combined PET/MRI system.

The group led by Bagci developed co-segmentation approach in multimodal medical imaging, using the Random Walker (RW) algorithm (Bagci et al. 2013) and unifying graph representation of each image modality in a single product lattice. The overall method results in a fully automatic framework, providing an automated object detection via interesting uptake region algorithm to avoid users foreground and background seed detection. Afterwards, prior to the initiation of the segmentation process, these identified seeds are propagated to the corresponding anatomical images. Although no significant anatomical and functional changes between the scans have to be assumed, the study used PET, PET/CT, MRI/PET, and fused MRI/PET/CT scans from 56 patients who had various lesions in different body regions.

Chapter 3

Image Segmentation in PET

Due to the nature of PET images (low spatial resolution, high noise and weak boundary), the BTV varies substantially depending on the algorithm used to delineate functional signal in PET images. Visual delineation is widely-used, but it is strongly operator-dependent, even if easily applicable. It is potentially inaccurate because susceptible to the window level settings and subject to both intra and inter-operator variability. In addition, the manual segmentation is time-consuming and impaired by inter and intra-observer variability. For this reason, the development and implementation of robust, fast, accurate, operator and scanner independent segmentation method is mandatory to properly use the crucial information provided by molecular imaging.

3.1 Random Walks on Graphs Method

Graph based methods are used to find the globally optimal segmentation of images. They provide a foreground and background recognition in images in which the seeds are specified by user input.

An undirected graph G can be represented as a pair G = (V,E) with nodes $v \in V$ and edges $e \in E \subseteq VxV$. A node v_i is a neighbour of another node v_j if they are connected by an edge e_{ij} with a weight w_{ij} ($w_{ij}=w_{ji}$ being an undirected graph).

The graph-based segmentation method represents a DICOM image as a graph in which the voxels are its nodes and the edges are defined by a cost function that maps a change in image intensity to edge weights.

The graph cut algorithm (Boykov et al. 2001) tries to partition the target and background edges detecting bottlenecks in a weighted graph. This approach is a computationally difficult algorithm and it has no exact solution. In addition, since this method tries to minimize the sum of edge weights in the cut, it may return very small segmentations as a result of low contrast or noise. To solve this problem, known as “small-cut”, many seeds are needed.

The Random Walks on Graphs (RW) algorithm was developed by Grady (L. Grady 2006) to resolve these issues in the computer vision domain and then was extended for image segmentation (Bagci et al. 2011). Aside from the fact that both graph cuts and RW are graph-based methods, they are actually quite different. Rather than treating the segmentation as a max-flow/min-cut problem, RW treats the segmentation as the solution to a linear system with an exact solution. In addition, RW, with respect to graph-cut, is less susceptible to "small cut" behaviours and is more efficient in terms of handling ambiguities among object boundaries.

RW is very accurate in noisy and low contrast images, such as PET images.

The PET image is then converted into a lattice where some voxels are known (nodes with labels specified by user input) and some voxels are not known. The delineation problem is to assign a label to unknown nodes. This is done by trying to find the minimum cost/energy among all possible scenarios in the graph to provide an optimal segmentation. The RW algorithm can be used to partition the nodes into two disjointed subsets representing lesion and background. The RW partition is derived from the probability that a "random walker" starting at a source node, first reaches a node with a pre-assigned label, visiting every node in the image. Then, the RW problem is to determine the highest probabilities for each pixel to reach the target node (target seed) and has the same solution as the combinatorial Dirichlet problem. The Dirichlet integral is defined as:

$$D[u] = \tfrac{1}{2} \int |\nabla U|2 \, d\Omega \qquad (1)$$

for a field U and region Ω. Its solution is given by a harmonic function that satisfies the Laplace equation:

$$\nabla 2U = 0 \qquad (2)$$

The combinatorial Laplacian matrix (L_{ij}) is a sparse matrix defined as:

- Σw_{ij} if i = j
- $- w_{ij}$ if i and j are adjacent nodes
- 0 otherwise

It can be partitioned into four sub-matrices:

$$\begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix} \qquad (3)$$

where L_U is a sub-matrix of edge weights for the un-labelled voxels and L_M is a sub-matrix of edge weights for the labelled nodes. B is a sub-matrix of edge weights corresponding to the labelled voxels.
The harmonic function can be found by solving a system of linear equations as follows:

$$L_U X = -B^T M \qquad (4)$$

M is a matrix of 0 or 1 corresponding to the boundary conditions of seeds. X is the probability for each node being a member of the labels: a threshold of 50% is chosen to discriminate the foreground from the background creating a voxel binary mask, so that:

- target node value = 1 if its probability $\geq$ 50%
- background node value = 0 if its probability < 50%.

This threshold implies that any pixel with less than a 50% chance of being in the foreground is rejected.

The weights w_{ij} between nodes, necessary for the walk moving on the graph, are constructed using a Gaussian function based on image intensity values:

$$w_{ij} = \exp(-\beta(g_i-g_j)\wedge^2) \qquad (5)$$

where both g_i and g_j are the image intensity values at voxels i and j; β is a free parameter depending on the user. RW measures the "betweenness" through the starting pixel (foreground seed) to the un-labelled pixel determining the highest probabilities for assigning labels to the nodes.

3.2 Random Walks on Graphs in PET imaging

The RW algorithm has been modulated to adapt to the PET application domain in order to incorporate metabolic information in the radiotherapy field. The Gaussian weighting function for the PET image is defined as:

$$w_{ij} = \exp(-\beta(SUV_i-SUV_j)\wedge^2) \qquad (6)$$

where

$$SUV = R/ (I / W) \quad (7)$$

R is the radioactivity activity concentration [kBq/ml], I is the decay-corrected amount of injected radiotracer [kBq], and W is the patient's weight [g]. The SUV, the most common semi-quantitative parameter used to estimate radiotracer accumulation within a lesion in clinical practice, normalizes the voxel activity considering acquisition time, administered activity and patient's weight. Hence the PET image is converted into a lattice where the SUV of each voxel is mapped to w_{ij}. Then, each random walker is driven by the gradient field and the SUV of PET images: the probability of each node is inversely proportional to the contrast between the voxels which does not allow the walks to cross the edges.

3.3 Clinical Applications

The accuracy of segmentation methods can be assessed using phantoms containing targets of simple shapes, such as spheres, simulating tumor lesions. Actual lesions are irregular, heterogeneous and can have complex shapes with respect to phantom spheres. For a realistic evaluation of delineation methods, it is important that irregular and inhomogeneous volumes are considered: the crucial step is to try delineation methods in actual patient situations. For this reason, two different solutions for BTV delineation in Head and Neck Cancer (HNC) and brain cancer have been implemented to overcome the RW issues:

1. the RW algorithm requires a set of pre-labeled seeds, which are generated interactively. The delineation is sensitive to the location of the pre-labeled voxels.
2. the RW algorithm is sensitive to the choice of β weighting factor in (5).

3.3.1 BTV delineation: Head and Neck Cancer

Radiotherapy is commonly used in the treatment of head and neck cancer. Figure 4 shows a bifurcated PET/CT lesion in an oncological patient with HNC; the volume of a complex metabolic lesion may evolve splitting into more parts or merging in a single part or both of these. Thus, a complex shape requires an accurate and efficient method of segmentation able to follow the lesion in its whole volume.
An enhanced version of the Random-Walk algorithm has been implemented:

1. The proposed algorithm incorporates k-means clustering to select the target seeds along the whole tumor volume performing a precise delineation of complex lesions
2. An adaptive probability threshold is proposed to discriminate between target and background regions rather the fixed one of the original RW method. The adaptive probability threshold takes into account the intensity changes between adjacent PET slices along the metabolic volume and it makes the algorithm not subjective.

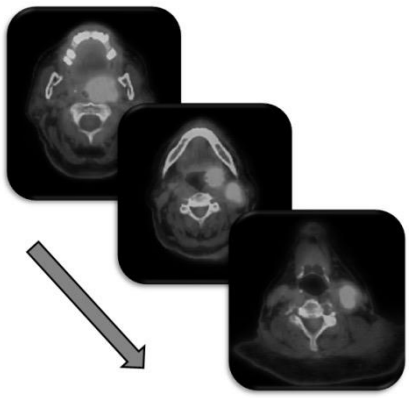

Figure 4 A PET lesion that evolves splitting into two parts and merging in a single part along its volume.

3.3.1.1 RW algorithm with K-means: K-RW

The RW delineation is sensitive to the location of the pre-labelled voxels. An automatic method of background and foreground seed detection is then proposed: the implemented method is based on k-means clustering to make the BTV delineation process feasible in complex and heterogeneous lesions, such as bifurcated lesions.

The user draws a line on the coronal PET image along the target, and the axial slice with maximum SUV (SUV_{max}) is automatically identified and showed to the user that draws a new line along the lesion. This approach allows to properly delineate the PET tumor, excluding false positives (normal structures such as brain, heart, bladder, and kidneys that normally have high FDG uptake).

The algorithm can be split in two main steps: the pre-segmentation step to automatically detect the RW seeds, and the segmentation step to delineate the final metabolic tumor. The initial target seeds are the voxels corresponding to the line drawn by the user (target seed line).

The delineation method is achieved by following these steps:

1. The target seeds with a SUV less than 30% of the SUV_{max} are removed to avoid eventual necrotic or background area.
2. The neighbourhood through searching in 8 directions with respect to the voxel with SUV_{max} is explored to detect background seeds. For all 8 directions, the location of the first voxel with a value less than 30% of the average of target seed SUVs is identified. Those 8 voxels are marked as background seeds.
3. The RW delineation performs a "roughly" pre-segmentation by utilizing the target seed line and the 8 background seeds. The probability threshold to discriminate target from background voxels is fixed at 50%: any voxel with less than a 50% chance of being in the foreground is rejected.
4. The k-means algorithm is used to automatically select k-cluster centres within the pre-segmented lesion. In a complex volume, a lesion can be divided into two or more areas with different hot peeks. This algorithm follows the evolution of the target in the whole volume identifying centroids of hot regions. In the case of a homogenous target (such as a sphere in phantom studies), the algorithm returns a single centroid. The *k* value is automatically inferred.
5. The centroids (one or more) and the voxels within the pre-segmented lesion with a SUV greater than 90% of SUV_{max} are identified as new target seeds.
6. The RW algorithm performs the segmentation by utilizing the seeds identified in step 2 and 5.
7. For the first slice (the slice with SUV_{max}), the user can manually change the probability threshold, rather the fixed one of 50%, to discriminate target from background voxels in order to select the value that optimizes the segmentation task: the probability threshold chosen by the user in this slice remains fixed for the whole volume.

This process is repeated slice by slice to obtain the whole lesion volume and it is performed in parallel for slices above and below the first one. In particular, the seeds are propagated in the next slices until no segmentation or abnormal increment of target seeds is verified.

An overview of the K-RW (K-means RW) seed localization method is outlined in figure 5.

About step 4, the k-means algorithm takes as input the matrix that contains the pre-segmented lesion. It returns the coordinates of the *k* centroids of hot regions.

The *k* value is automatically inferred:

- k=2 if the target seed line contains voxels with SUV greater than 30% of the

SUV_{max}. This check allows excluding necrotic or background area. Lesion and background regions will represent the two clusters.

- k=n+1 if the start target seed line contains seeds with SUV less than 30% of the SUV_{max}. These voxels are considered as belonging to the background or necrotic region: the seed line passes from a "hot" region to another "hot" region. k= n+1 indicates the n "hot" regions corresponding to the segment number of the target seed line after the thresholding step, and the background region (various background regions correspond to a single region).

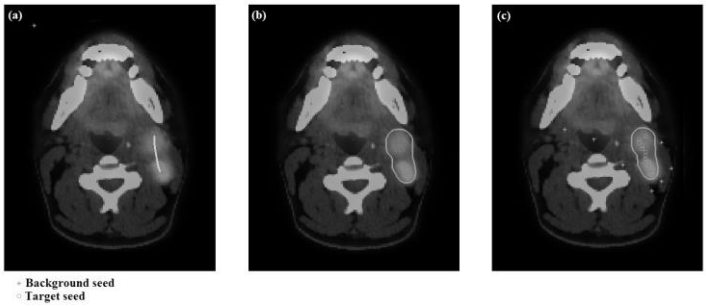

Figure 5 Pre-segmentation and segmentation steps in the $slice_{max}$ are shown in (a–c). The user draws a line on the lesion (a). Then, the RW algorithm performs a rough pre-segmentation step (region of interest in (b)) to detect the target and background seeds used to perform the final segmentation step (c).

3.3.1.2 RW algorithm with adaptive probability threshold: AK-KW

The RW method depends on the choice of the β factor in the Gaussian weighting function in equation 5. In the paper reported in (Bagci et al. 2013), the authors improved RW robustness in PET imaging but they did not deal with the dependence of this parameter. In (Onoma et al. 2014), the authors take into account this limitation suggesting the use of the Euclidean distance between adjacent nodes. Nevertheless, the authors do not take in consideration the issue of complex lesions that may evolve splitting into more parts or merging in a single part or both of these, such as HNC in which the Euclidean distance is not an optimal solution. To overcome these limitations, we propose an approach based on the RW probability matrix to obtain an adaptive probability threshold rather than a fixed one of 50%. In this way, the proposed method becomes independent of the choice of β factor because β influences how quickly the probability decreases with increasing intensity: a high β value reduces the weight of walker, which weakens the connection between the adjacent

nodes and underestimates the target volume. Vice versa, a low β value increases the weight, which overestimates the target volume. The adaptive probability threshold takes into account the intensity changes between adjacent PET slices along the whole metabolic volume because the probability threshold changes slice after slice to discriminate target from background voxels.

The algorithm is the same as the previous one descripted in paragraph 3.3.1.1 except for step 7 where the probability threshold is automatically inferred by the system, slice after slice (the probability threshold changes during volume delineation). The probabilistic output of segmentation is processed to obtain an adaptive threshold value (P) by following these steps:

1. Calculate the mean (M) of the probability values inside a large pre-segmented lesion obtained using a fixed probability threshold of 80%. Identification of two groups of voxels:
    - Voxels with a probability < M.
    - Voxels with a probability ≥ M.
2. Calculate the probability means (P1 and P2) of the two groups.
3. The adaptive threshold value is then calculated as P=½ (P1+ P2).

Any voxel with less than a P% chance of being in the target is rejected:

- target voxel value = 1 if its probability ≥P%
- background voxel value = 0 if its probability < P%

This method follows the whole lesion volume taking into account the gradient of intensity and contrast changes of the PET lesion in different slices.

3.3.2 BTV delineation: Brain Cancer

FDG PET is not able to detect metabolic active regions in cerebral area, because the whole brain has a high uptake of glucose appearing as a hyper-intense area. Consequently, another radio-tracer must be used in order to have functional information about brain lesions. 11C-labeled Methionine (MET) is an amino acid that shows a greater transport within the active cancer cells. It has been reported that the extent of tumour cell invasion can be detected more clearly by MET PET than by CT or MRI (Nariai et al. 2005). Using MET PET imaging, it is possible to define the BTV to apply a strategy that will deliver radiation to these regions.

A fully automatic and operator independent method for the BTV delineation of brain metastases based on an extension of the RW algorithm is then proposed. Target and background RW seeds are automatically identified and an adaptive threshold,

likewise to that used in HNC patient study, is proposed. Due to these improvements proposed, the method is called automatic and adaptive random walk (AA-RW).

3.3.2.1 Fully Automatic Method

The fully automatic identification of target and background seeds is feasible because the MET PET datasets for Gamma-Knife treatments include only the brain area avoiding the possible presence of false positives in other anatomic regions. Differently, in total body PET examinations, such as those carried out in our previous clinical study, a user interaction to manually identify the target lesion is always needed (normal structures such as brain, heart, bladder, and kidneys normally have high FDG uptake).

To obtain a fully automatic and operator independent method, the target and background seeds are localized by following these steps:

- The PET slice with the highest SUV (SUV_{max}) is identified
- The n nodes with a SUV greater than 95% of SUV_{max} are marked as target seeds
- The neighbourhood of the node with SUV_{max} through searching in 8 directions is explored to identify the nodes with a SUV less than 30% of the average of target seed SUVs. In this way, 8 background nodes are identified.

3.3.2. Adaptive probability threshold: AA-RW

The probabilistic output of segmentation is processed to obtain an adaptive threshold value (P) by following these steps:

1. Calculate the mean (M) of the probability values inside a large pre-segmented lesion obtained using a fixed probability threshold of 80%. Identification of two groups of voxels:
    - Voxels with a probability < M.
    - Voxels with a probability ≥ M.
2. Calculate the probability means (P1 and P2) of the two groups.
3. The adaptive threshold value is then calculated as P=½ (P1+ P2).

Any voxel with less than a P% chance of being in the target is rejected:

- target voxel value = 1 if its probability ≥P%
- background voxel value = 0 if its probability < P%

This method follows the whole lesion volume taking into account the gradient of intensity and contrast changes of the PET lesion in different slices.

Chapter 4

Metabolic Evaluation of Oncological Treatments

FDG-PET imaging has been a fundamental impact in many fields of oncology and has been recognized as an adequate staging tool. The uptake quantification of oncological lesion is requested for a more accurate characterization of the tumour malignancy and a better patient monitoring after chemotherapy or radiotherapy treatments. Furthermore, the availability of in vivo quantitative biomarkers for cancer is crucial in perspective of personalized medicine and predictive medicine. Nevertheless, some drawbacks are evident when PET quantitative analysis are performed. PET quantification is in fact strongly affected by the PVE (Soret et al. 2007), as a consequence of the PET poor spatial resolution. For this reason, a PVE correction method has been implemented and evaluated in treatment monitoring (Gallivanone et al. 2011; Stefano et al. 2014) (see figure 6).

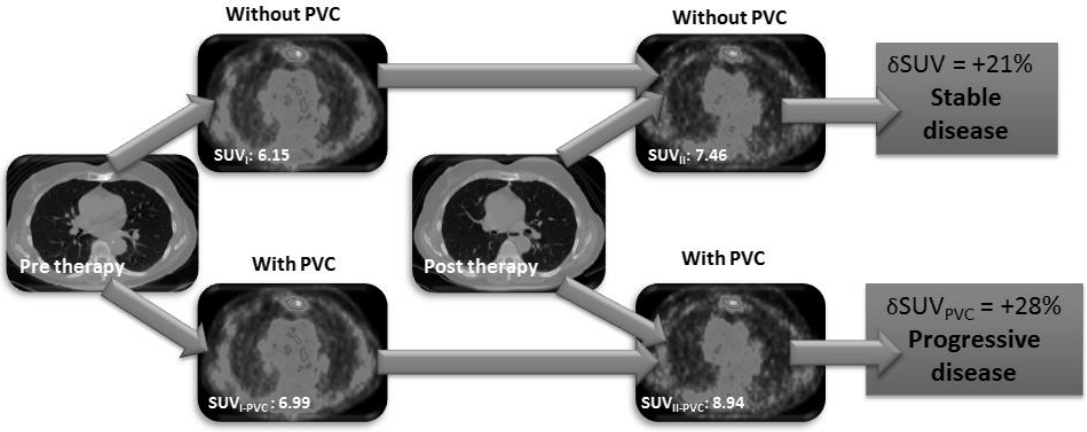

Figure 6 PVE has a critical impact on PET SUV treatment monitoring.

In addition, metabolic changes assessed by PET imaging are often faster and more indicative of the effects of therapy with respect to morphological changes, providing a more rapid evaluation of the response to treatment (Wahl et al. 2009). Current generation PET/CT systems allow the integration of functional and morphological data leading to a better characterization and localization of the tumors and, consequently, a better diagnostic performance.

According to RECIST (Response Evaluation Criteria in Solid Tumors)

(Eisenhauer et al. 2009), therapy assessment on radiological imaging is based on a change in lesion size. When more than one measurable lesion is present at baseline all lesions up to a maximum of five lesions total representative of all involved organs should be identified as target lesions. A sum of the diameters for all target lesions is then calculated.

Evaluation of therapy response according to RECIST:

- Complete Response (CR): Disappearance of all target lesions.
- Partial Response (PR): At least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum diameters.
- Progressive Disease (PD): At least a 20% increase in the sum of diameters of target lesions. In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm. The appearance of one or more new lesions is also considered progression.
- Stable Disease (SD): In another cases.

PET images are analysed semi-quantitatively by using the SUV. The European Organization for Research and Treatment of Cancer (EORTC) has proposed metabolic criteria to evaluate the response after therapy by PET studies (Young et al. 1999). In the same way, new standardized criteria to metabolically assess treatment response, PET Response Criteria In Solid Tumours (PERCIST), was proposed in 2009 (Wahl et al. 2009). Key elements of PERCIST, with respect to EORTC, include performance of PET scans by a method consistent with the National Cancer Institute recommendations and with those of the Netherlands multicentre trial group, on well-calibrated and well-maintained scanners 18.

Therapy response classification	**EORTC**		**PERCIST**
	$EORTC_{15\%}$	**$EORTC_{25\%}$**	
Progressive Metabolic Disease (PMD)	ΔSUV > 25% or visible increase in the tumor uptake extent or new lesion	ΔSUV > 25% or visible increase in the tumor uptake extent or new lesion	ΔSUL>30% or visible increase in the tumor uptake extent or appearance of new lesions
Stable Metabolic Disease (SMD)	-15%<ΔSUV<25% and no visible increase in tumor uptake extent	-25%<ΔSUV<25% and no visible increase in tumor uptake extent	-30%< ΔSUL< 30% and no visible increase in tumor uptake extent
Partial Metabolic Response (PMR)	ΔSUV< -15%	ΔSUV< -25%	ΔSUL<-30%
Complete Metabolic Response (CMR)	Disappearance of all lesions to background level	Disappearance of all lesions to background level	Disappearance of all lesions to back-ground level

Figure 7 EORTC and PERCIST classifications.

According to the PERCIST protocol, the SUL_{peak} is the SUV normalized to lean body mass calculated as the average within a 1 cm^3 spherical VOI centred at the portion of the lesion with the highest 18F-FDG uptake in each patient. The SUL

threshold for a lesion to be reportable is (1.5xSUL_{mean}) + (2xSD of SUL_{mean}), being SUL_{mean} estimated on a reference VOI positioned on the right lobe of the liver. In cases of diseased liver, the VOI is positioned on the descending aorta and the threshold is (2xSUL_{mean}) + (2xSD of SUL_{mean}). Finally, in order to take into account non-pathologic variability PERCIST recommends comparing baseline and follow-up PET studies when the liver/aorta SUL variation is less than 20 percent.

Larson et al. (Larson et al. 1999), in 1999, introduced the Metabolic Tumour Volume (MTV) and the Total Lesion Glycolysis (TLG) to evaluate therapeutic response. TLG, defined as MTV x (average SUV within the MTV), combines the volumetric and SUV information, to try to obtain a better evaluation of the treatment response. These parameters can be considered prognostic markers, and are correlated with disease-free survival (DFS) and overall survival (OS). Nevertheless, the metabolic volume and, consequently, the TLG vary substantially depending on the method used to delineate functional positive tissues, due to the low spatial resolution and high noise level of PET images. For this reason, the developed graph-based algorithm, introduced in the third chapter, has been used.

The calculation of SUL over lesion and liver/aorta, accounting for all voxel data inside the VOI, the liver/aorta SUL comparison in different time points, and the lesion classifications, are a cumbersome procedure and needs the development of customized software. Ad-hoc stand-alone tool has been developed to implement RECIST, EORTC, PERCIST, MTV and TLG as a Medical Decision Support System to help clinicians in treatment response evaluation of oncological patients: the customized framework developed to carry out clinical studies is shown in figures 8, and 9.

4.1 Lung Cancer

4.1.1 Evaluation of Erlotinib treatment response

In this clinical application, the value of PET for early prediction of tumour response to Erlotinib (Tarceva, Astellas Pharma Inc.) in patients with advanced or metastatic non-small cell lung cancer (NSCLC) after failure of at least one prior chemotherapy regimen is evaluated. In NSCLC the tyrosine kinase receptors are over-expressed or non-regulated: these receptors are therefore recognized as targets for cancer therapy. Erlotinib is an epidermal growth factor receptor (EGFR)-tyrosine kinase inhibitor (TKI) that has been gradually used in the treatment for advanced NSCLC after the failure of more than one or two previous chemotherapeutic

regimens. Glucose metabolic activity reflects the response to EGFR TKI.

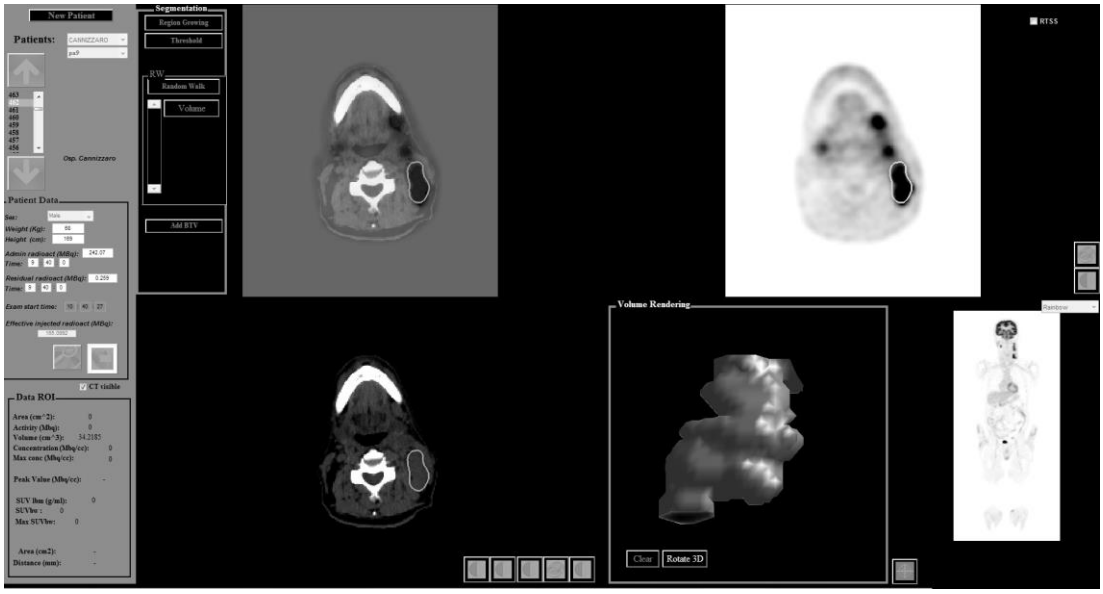

Figure 8 The developed framework to automatically assess the treatment response.

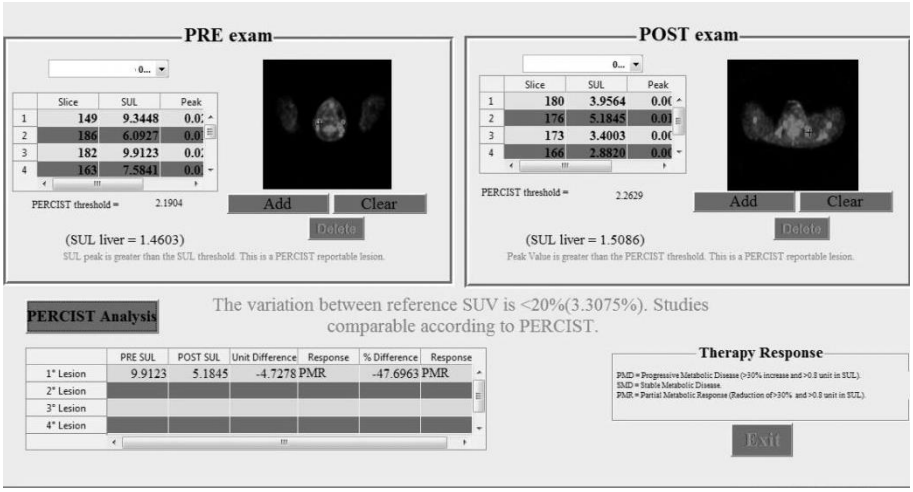

Figure 9 PERCIST classification of a treated patient.

Patients receiving Erlotinib neoadjuvant chemotherapy treatment (150 mg orally once daily for 45 days) undergo a FDG PET/CT scan before treatment (PETI) and after administration of 3 doses of Erlotinib at 48 h (PETII) for early monitoring. A further metabolic follow-up is planned at the end of Erlotinib treatment, 45 days

from the start of therapy (PETIII). CT study, after the administration of 100 ml of intravenous iodate contrast material (Iomeprolo, Iomeron 300), is performed before and 45 days after the initialization of treatment.

For each PET study, the lesion presenting with the highest 18F-FDG uptake is considered (not necessarily the same lesion on three PET studies) according to EORTC, PERCIST and RECIST classifications to assess metabolic and anatomic response. In addition, the metabolic volume is calculated using the method described in the previous chapter (section 3.3.1).

4.1.2 Evaluation of Platinum-based therapy response

The purpose of this clinical application, is to evaluate the PET value for early prediction of tumour response to platinum-based therapy in patients with NSCLC. Patients receive a platinum-based chemotherapy according to the American Society for Clinical Oncology or European Society for Medical Oncology guidelines, with carboplatin or cisplatin in combination with another third generation agent.

Patients undergo a FDG PET/CT examination without contrast before platinum based chemotherapy and ~28 days after the initiation of therapy for early monitoring. Morphological diagnostic CT scans before and after the initiation of treatment are performed. The evaluation is carried out comparing the standard treatment response using RECIST with metabolic treatment response according to EORTC, PERCIST, TLG and MTV.

4.2 Brain Cancer

Leksell Gamma Knife® (Elekta, Stockholm, Sweden) is a stereotactic radio-surgical device to treat different brain disorders that are often inaccessible for conventional surgery, such as benign or malignant tumours. Stereotactic radiosurgery allows an accurate external irradiation (with a single, high dose and sharp dose gradient) to minimize doses given to adjacent critical brain structures. The gamma rays (generated by cobalt-60 [60Co] radioactive sources) are focused on the target through a metal helmet. It provides a safe and effective way of treating inaccessible cerebral tumours.

An examination to differentiate malign and benign tissue in brain tumours with great preciseness is the PET with the amino acid tracer 11c-methionin (MET).

Numerous studies have shown that the specificity of the MET PET for marking

tumour delineation and for the differentiation relapse versus radiation necrosis is higher compared with MRI.

The implemented segmentation method described in section 3.3.2 is used in patients with cervical metastases to evaluate therapeutic response in sequential scans. Patients undergo MET PET examinations before and 2 months after the Gamma Knife treatment. Lesion segmentation is used to evaluate therapeutic response using SUV_{max}, MTV, and TLG variations in sequential scans.

Chapter 5

Multimodal Image Segmentation

Anatomical GTV often does not match with BTV at all. Metabolic imaging could be used to provide additional information useful for treatment planning and enhanced tumour characterization. In this way, the BTV can be used to modify the GTV in order to treat the actual cancer region more precisely combining the complementary information of tissues from both anatomical and functional domains. For this reason, a robust joint anatomical/functional segmentation method, which simultaneously segments lesions in each image domain, is required. This task, named co-segmentation, is a challenging problem due to:

- unique demands and peculiarities brought by each imaging modality, and
- lack of one-to-one region and boundary correspondences of lesions in different imaging modalities.

5.1 Gamma Knife Treatment

Leksell Gamma Knife (Elekta) radiosurgery is a mini-invasive technique defined as the delivery of a single, high dose of radiation to obtain a complete destruction of brain lesions with an abrupt drop of dose to healthy tissue. It provides a safe and effective way of treating inaccessible brain tumours. The high preciseness of the irradiation is achieved by special stereotactic fixing, localization and positioning device (Leksell 1983).

Nowadays, the GTV for radiation treatment planning is identified on MRI. MRI is preferred over CT when soft-tissue contrast resolution must be highly detailed such as in brain malignancies. Even if MRI is characterized by a high spatial resolution, it contains just morphological information, without providing tumour biology information. GTV will not match with the functional area of the tumour, and metabolic imaging must be considered to assist the radiation oncologist in treatment planning (Levivier et al. 2002).

FDG-PET for BTV delineation shows limited clinical value in brain tumours. Only in a few patients, additional information are derived from FDG-PET for radiation treatment planning because of the low contrast between viable tumours and normal brain tissue, although FDG uptake is regionally related to anaplastic areas. In

fact, FDG is used as a tracer of glucose metabolism. The distribution of FDG is thus not limited to malignant tissue because FDG enters the cells according to glucose transport mechanism: it is known that FDG PET is not able to detect metabolic active regions in cerebral area, because the whole brain has a high uptake of glucose appearing as a hyper-intense area. Consequently, another radio-tracer must be used in order to have functional information about brain lesions.

[11C]-Methionine (MET) is a natural amino acid avidly taken up by brain cancer cells, whereas its uptake by normal brain is low. MET-PET discerns malign and benign tissue in brain tumours with great sensitivity and specificity, by localizing selectively in cancer regions of the brain. Numerous studies have shown that the specificity of the MET-PET for marking tumour delineation and for the differentiation relapse versus radiation necrosis is higher compared with MRI. In the paper reported in (Grosu et al. 2005), metabolic imaging was used for biological target delineation in 36 patients that showed a significantly longer median survival compared with the group of patients, in which target volume was merely defined by MRI. In addition, the extent of tumour cell invasion can be detected more clearly by MET-PET than by CT or MRI (Nariai et al. 2005).

Using MET PET imaging, within the GTV, it is possible to define the BTV to apply a strategy that will deliver radiation to these regions. Integration of PET in Gamma Knife treatment may contribute to a better management of brain metastases.

5.2 Joint PET/MRI Segmentation Approach

The proposed multimodal method exploits PET and MRI computer-assisted segmentation methods, which have been properly combined together and improved. An image co-registration stage is mandatory in order to integrate and quantitatively compare MRI and PET modality imaging data. Figure 10 outlines the overall flow diagram of the proposed PET/MRI joint segmentation method.

5.2.1 BTV delineation

The fully automatic segmentation method described in section 3.3.2 is used to obtain the BTV.

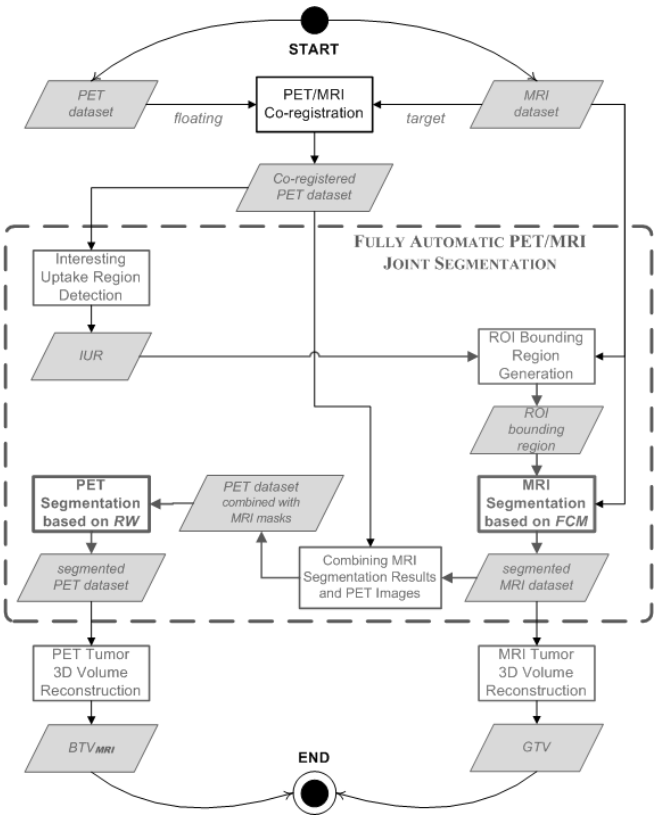

Figure 10 Flow diagram of the proposed fully automatic PET/MRI joint segmentation method.

5.2.2 GTV delineation

The MRI tumour segmentation approach, based on unsupervised Fuzzy C-Means Clustering (FCM) and proposed in (Militello et al. 2015), is outlined. Nevertheless, the method is semi-automatic. User intervention is reduced to the selection of a ROI bounding area containing the cancer zone and no parameter setup is required. All the subsequently processing stages are accomplished on the MR images masked with this ROI bounding region.

Firstly, some pre-processing operations are applied:

1. contrast stretching operation, by means of a linear intensity transformation, to convert the input intensity values into the full dynamic range;
2. smoothing operation, using an average filter, to remove the MRI acquisition noise (i.e. due to static magnetic field non-uniformities).

The ROI segmentation based on unsupervised clustering is performed on MR images including brain tumours. FCM cluster analysis is an iterative machine

learning technique that classifies a dataset (i.e. a digital image) into groups (regions) (Bezdek et al. 1984). Mathematically, the goal of this algorithm is to divide a set of objects (statistical samples represented as vectors belonging to n-dimensional Euclidean space) into clusters (partitions of the input dataset), represented by centroids. A fuzzy partition is defined as a fuzzy set family such that each object can have a partial membership to multiple clusters. The matrix defines a fuzzy C-partition of the set through membership functions, whose values represent membership grades of each element to the i^{th} fuzzy set (cluster).

FCM is an optimization problem where the objective function must be iteratively minimized using a least-squares method. Two clusters are selected in order to suitably classify a hyper-intense lesion from the healthy part of the brain. In fact, in contrast-enhanced MR brain scans, metastases have a brighter core than periphery and distinct borders distinguishing them from the surrounding healthy brain tissue. As the brain tumours are imaged as hyper-intense enhancement regions, during the defuzzification step, the pixels that have the maximum membership with the brightest cluster are selected.

Since sometimes edema or necrosis (hypo-intense regions) could be present in some lesions and must be included in the planned target volume for radio-therapeutic purposes, the following post-processing refinements are required:

1. small area removal, to delete any unwanted connected-components included by the highest intensity cluster. These small regions may be due to anatomical ambiguities when the lesion is located near high valued pixels, i.e. cranial bones, head skin or the corpus callosum;
2. convex hull operation, to envelope the segmented lesion into the smallest convex polygon that contains the automatically segmented region, since brain metastases have a nearly spherical or pseudo-spherical appearance (Ambrosini et al. 2010). This operation comprises hole filling too, in order to include any necrotic core inside the segmented ROI;
3. logical and morphological operations based on cluster analysis (FCM on the pre-processed image with clusters), to include internal or adjacent necrotic areas do not detect by the previous convex hull operation.

5.2.3 BTV and GTV Co-registration

An image co-registration step is mandatory in order to bring the different MRI and MET-PET datasets, concerning the same patient, into the same space. In this way, it will be possible to make quantitative and meaningful comparisons between

the brain lesion segmentation results achieved by both MRI and PET segmentation methods. MRI is used as reference (target) image while the PET is the source (floating) image, because MRI conveys more anatomical information than PET. Indeed, PET imaging is characterized by weak boundaries and have a lower spatial resolution than MRI. Realignment and re-slicing operations are thus required to get an one-to-one mapping between PET and MRI slices. From the algorithmic perspective, image co-registration involves finding parameters (i.e. geometric transformation matrix) that either maximize or minimize some objective function. However an accurate interpolation method is clearly required, by which the floating and target images are sampled when being represented in different spaces (Pluim et al. 2003).

PET/MRI inter-modal 3D registration was performed using SPM 12 (Statistical Parametric Mapping, Wellcome Trust Centre for Neuroimaging, University College, London, UK), a software package designed for the analysis of brain imaging data sequences (Friston 2006).

The registration method used by SPM is based on the work by Collignon et al. (Collignon et al. 1995), where the original interpolation method has been changed in order to give a smoother cost function. The images are also smoothed slightly, by means of their histograms. This is all in order to make the cost function as smooth as possible, to give faster convergence and less chance of local minima. A 3D rigid-body model, parameterized by three translations and three rotations about the different axes, is used by means of voxel-to-voxel affine transformations. This approach is very efficient for brain anatomical district.

We chose Normalized Mutual Information (NMI) as the cost function to be optimized. The Mutual Information registration criterion states that the mutual information of the image intensity values of corresponding voxel pairs is maximal if the images are geometrically aligned. The results presented by Maes et al. (Maes et al. 1997) demonstrated that sub-voxel registration differences with respect to the stereotactic reference solution can be obtained for PET/MRI matching without using any prior knowledge about the gray-value content of both images and the correspondence between them. Mutual Information is the most intensively investigated criterion for registration of intra-individual human brain images. In addition, PET/MRI registration misalignment can be large with respect to the imaged field of view, then a criterion invariant to image overlap statistics is very important, such as the NMI proposed in (Studholme et al. 1999).

Since MET-PET and MRI multimodal images are written in a different space, an efficient resampling and interpolation method must be utilized. Although Nearest-Neighbor and Trilinear interpolations are faster, B-Spline interpolation is

recommended when subject movement can occur in the different PET and MRI scans, because they are not acquired simultaneously. In particular, we used the 4th Degree B-Spline interpolation.

Two instances of PET/MRI co-registration are shown in figure 11, where the quality of the achieved registration can be qualitatively appreciated with the checkerboard images (figure 11a and figure 11d) as well as fused PET/MR images (figure 11b and figure 11e). In addition, joint histogram, that is a feature space constructed by counting the number of times a combination of grey values occurs contextually on source (MET-PET) and target (MRI) images was also computed and plotted (figure 11c and figure 11f). Along the axes of each joint histogram the grey values of the two images are represented: from left to right for source (MET-PET) and from bottom to top for target (MRI).

Original and final joint histograms prove the quality of the Normalized Mutual Information registration process, because pixel values are mainly redistributed on the histogram diagonal.

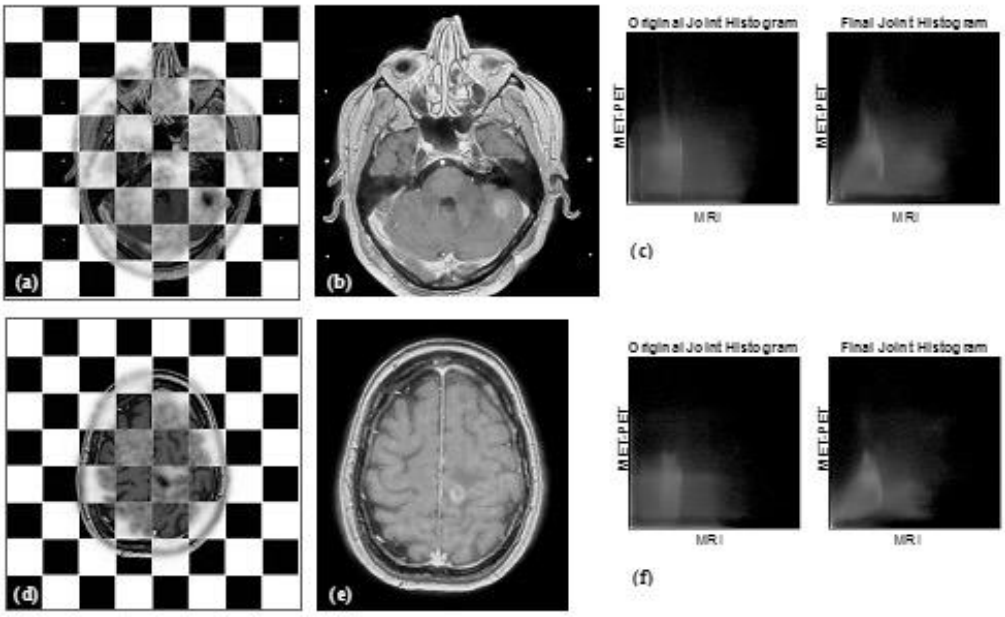

Figure 11 Instances of PET/MRI co-registration using SPM via maximization of NMI: (a, d) checkerboard images of co-registered PET and MRI; (b, e) fused images; (c, f) corresponding original and final joint histograms for (b) and (e), respectively.

5.2.4 Fully Automatic Co-Segmentation

It is worth noting, the smart combination of the two single modality pipelines in the flow diagram shown in 10. In particular, PET and MRI segmentation results are mutually exploited each other.

Firstly, the Interesting Updating Regions (IURs), obtained on PET images by the graph-based tumour segmentation method, are used for the generation of a properly area including the tumour on MR brain images. These ROI bounding regions are calculated adaptively on MR images using a Level Set method and then utilized by the MRI brain lesion segmentation method based on FCM clustering technique.

Lastly, MRI GTV masks are also combined with PET images, which are processed by the RW algorithm, in order to influence and refine tumour segmentation on PET images.

5.2.4.1 Interesting Uptake Region Detection

The fully automatic PET/MRI joint segmentation starts with the automatic identification of the PET slice with the highest SUV (SUV_{max}) on PET dataset previously co-registered with the corresponding MRI. The voxels with a SUV greater than 95% of SUV_{max} are marked as target seeds. The 8-neighborhood of the node with SUV_{max}, through searching in all 8 directions, is explored to identify the voxels with a SUV less than 30% of the average of target seed SUVs. In this way, 8 background nodes are identified. Once the foreground and background seeds are automatically localized, the PET lesion is segmented using our enhanced RW method: the probability threshold to discriminate between target and background voxels is obtained slice by slice to follow the whole lesion volume. In this way, the method takes into account the intensity gradient and contrast changes of the metabolic lesion over the entire range of PET slices. Once the BTV has been extracted, the ROI of the first PET slice (the PET slice with the SUV_{max}) is propagated to the corresponding MRI dataset. In addition, the range of segmented PET slices with high uptake regions is also provided to the next processing phase, in order to generate efficiently a ROI bounding region for the tumour on MR images.

This information defines the so-called Interesting Uptake Region (IUR).

5.2.4.2 ROI Bounding Region Generation

A coarse lesion delineation on MR images is required in order to define automatically a bounding region that encloses the actual GTV on MR slices. Hence user interaction, needed by the methods in (Militello et al. 2015) through a ROI selection tool, will be completely prevented and avoided.

Because of the variability of brain lesion in terms of location and intensity values, the definition of a valid ROI bounding region is not a trivial problem. A bad choice of the ROI bounding region on MR image can affect the whole FCM clustering process. Therefore, this task must be carried out not in a static way (e.g. using a simple morphological dilation or closing with a fixed structuring element), but rather adaptively by processing MRI input data.

This adaptive procedure starts on the lesion ROI found on SUV_{max} slice. This IUR is first dilated (using a structuring element represented by a disk of 3-pixel radius) and then utilized as a binary step function to initialize a Level Set method. A rough and over-estimated MRI brain lesion segmentation, based on the Distance Regularized Level Set Evolution (DRLSE) formulation proposed in (Li et al. 2010), is used to generate dynamically a bounding region that includes the brain lesion in the MR slice corresponding with the PET slice. Unlike Region Growing algorithm, DRLSE method allows a controlled and regular Level Set Function (LSF) evolution, by avoiding leaking in the brain tissue area. In short, level set methods represent an active contour as the zero level set of a higher dimensional function and determine the evolution of the contour.

Let $\phi : \Omega \to \Re$ be a LSF defined on a domain Ω, the energy functional is defined by:

$$E_\varepsilon := \mu R_p(\phi) + \lambda L_g(\phi) + \alpha A_g(\phi) \tag{8}$$

where:

- $\mu > 0$, $\lambda > 0$ and $\alpha \in \Re$ are the coefficients of the energy functional terms R_p, L_g and A_g, respectively;
- $R_p(\phi) := \mu \int_\Omega p(|\nabla \phi|) d\mathbf{x}$ is the level set regularization term, where $p : [0, +\infty) \to \Re$ is a potential function. In order to smooth the LSF effect and ensure accurate computation for curve evolution by maintaining the signed distance property $|\nabla \phi| = 1$, a potential function $p(s)$ with a minimum point at $s = 1$ is chosen. In

particular, a double-well potential, with two minimum points at $s=0$ and $s=1$, was used: $p(s)=\begin{cases}(1/(2\pi)^2)(1-\cos(2\pi s)) & \text{if } s\le 1\\ (1/2)(s-1)^2 & \text{if } s>1\end{cases}$;

- $L_g(\phi):=\lambda\int_\Omega g\delta(\phi)|\nabla\phi|\,d\mathbf{x}$ is an energy functional that computes the line integral of the edge indicator function $g=1/(1+|\nabla G_\sigma * I|^2)$ (G is a Gaussian kernel with standard deviation σ that is convoluted with the image $I\in\Omega$ to reduce the noise) along the zero level contour of ϕ with the Dirac delta function δ;
- $A_g(\phi):=\alpha\int_\Omega gH(-\phi)\,d\mathbf{x}$ is an energy functional that calculates weighted area of the region $\Omega_\phi^-=\{\mathbf{x}:\phi(\mathbf{x})<0\}$. This term is introduced to speed-up the motion of the zero level contour in the LSF evolution process, which is necessary when the initial contour is far away from the desired object boundary. Lastly, H is the Heaviside step function, whose derivative is the Dirac delta function δ.

In practice and for implementation purposes, the Dirac delta function δ and Heaviside step function H are approximated by their respective smooth version δ_ε and H_ε, defined as:

$$\delta_\varepsilon(x)=\begin{cases}\frac{1}{2\varepsilon}\left(1+\cos\left(\frac{\pi x}{\varepsilon}\right)\right) & \text{if } |x|\le\varepsilon\\ 0 & \text{if } |x|>\varepsilon\end{cases} \tag{9}$$

$$H_\varepsilon(x)=\begin{cases}\frac{1}{2}\left(1+\frac{x}{\varepsilon}+\frac{1}{\pi}\sin\left(\frac{\pi x}{\varepsilon}\right)\right) & \text{if } |x|\le\varepsilon\\ 1 & \text{if } x>\varepsilon\\ 0 & \text{if } x<-\varepsilon\end{cases} \tag{10}$$

where ε (usually set to 1.5) is a parameter that specifies the width of the Dirac Delta function and the slope of the Heaviside step function.

The overall energy functional in (8) can be minimized, given an initial LSF $\phi(\mathbf{x},0)=\phi_0(\mathbf{x})$, by solving the following gradient flow in the temporal variable $t\in[0,+\infty)$:

$$\frac{\partial\phi}{\partial t}=\mu\,\mathrm{div}\left(d_p\left(|\nabla\phi|\right)\nabla\phi\right)+\lambda\delta_\varepsilon(\phi)\,\mathrm{div}\left(g\frac{\nabla\phi}{|\nabla\phi|}\right)+\alpha g\delta_\varepsilon(\phi) \tag{11}$$

where div(·) is the divergence operator and the function d_p, defined as $d_p=p'(s)/s$, has a diffusion effect on the LSF. In particular, in our approach, we used

$\phi_0(\mathbf{x}) = \begin{cases} -c_0 & \text{if } \mathbf{x} \in IUR_0 \\ c_0 & \text{otherwise} \end{cases}$, where $c_0 > 0$ is a constant and $IUR_0 \in \Omega$ is the initial IUR.

Therefore, the utilized LSFs take negative values inside the zero level contour and positive values outside. This choice causes the coefficient α to take negative value to expand the active contour. The DRLSE and the temporal partial derivative $\partial\phi/\partial t$ were implemented as a finite difference numerical scheme.

Equation (8) is an edge-based geometric active contour model and represents an application of the general DRLSE formulation. This variational formulation demonstrates good adaptation segmentation results in different image processing contexts, thanks to the available free parameters.

In our implementation, on theoretical basis and experimental trials, we set the parameters as follows: $\mu = 0.2$, $\lambda = 5$, $\alpha = -1.0$, $c_0 = 3$, $\sigma = 0.8$, and a time step $\Delta t = 1.0$ It is worth to note that the parameter α needs to be carefully tuned. A non-zero value forces the motion of the contour, but the resulting final contour may slightly deviate from the true object boundary due to the shrinking or expanding effect of the weighted area term. To avoid this, the final contour can be refined by further evolving the contour for a few iterations with $\alpha = 0$ (e.g. ten iterations). Since it is required a first rough and over-estimated segmentation to define a ROI boundary region on MR images and we deal with weak object boundaries, a large value of α may cause boundary leakage. For this reason, a relatively small α was chosen to reduce boundary leakage.

These computations are performed just once for each dataset. In fact, only one bounding region is determined via DRLSE-based segmentation and it is reported onto the other candidate MR slices. For this purpose, the range of PET slices with high uptake regions is also provided by the IUR detection steps, in order to construct a cylindroid (i.e. a cylinder with irregular-shaped bases) including the whole GTV brain lesion. Since PET and MRI tumour volumes may be distributed along different axial slices, the range of PET active slice is augmented by three slices on upper and lower sides, in order to ensure the total enclosing of the tumour into the cylindroid defined on the MRI dataset.

Three examples of ROI bounding regions delineated, on original MR images, by the DRLSE-based method are shown in 12.

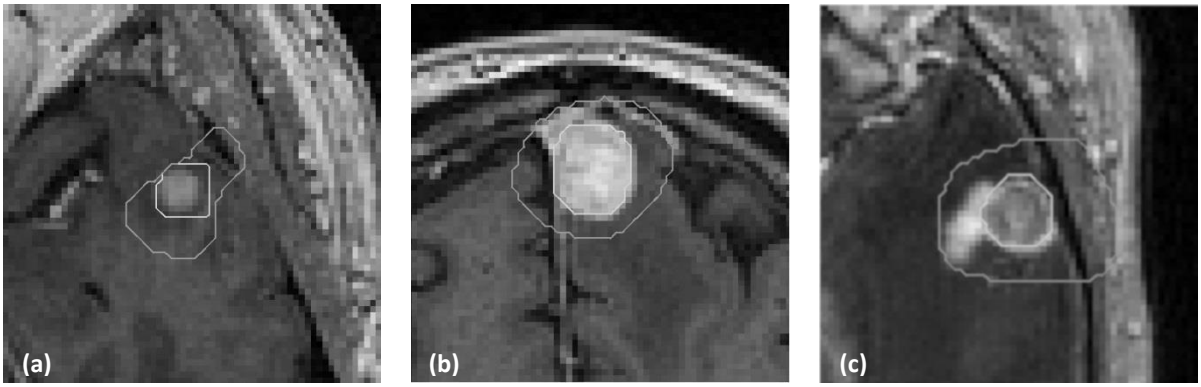

Figure 12 ROI bounding region generation on MR images: three examples of bounding region determined via DRLSE-based method, starting from IUR imported from the corresponding SUV_{max} PET slice.

5.2.4.3 MRI Segmentation based on FCM clustering

As stated before in section 5.2.2, brain tumour segmentation on MR images is accomplished using the validated semi-automatic method described in (Militello et al. 2015). However, due to the automation of the segmentation process, the direct control of the human operator is lost during ROI bounding region manual delineation. Some MR images included in the cylindroids (defined by the steps in section 5.2.4.2), especially the most distant slices from the IUR, might include no brain tumours as well as anatomical parts with pixel values similar to enhancement regions (such as bone or epidermal tissue).

For the former case, MRI brain lesion segmentation method yields an empty ROI. For the latter case, a shape-based control on connected-components belonging to the cluster with the highest intensity pixels is also defined, by taking advantage of the pseudo-spherical appearance of metastatic tumours. When either skull bone or skin are included in the ROI bounding region, since IUR is located near brain boundary, other high-valued pixel areas may be erroneously segmented by the unsupervised FCM clustering algorithm. These areas are often characterized by a lengthened shape and may be removed by checking eccentricity (ratio between the foci distance related to an ellipse and its own major axis length) and extent (ratio between the pixels of the region and the bounding-box pixels) of the connected-components. Theoretically suitable and experimentally validated feature values for this refinement step are:

$$\begin{cases} \text{Extent} \geq 0.6 \\ 0.0 \leq \text{Eccentricity} < 0.8 \end{cases} \quad (12)$$

Figure 13 shows segmentation results on the images reported in figure 12. In particular, refinement steps can be appreciated in Figure 12c, where the region composed of the skin highest pixels included in the ROI bounding region is correctly removed by the shape-based controls.

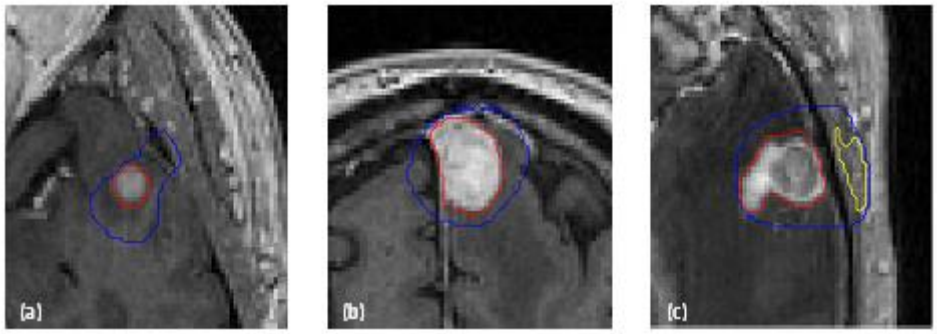

Figure 13 Examples of brain tumour segmentation on MR images (red contour), starting from the ROI bounding region (blue contour). A difficult case is shown in **(c)**, where a lengthened hyper-intense region of the head skin is detected by FCM clustering (yellow contour), but it is removed using the shape-based control.

5.2.4.4 Combining MRI Segmentation Results and PET Images

Lesion segmentation results on the morphologic brain MRI data are combined with the metabolic information of the co-registered PET images. This PET dataset represents the input of the RW algorithm. In this way, MRI GTV is utilized to combine the superior contrast of PET images with the higher spatial resolution of MR images. Each PET slice is weighted pixel-by-pixel according to the corresponding MRI binary masks. Nevertheless, an assumption that there is a one-to-one correspondence between BTV and GTV is unrealistic. Lesions may have smaller uptake regions compared to GTV. In the same way, the PET lesion can show additional area compared to lesion boundaries in MR images. For this reason, we incorporate, with extreme caution, the MRI GTV binary mask in the RW segmentation procedure. In particular, PET pixels inside the corresponding MR target are multiplied by a 15% gain factor to slightly emphasize probable lesion area; PET pixels outside the corresponding MRI ROI masks are underestimated with a 5% factor in order to correct radioactivity spill-over and spill-out effects between tumour

and surrounding tissues taking advantage of the higher spatial resolution of MR images.

5.2.4.5 PET Segmentation based on Random Walker algorithm

The RW-based segmentation process on PET dataset is updated and repeated on PET dataset combined with MRI binary masks in order to achieve a new BTV, in this paper written also as BTV_{MRI}.

Chapter 6

Experimental Results

6.1 Image Segmentation in PET

The results reported in this section have been used to produce the following research papers:

- A. Stefano, S. Vitabile, D. D'Urso, G. Russo, M. Ippolito, M. Sabini, O. Gambino, R. Pirrone, E. Ardizzone, and M. Gilardi, *Head and Neck Cancer Tumor Volume Segmentation in PET images Using an Enhanced Random Walk Algorithm*, Under Review at Medical & Biological Engineering & Computing.
- A. Stefano, Vitabile S, Russo G, Ippolito M, Sardina D, Sabini MG, Gallivanone F, Castiglioni I, and Gilardi MC. *A Graph-Based Method for PET Image Segmentation in Radiotherapy Planning: A Pilot Study*, A. Petrosino (Ed.): ICIAP 2013, Part II, LNCS 8157, pp. 711–720, 2013.
- A. Stefano, S. Vitabile, G. Russo, M. Ippolito, F. Marletta, C. D'Arrigo, D. D'Urso, M. Sabini, O. Gambino, R. Pirrone, E. Ardizzone, and M.C. Gilardi, *An Automatic Method for Metabolic Evaluation of Gamma Knife Treatments*, A. Petrosino (Ed.): ICIAP 2015, Part II, LNCS 9279, pp. 579–589, 2015.
- A. Stefano, S. Vitabile, G. Russo, M. Ippolito, F. Marletta, C. D'Arrigo, D. D'Urso, M. Sabini, O. Gambino, R. Pirrone, E. Ardizzone, and M. Gilardi, *A Fully Automatic Method for Biological Target Volume Segmentation of Brain Metastases*, International Journal of Imaging Systems and Technology, In press.

6.1.1 Phantom Study Description

PET experimental measurements have been carried out by using radioactive National Electrical Manufacturers Association International Electrotechnical Commission (NEMA IEC) body phantom to estimate the accuracy of the proposed

PET segmentation method. The phantom consists of an elliptical cylinder (D1 = 24 cm, D2 =30 cm, h= 21cm) with six spheres of different diameters (d1= 10 mm, d2 = 13 mm, d3 = 17 mm, d4 = 22 mm, d5 = 26 mm, d6 = 37 mm) positioned at 5.5 cm from the centre of the phantom. Spheres, and background are filled with radiotracer with different ratios between measured sphere radioactivity concentration and measured background radioactivity concentration (S/B). The proposed segmentation method is evaluated by matching the sphere delineation with the ground truth in the CT images. The image acquisition is performed on Discovery 690 with time of flight and Discovery STE PET/CT scanners by General Electric Medical Systems. A single PET slice consists of a matrix of 256x256 voxels of 2.73x2.73x3.27 mm3 voxel size, while a single CT slice consists of a matrix of 512x512 voxels of 1.36x1.36x3.75 mm^3 voxel size.

6.1.2 Patient Study Description

6.1.2.1 HNC

Eighteen patients with HNC, that have been referred for a diagnostic PET/CT scan before radiotherapy treatments, are enrolled. Patients fast for 12 h before the PET examination and are intravenous injected with FDG. The PET/CT oncological protocol begins 60 min after the injection. Patients breathe normally during the PET and CT examinations, and scanning is executed from the top of the skull to the middle of the thigh with the arms along the body. The BTV is manually defined by two expert radiation oncologists in consensus with the nuclear medicine physician drawing a 2D outline slice by slice: the active tumour volume includes the tumour area with an intense tracer uptake respect to background activity level.

6.1.2.2 Brain Cancer

Twelve brain acquisitions without head fixation are performed using the same PET/CT scanner used in phantom studies. Patients have fast for 4 hours before PET exam and are intravenous injected with MET. The PET protocol begins 10 minutes after the injection. The BTV is manually defined by an expert radiation oncologist in consensus with the nuclear medicine physician drawing a 2D outline slice by slice: the active tumour volume includes the tumour area with an intense tracer uptake

respect to background MET activity level. The clinical feasibility evaluation of the proposed method is done by comparing the automatic with the manual BTV segmentation.

6.1.3 Evaluation Metrics

The validation of the proposed method is achieved on phantom experiments. The same algorithm is then applied in patients that have been referred for a diagnostic PET/CT scan before treatment to assess the clinical feasibility of our segmentation method. In phantom studies the CT region must match with the PET region: the proposed segmentation method is evaluated by matching the sphere delineation with the ground truth in the CT images.

The segmentation performance is evaluated by the Dice Similarity Coefficient (DSC), median Hausdorff Distance (HD), and True Positive Volume Fractions (TPVF) (Udupa et al. 2006). The specificity (100 – false positive volume fractions) is not considered because the target voxel number with respect to the background voxel number is very small (a single PET slice consists of 65536 voxels while the largest spheres -or lesions- in a single slice is < 400 voxels).

DSC is a measurement of spatial overlap between segmented and manual BTV: a DSC of one describes a perfect match between the two volumetric segmentations, while a DSC of zero describes no overlap between the BTVs.

DSC is defined as:

$$DSC = \frac{2\left| S_{PET} \cap S_{MRI} \right|}{\left|S_{PET}\right| + \left|S_{MRI}\right|} \times 100\% \qquad (13)$$

HD is a shape dissimilarity metric measuring the most mismatched boundary pixels between automatic and manual BTV: a small average of HD value means an accurate segmentation, while a large average of HD value means no accuracy.

TPVF indicates the fraction of the total amount of tissue inside the target lesion (sensitivity).

6.1.4 Results

6.1.4.1 K-RW and AK-RW Method Validation

Phantom images are used for validation of the methods proposed in section 3.3.1.1 (K-RW method) and 3.3.1.2 (AK-RW method).

Figure 14 shows the DSC averaged on different spheres in each NEMA IEC body phantom experiment. The size of the spheres is known, and the accuracy of delineation methods is evaluated. Phantoms have been studied with an S/B of 2-3 for the NEMA (a), 3-5 for the NEMA (b), 5-6 for the NEMA (c), and 6-7 for the NEMA (d). The 30% and 90% thresholds of the SUV_{max} used to find the background and target seeds in section 3.3.1.1 are the best among those evaluated (threshold ranges from 10% to 40% and from 70% to 99%, respectively for background and target seeds).

The DSC range of PET delineation using the K-RW algorithm is found to be from 83.51% (phantom "d") up to 99.86% (phantom "a") for the sphere < 17 mm diameter with a mean of 92.95±5.90% and from 87.71% (phantom "a") up to 99.43% (phantom "d") for the sphere > 17 mm diameter with a mean of 96.17±3.48%. Considering all spheres, phantom "b" has the best mean (97.33±1.87%) and phantom "a" the worst one (93.29±5.44%).

As regards the AK-RW algorithm, the DSC range is found to be from 80.92% (phantom "c") up to 99.98% (phantom "b") for the sphere < 17 mm diameter with a mean of 89.55±7.16% and from 93.42% (phantom "a") up to 99.39% (phantom "b") for the sphere > 17 mm diameter with a mean of 97.32±1.73%. Considering all spheres, phantom "b" has the best mean (97.23±3.56%) and phantom "a" the worst one (93.45±4.77%).

The HD range of PET delineation using the K-RW algorithm is found to be from 1.27 mm (phantom "c") up to 2.71 mm (phantom "b") for the sphere with a diameter < 17 mm (HD mean = 1.72±0.28 mm) and from 0.1 mm (phantom "a") up to 2.73 mm (phantom "b") for the sphere with a diameter > 17 mm (HD mean = 0.8±0.13 mm).

The AK-RW algorithm HD range is found to be from 1.35 mm (phantom "c") up to 2.81 mm (phantom "b") for the sphere with a diameter < 17 mm (HD mean = 1.81±0.19 mm) and from 0.1 mm (phantom "a") up to 2.73 mm (phantom "b") for the sphere with a diameter > 17 mm (HD mean = 0.9±0.11 mm).

The average TPVF using the K-RW algorithm is found to be 93.18±5.73% for phantom "a", 98.70±3.20% for phantom "b", 95.85±4.28% for phantom "c", 96.81±3.71% for phantom "d".

The average TPVF using AK-RW algorithm is found to be 93.08±9.66% for phantom "a", 98.37±2.13% for phantom "b", 97.30±1.11% for phantom "c", 98.74±2.32% for phantom "d". The average specificity (100 – FPVF) is ~ 100% in all experiments.

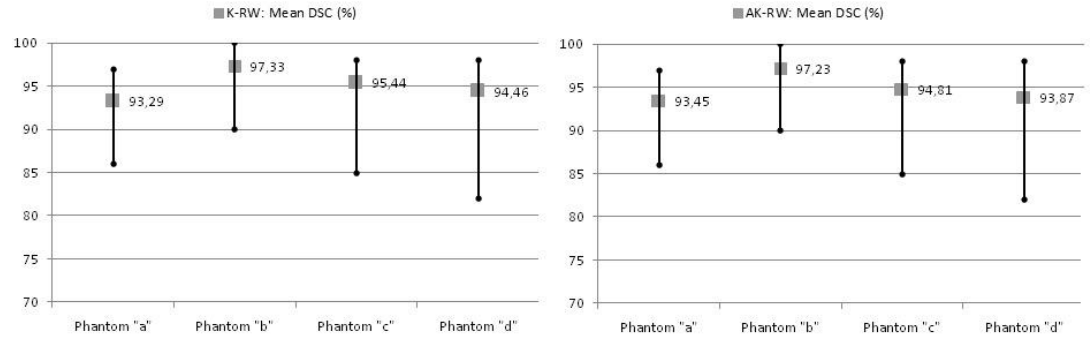

Figure 14 Mean DSC obtained from NEMA IEC body phantoms. Phantoms have a measured S/B of 2-3 for phantom "a", 3-5 for phantom "b", 5-6 for phantom "c", and 6-7 for phantom "d".

An analysis of the time performance showed that both algorithms are fast: the segmentation time for larger spheres (the whole volume consists of 12 slices) is around 4 seconds. Obviously, the time necessary for the choice of the probability threshold by the user in K-RW delineation time is excluded.

6.1.4.2 HNC

Manual delineation is obtained by averaging the segmentations performed by two nuclear medicine physicians with an inter-observer agreement of 86.51±3.65%.

Figure 15 reports the quantitative comparison between semi-automatic and manual segmentation. Results based on the DSC, HD, and TPVF show that the K-RW and AK-RW algorithms outperform the best among other tested algorithms (original RW, Threshold 40%, and Region Growing). In addition, region growing and RW methods often failed to properly delineate bifurcated lesions: figure 16 shows the segmentation task of two lesions with a complex shape obtained using the different methods. In particular, the figure shows the PET slice where the target lesion splits into two regions. In both examples, AK-RW (cyan) and K-RW

(magenta) methods correctly delineate the bifurcated lesions while region growing and RW methods fail in delineating the bifurcation. The threshold (yellow) method correctly delineates the first bifurcated lesion but it requires an accurate VOI (volume of interest) definition by the user to enclose the lesion volume and to restrict the delineation bounds.

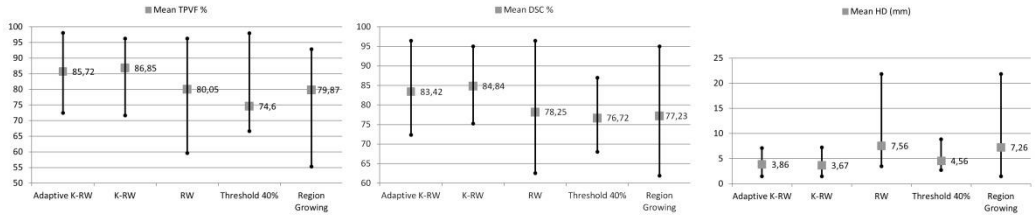

Figure 15 Mean DSC obtained from NEMA IEC body phantoms. Phantoms have a measured S/B of 2-3 for phantom "a", 3-5 for phantom "b", 5-6 for phantom "c", and 6-7 for phantom "d".

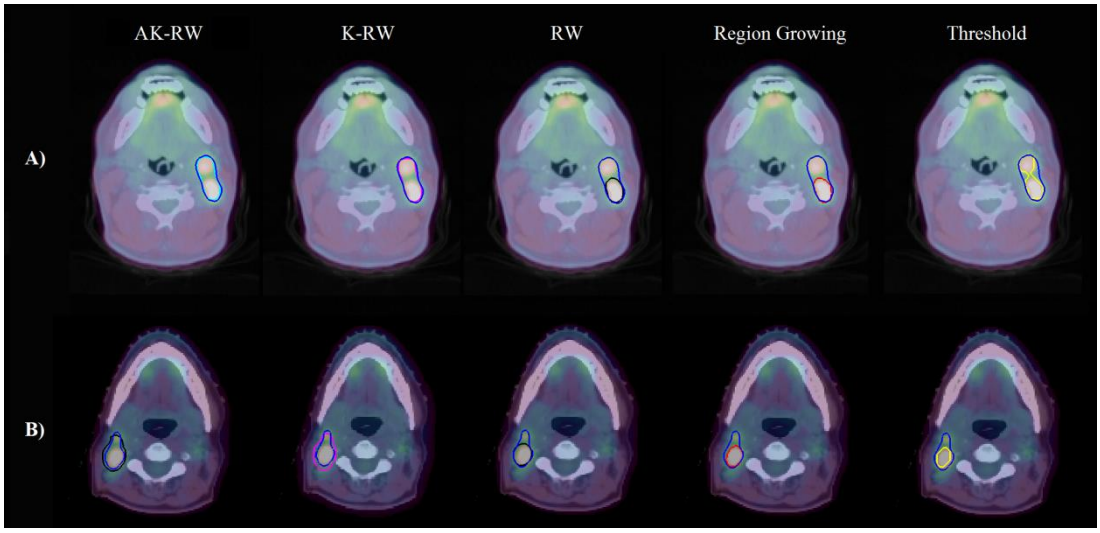

Figure 16 Segmentation examples of two bifurcated lesions (A and B rows) are shown. First column: AK-RW method (cyan) and ground truth (blue) are overlaid. Second column: K-RW method (magenta) and ground truth (blue). Third column: the RW method (black) and ground truth (blue). Fourth column: the region growing method (red) and ground truth (blue). Fifth column: the threshold method (yellow) and ground truth (blue).

In this way, false positives are removed, but the segmentation time increases considering the need to delineate the VOI. Instead, the proposed methods are able to follow the bifurcation identifying the centroids of hot regions slice after slice; also

qualitative assessment indicates that AK-RW and K-RW methods are better than other approaches to properly follow the whole PET lesion volume.

6.1.4.3 AA-RW Method Validation

Phantoms are studied with an S/B 6-7 for the NEMA (a), 5-6 for the NEMA (b), and 3-5 for the NEMA (c).

For the small sphere (diameter < 17 mm), the DSC range of the AA-RW delineation is found to be from 81.23% up to 99.99%, with a mean of 89.44±6.19%. For the large spheres (diameter ≥ 17 mm), the AA-RW DSC range is found to be from 95.01% up to 99.89%, with a mean of 98.30±1.66%. The average TPVF is 98.74±2.32% for phantom "a", 97.30±1.11% for phantom "b", and 98.37±2.13% for phantom "c". HD ranges from 2.03 mm up to 3.66 mm for the small spheres, with a mean of 2.57±0.46 mm. In the same time, HD ranges from 1.69 mm up to 2.73 mm for the large spheres, with a mean of 1.96±0.28 mm.

The resultant DSC, TPVF, and HD scores of all implemented methods (original RW, Threshold 40%, Region Growing, and Fuzzy C-Means) are shown in figure 17. Mean values and standard deviations of the DSC, TPVF, and HD for small and large spheres and for the seven methods are given in table 1.

Time performance analysis shows that the algorithm is fast: the segmentation time for the larger sphere (the whole volume consists of 12 slices) is about 3 seconds.

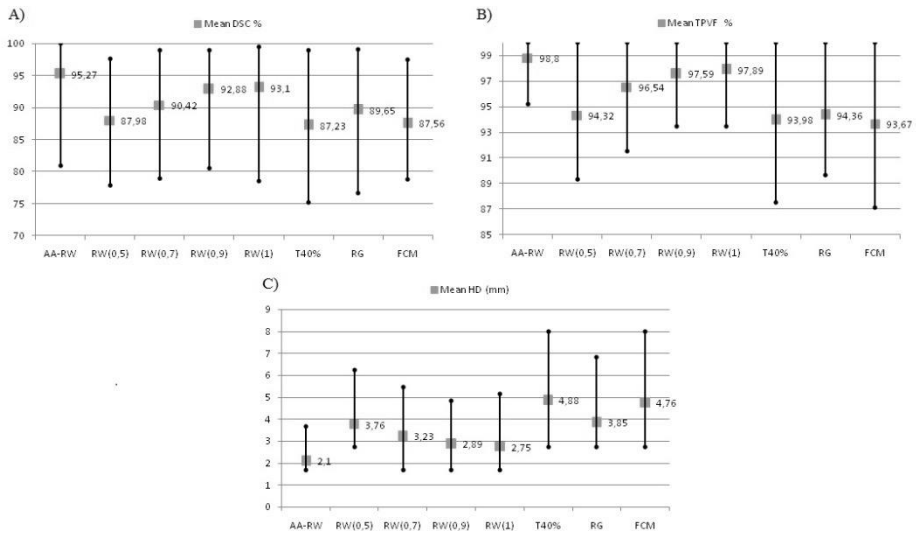

Figure 17 Mean DSC, TPVF, HD and the standard deviation obtained for phantom studies.

	Small spheres (d<17mm)				*Large spheres (d≥17mm)*		
Method	**DSC(%)**	**TPVF(%)**	**HD(mm)**		**DSC(%)**	**TPVF(%)**	**HD(mm)**
AA-RW	89.44±6.19%	99.57±1.06	2.57±0.46		98.30±1.66%	97.92±1.03	1.96±0.28
RW(0.5)	82.74±8.17%	96.17±1.76	5.24±0.68		91.04±2.19%	93.92±1.15	3.06±0.66
RW(0.7)	84.77±7.69%	98.57±1.11	4.89±0.91		93.46±2.16%	94.92±1.13	2.95±0.35
RW(0.9)	85.94±6.83%	99.03±1.21	4.01±0.29		94.24±1.23%	96.18±1.24	2.38±0.36
RW(1)	85.71±7.35%	98.95±1.15	3.86±0.48		95.36±1.19%	96.83±1.13	2.19±0.46
T40%	81.23±8.90%	95.07±1.25	6.71±0.59		90.44±2.02%	89.92±2.25	4.26±0.58
RG	83.40±8.04%	97.81±1.04	4.99±0.88		91.25±1.19%	94.22±1.82	3.35±0.48
FCM	81.93±6.90%	95.97±1.29	6.14±0.73		89.91±3.12%	88.22±2.89	4.43±0.60

Table 1 Mean values of the DSC, TPVF, and HD values for small and large spheres.

6.1.4.1 Brain Cancer

The AA-RW delineation is not subjected to both intra and inter-operator variability. All segmented radioactive lesions are homogenous (one single peek is visually detected on the histogram of lesion voxel values). The metabolic volume is 3.02±2.24 ml (range: 1.19-10.44 ml).

The DSC range of PET delineation using the AA-RW algorithm ranges from 81.7% up to 92.6%, with a mean of 88.57±3.22%. The HD ranges from 1 mm up to 3.67 mm, with a mean of 1.96±0.62 mm. The average TPVF is found to be 91.13±6.75% (range: 76.3% - 100%). High DSC and TPVF values, and a low HD value confirm the accuracy of the proposed method.

The difference between manual and automatic contour delineation in three MET-PET examinations is shown in figure 18.

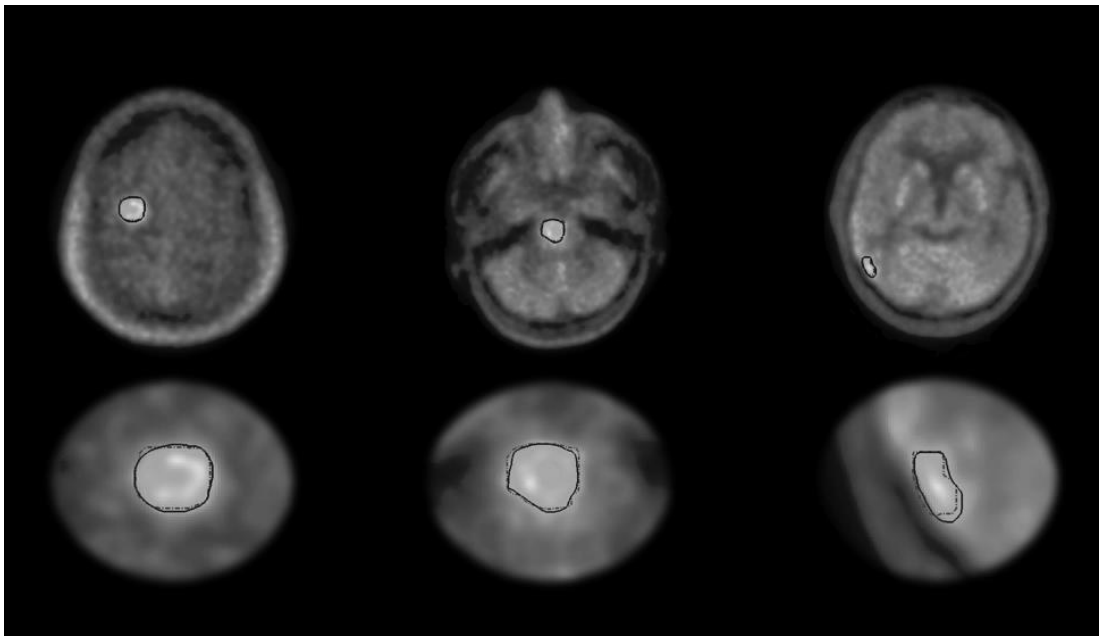

Figure 18 Three different segmentation examples of MET uptake regions. AA-RW method (non-continuous line) and manual segmentation (continuous line) are overlaid.

6.1.5 Discussion

PET delineation is a critical task due to physical factors (size of the detector crystals, positron range in matter, non-collinearity of annihilation gamma-rays and detector scatter) which leads to a blurred margin of the lesions. To date, manual contouring is the most commonly used method in radiotherapy planning. The manual segmentation approach is dependent on the experience of the nuclear physician limiting the measurement accuracy. In addition, the manual segmentation is time-consuming and impaired by inter and intra-observer variability. Clinical analysis of PET studies is performed on software platforms with predefined window-level setting. This setting can be liberally modified by the nuclear physician, altering the visual appearance of PET lesions. To reduce these issues, several automatic methods have been presented, but few clinical studies are available and there is no consensus for proper BTV determination. Efforts to generate an automatic or semi-automatic

framework for a precise PET contouring are needed in order to increase the BTV reproducibility and reliability in the planning of radiotherapy treatment.

The aim of this thesis is to validate a PET image segmentation method based on RW. An adaptive probability value is introduced to make the proposed method independent from the choice of the β factor in the Gaussian weighting function (5).

In addition, the proposed algorithm identifies the RW foreground and background seeds to obtain an automatic method, satisfying a very critical requirement in a clinical environment. The fully automatic identification of target and background seeds is feasible in brain studies because the MET PET datasets for Gamma-Knife treatments include only the brain area avoiding the possible presence of false positives in other anatomic regions. Differently, in total body PET examinations, such as those carried out in HNC study, a user interaction to manually identify the target lesion is always needed (normal structures such as brain, heart, bladder, and kidneys normally have high FDG uptake).

In addition, the proposed methods are very fast if compared with the time needed for manual segmentation in a clinical environment.

High DSC and TPVF, and low HD values confirm the accuracy of the RW-based methods. In HNC study, the AK-RW algorithm is slightly less accurate than the K-RW method, but it is expected because K-RW is a semi-supervised algorithm: the user can choose the best among probability threshold values to properly delineate the PET spheres. AK-RW obtains good segmentation results besides having the benefit of requiring a low user interaction effort and low levels of user's specialist knowledge being independent of the choice of the probability threshold value to discriminate target from background regions. The development of user independent techniques able to perform good segmentation is crucial in a clinical environment. In addition, AK-RW method follows the whole lesion volume taking into account the gradient of intensity and contrast changes of the PET lesion in different slices.

A slight drop in the accuracy in phantom experiments occurs for lesions with diameter < 1.7 cm. In patient's studies, the mean metabolic volume is greater than 3 ml (lesions with a sphere-equivalent diameter > 1.8 cm): since target dimensions are larger than 2–3 times the FWHM (full width at half maximum) of the point spread function of the PET scanner, the underestimation of metabolic activity caused by PVE can be assumed to be negligible. In the lesions with a sphere-equivalent diameter < 1.7 cm a partial volume correction method could be included in the algorithm.

Despite the fact that many studies use the information of co-registered PET and CT images to identify features for distinguishing a lesion from the background and, consequently, for PET image segmentation, we believe that BTV extraction is

independent from anatomical segmentation being inappropriate to consider a matching between anatomical and metabolic regions, with special reference in HNC district. Head and neck lesions may have different uptake margins compared to anatomical margins; PET and CT show complementary features. In a phantom study the CT region must match the PET region. This is not true in HNC patient studies. It is not appropriate to consider a one-to-one relationship between anatomical and functional images. The metabolic volume cannot match the tumor anatomic extension, showing different and additional information including not visible metastases using traditional radiologic techniques. In addition, mis-registration between the two series can occur due to patient's motion. Then, the assumption of identical HNC boundary in PET and CT images is questionable and unrealistic. Uptake regions may be larger\smaller than the anatomical boundary of the tumor due to cellular activation in target tissues. I believe that a correct delineation of BTV must be obtained without incorporating anatomical information or incorporating them with great attention. For example, although the patients with brain cancer undergo a PET/CT examination, CT has strong limitations for target delineation of brain metastases and this makes impossible to apply a multi-modal PET/CT image segmentation. Differently, the high resolution of the MRI is important for the illustration of the anatomy, and a co-registration PET/MRI is essential in the high-precision radiotherapy of brain tumors: a study to determine simultaneously tumor contours on PET and RM images is the topic of the fifth chapter.

6.2 Metabolic Evaluation of Oncological Treatments

The results reported in this section have been used to produce the following research papers:

- Stefano, Russo G, Ippolito M, Cosentino S, Murè G, Baldari S, Sabini MG, Sardina D, Valastro L, Bordonaro R, Messa C, Gilardi MC and Soto Parra H. *Evaluation of erlotinib treatment response in non-small lung cancer using metabolic and anatomic criteria*, QJ Nucl Med Mol Imag, EPUB ahead of print.
- Stefano, N. Porcino, G. Banna, G. Russo, V. Mocciaro, G. Anile, S. Gieri, S. Cosentino, G. Murè, S. Baldari, M.G. Sabini, D. Sardina, F. Fraggetta, S. Vitabile, M.C. Gilardi, and M. Ippolito. *Metabolic Response Assessment in Non-Small Cell Lung Cancer Patients after Platinum-Based Therapy: A*

Preliminary Analysis, Current Medical Imaging Reviews, vol. 11, n. 4, pp. 218-227, 2015.

- Stefano, S. Vitabile, G. Russo, M. Ippolito, F. Marletta, C. D'Arrigo, D. D'Urso, M. Sabini, O. Gambino, R. Pirrone, E. Ardizzone, and M.C. Gilardi, *An Automatic Method for Metabolic Evaluation of Gamma Knife Treatments*, A. Petrosino (Ed.): ICIAP 2015, Part II, LNCS 9279, pp. 579–589, 2015.

6.2.1 Clinical Case Description

6.2.1.1 Erlotinib treatment

Twenty patients, with NSCLC stage IV after at least one course of chemotherapy, with a performance status of asymptomatic, or symptomatic but completely ambulatory, scheduled to undergo Erlotinib therapy are considered (11 men and 9 women, mean age: 64.2±11.8 years, range 38 to 79 years). Twelve patients (60%) have a history of smoking. The histopathologic diagnosis is: 4 adenocarcinomas (20%), 12 non-adenocarcinomas (60%), 4 NOS (not otherwise specified) (20%).

All patients receive the same Erlotinib neoadjuvant chemotherapy treatment (150 mg orally once daily for 45 days) at the Department of Oncology, Garibaldi Nesima Hospital of Catania. Patients undergo a FDG PET/CT scan before treatment (PET_I) and after administration of 3 doses of Erlotinib at 48 h (PET_{II}) for early monitoring. A further metabolic follow-up is planned at the end of Erlotinib treatment, 45 days from the start of therapy (PET_{III}).

Six patients drop out at the last PET scan (PET_{III}): 2 patients undergo PET/CT scan in a different PET institution, 2 patients interrupt therapy, and 2 patients need of additional therapy. For the 14 patients who completed the protocol, morphological diagnostic CT scans after the administration of 100 ml of intravenous iodate contrast material (Iomeprolo, Iomeron 300) are performed before and 45 days after the initialization of Erlotinib treatment.

For each PET scan, the lesion presenting with the highest FDG uptake is considered (not necessarily the same lesion on three PET studies) according to EORTC, PERCIST and RECIST classifications to assess metabolic and anatomic response. The ratio (SUV_{ratio}) between SUV_{max} and mean SUV in the liver is considered, in order to limit intra-subject uptake variability. The region of interest in the liver is placed on the central region of the right lobe. EORTC and EORTC ratio

criteria are applied to classify SUV_{max} and SUV_{ratio}, respectively.

In addition, the metabolic volume is calculated using the method described in the third chapter. Median values of metabolic (EORTC, and PERCIST) and morphological (RECIST) response classifications are statistically compared using Wilcoxon signed-rank test. Disease-Free Survival (DFS) and Overall Survival (OS) time of response classifications are calculated by the Kaplan-Meier test. Median OS and median DFS are compared between responders and non-responders using a long-rank test. A p value less than 0.05 is considered statistically significant.

6.2.1.2 Platinum-based therapy

Seventeen inoperable patients with stage IV NSCLC are enrolled. PET studies is carried out before the initiation of platinum-based therapy and after the first cycle of chemotherapy for an early therapy monitoring. Patients received a platinum-based chemotherapy according to the American Society for Clinical Oncology or European Society for Medical Oncology guidelines, with carboplatin or cisplatin in combination with another third generation agent, with possible addition of bevacizumab. PET/CT measurements without contrast are performed on a time-of-flight Discovery 690 scanner (General Electric Medical Systems). Patients fast for 12 h before the PET examination and are intravenous injected with FDG. The PET/CT oncological protocol begins 60 min after the injection. Patients breathe normally during the PET and CT examinations, and scanning is executed from the top of the skull to the middle of the thigh with the arms along the body. Patients underwent a 18F-FDG PET/CT examination without contrast before platinum based chemotherapy (PET0) and 28±6 days after the initiation of therapy (PETI) for early monitoring. Morphological diagnostic CT scans before and between day 49 and 88 (median, 68 days) after the initiation of treatment are performed.

The lesions with the highest uptake in each patient are evaluated according to EORTC considering a cut-off of 15% (EORTC15%) and 25% (EORTC25%) to discriminate between patients who respond from those who do not respond to treatment. Moreover, PERCIST and RECIST classifications are evaluated too.

Receiver operating characteristic curves are used to obtain cut-off points for therapy evaluation based on variations of TLG and MTV in sequential scans. Overall Survival (OS) time is calculated by the Kaplan-Meier test. A p value less than 0.05 is considered statistically significant.

6.2.1.3 Gamma Knife Treatment

Eighteen patients with brain metastases that have been referred for a diagnostic PET/CT scan before Gamma Knife treatment are enrolled. PET brain acquisitions without head fixation are performed. Patients have fast for 4 hours before PET exam and are intravenous injected with MET. The PET protocol begins 10 minutes after the injection.

Lesion segmentation is used to evaluate therapeutic response using SUV_{max}, MTV, and TLG variations in sequential scans. MTV provides metabolic volumetric information of the tumors; TLG, defined as MTV x (average SUV within the MTV), combines the volumetric and SUV information, to try to obtain a better evaluation of the treatment response.

Variations (Δ) in PET parameters are normalized to baseline:

Δ(%) = 100 x (post-treatment – baseline)/baseline.

6.2.2 Results

6.2.2.1 Erlotinib treatment response

Metabolic Response Evaluation at 48 h

As regards the EORTC and EORTC ratio classifications, some patients are classified differently, but the classes of metabolic response are overall equal: 7 patients show a Partial Metabolic Response (PMR), 10 patients a Stable Metabolic Disease (SMD), and 3 patients a Progressive Metabolic Disease (PMD).

According to PERCIST: 2 PMR, 17 SMD, and 1 PMD.

There is a significant difference in response evaluation between EORTC and PERCIST (Wilcoxon signed-rank test, $p<0.02$). To perform the Kaplan-Meier analyses, the patients are split into responders (PMR) and non-responders (SMD and PMD). Figure 19 shows Kaplan-Meier curves for EORTC, EORTC ratio, and PERCIST classifications. The analyses show that only EORTC proved a significant prognostic factor for predicting DFS (hazard ratio [HR] =0.364; 95% confidence interval [CI], 0.149-0.889; p=0.011) and OS (HR = 0.332; 95% CI, 0.133-0.827; p=0.009) with a median of 103 and 410 days, respectively, in responder patients. Non-responders have a median DFS of 47 days and a median OS of 104 days.

EORTC ratio and PERCIST are not a significant factor associated with DFS (p-value=0.054 and p-value=0.89, respectively) and OS (p-value=0.19 and p-value=0.68, respectively) in this study.

Table II summarizes median DFS and OS days for responder and non-responder patients.

Metabolic and Morphological Response Evaluation at 45 days

The morphological responses according to RECIST are as follows: 4 Partial Response (PR) and 10 Progressive Disease (PD). Metabolic responses according to EORTC and EORTC ratio criteria are coincident (all patients are equally classified): 5 PMR, 5 SMD, 4 PMD. According to PERCIST: 4 PMR, 6 SMD, and 4 PMD. There is a significant difference in the results of response classification between metabolic criteria (EORTC, EORTC ratio, and PERCIST) and RECIST ($p<0.04$).

RECIST is a significant prognostic factor for DFS (HR =4.070; 95% CI, 1.383-11.972; p=0.002) and OS (HR =2.620; 95% CI, 0.8806-7.795; $p<0.046$). Both EORTC and PERCIST show significant prognostic factors for DFS (EORTC: HR = 0.319; 95% CI, 0.061-1.667; p=0.028; PERCIST: HR =0.245; 95% CI, 0.0835-0.722; p=0.001), but only the PERCIST criteria show a good agreement with OS (EORTC: p=0.65; PERCIST: HR =0.3817; 95% CI, 0.1283-1.1356; p=0.046).

For responder patients, median DFS is 287.5 days for the RECIST and PERCIST and 79.5 days for EORTC. For non-responder patients, median DFS is comparable for all response criteria (50 days for RECIST and PERCIST, and 49 for EORTC). RECIST and PERCIST show a good prognostic factor for the median OS: 480 vs 174 days (for responders versus non-responders).

Table III summarizes median DFS and OS days for responders and non-responders. Figure 20 shows Kaplan-Meier curves for RECIST, EORTC, and PERCIST classifications.

Two representative PET lesions for a patient affected by multiple metastases are presented in figure 21 (PET_I, PET_{II}, PET_{III}). The lesion with the highest 18F-FDG uptake is the same in PET_I and PET_{II} but different in PET_{III}. Metabolic response at 48 h is PMD according to EORTC ($\Delta SUV=+53.3\%$), and SMD according to PERCIST ($\Delta SUL_{peak}=+21.4\%$). Metabolic response at 45 days is PMD for EORTC and PERCIST ($\Delta SUV=+25.87\%$ and $\Delta SUL_{peak}=+41.2\%$), whereas morphological response according to RECIST is PD. The patient presents recurrence of disease at 51 days and died at 147 days.

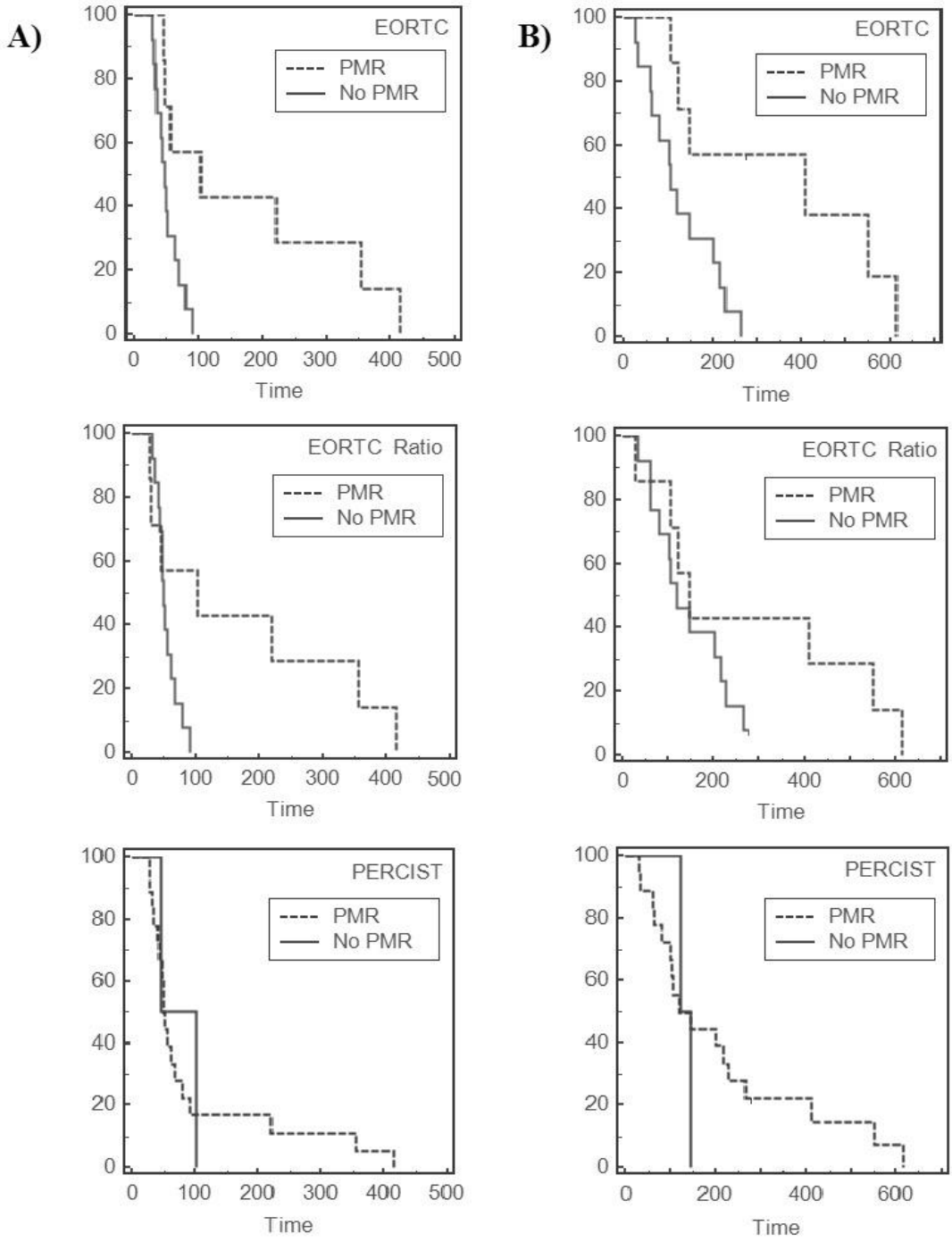

Figure 19 Kaplan-Meier analyses at 48 h. DFS (column A) and OS (column B) in patients after administration of 3 doses of Erlotinib according to EORTC, EORTC ratio and PERCIST based on the classification between responders (PMR) and non-responders (SMD and PMD). Only EORTC is a significant prognostic factor for predicting DFS and OS ($p<0.011$).

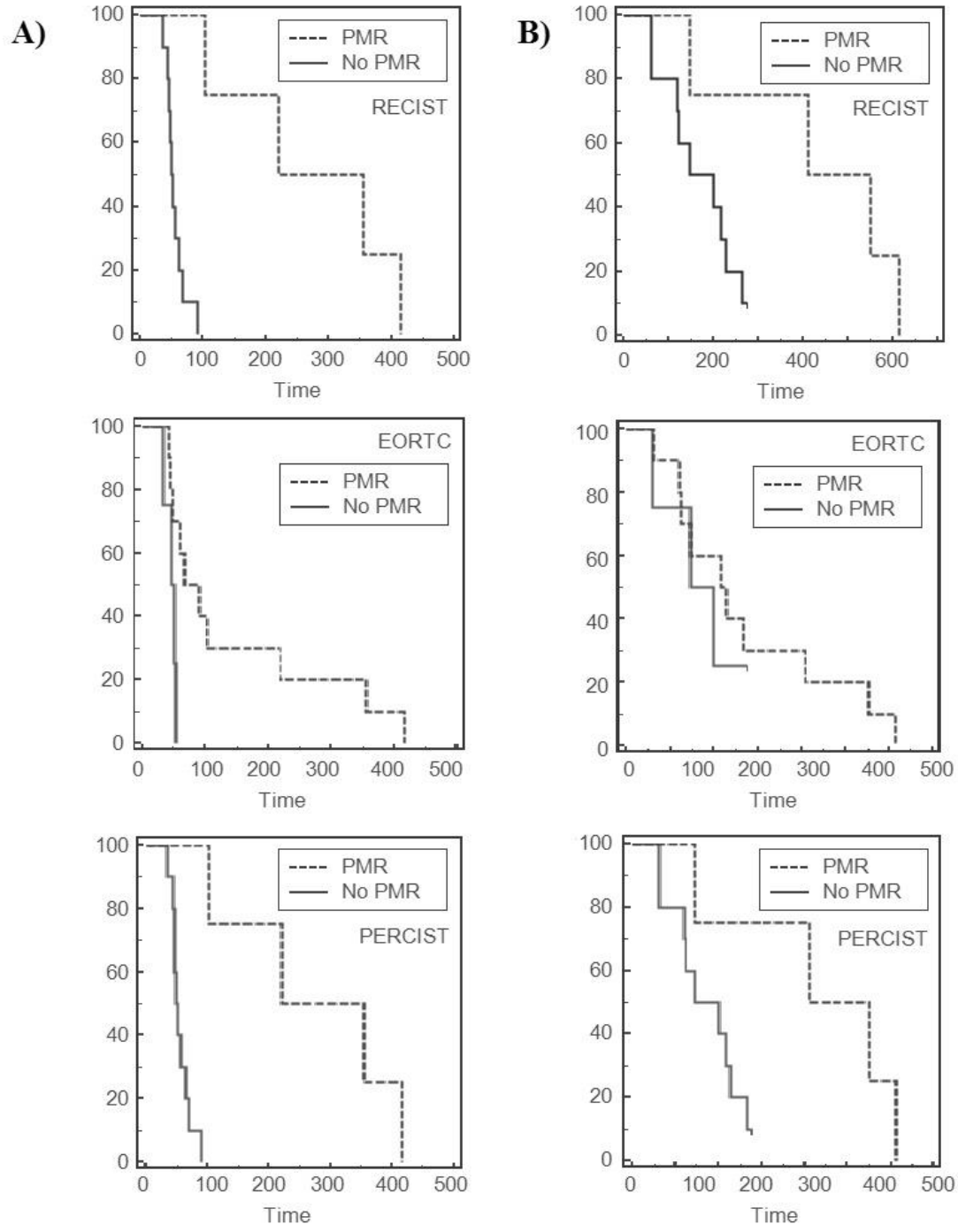

Figure 20 Kaplan-Meier analyses at 45 days. DFS and OS in patients according to RECIST, EORTC, and PERCIST based on the classification between responders and non-responders. RECIST and PERCIST are significant prognostic factors for predicting DFS and OS ($p<0.05$). EORTC is a significant prognostic factor for predicting DFS ($p=0.028$) but not OS ($p=0.65$).

At 48 h	DFS responders	DFS Non-responders	OS responders	OS Non-responders
	Median Days			
EORTC	103	47	410	104
EORTCratio*	103	49	147	119
PERCIST*	74	49	134	119

*No significant prognostic factor for DFS and OS

Table II Comparison of Metabolic Response at 48 h.

At 45 days	DFS responders	DFS Non-responders	OS responders	OS Non-responders
	Median Days			
RECIST	287.5	50	480	174
EORTC*	79.5	49	222	174
EORTC ratio*	79.5	49	222	174
PERCIST	287.5	50	480	174

*No significant prognostic factor for OS

Table III Comparison of Metabolic and Morphological Response at 45 days.

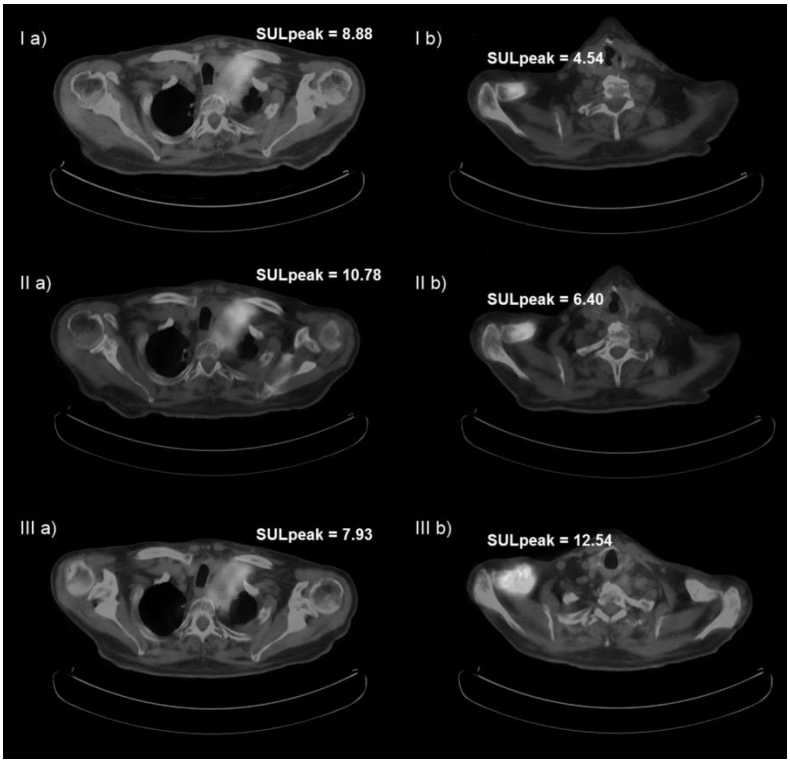

Figure 21 Two representative PET lesions for a patient affected by multiple metastases.

6.2.2.2 Platinum-based therapy response

For the 17 patients who completed the protocol, 11 are men and 6 women (mean age: 65±11 years, range 35 to 81 years). Fifteen (88.23%) had a history of smoking and among these 8 patients (52.94%) are current smokers. The histopathologic diagnosis is: 11 adenocarcinomas (64.71%), 4 squamous carcinomas (23.53%), and 2 not otherwise specified (11.76%).

According to $EORTC_{25\%}$ 5 patients show a PMR, 9 patients a SMD, and 3 patients a PMD; according to $EORTC_{15\%}$ 9 PMR, 5 SMD, and 3 PMD; according to PERCIST 4 PMR, 10 SMD, 3 PMD; according to RECIST 1 CR, 4 PR, 4 SD, and 8 PD. MTV and TLG averages of responder patients are lower than those of non-responder patients for all classifications. $EORTC_{15\%}$ and RECIST classifications show SUV and SUL averages lower in patients who responded to treatment rather than in non-responders.

The Kaplan-Meier analyses show that a statistically significant difference between responder and non-responder patients is recognized according to $EORTC_{15\%}$ (HR=2.97 ; CI, 0.89-9.95; p =0.0251) and RECIST (HR=4.80 ; CI, 1.61- 14.27; p=0.0154) classifications. No statistically significant difference between responders and non-responders is observed in the $EORTC_{25\%}$ and PERCIST classifications (p=0.13 and p=0.18, respectively).

Figure 22 shows Kaplan-Meier curves for all classifications.

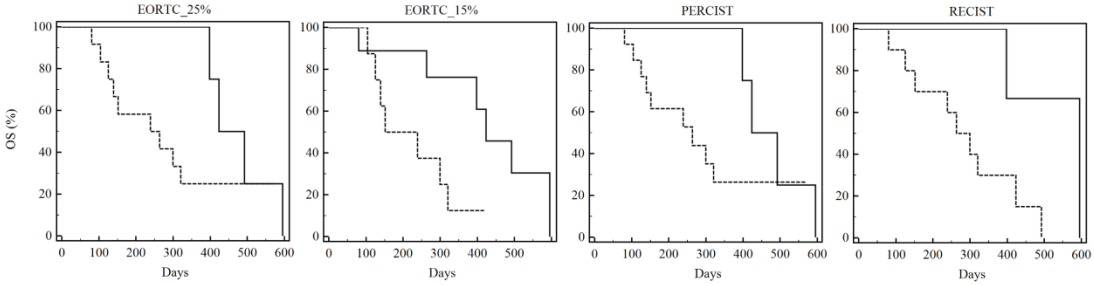

Figure 22 OS curves according to $EORTC_{25\%}$, $EORTC_{15\%}$, PERCIST and RECIST based on the division between responders (continuous line) and not responders (non-continuous line).

For responders, median OS is 423 days for the EORTC15%, and 595 for RECIST. For non-responders, median OS was 194.5 and 238 for $EORTC_{15\%}$ and RECIST, respectively.

The ROC study suggests a ΔTLG cut-off value of -36%, and a ΔMTV cut-off value of -8%. ΔTLG (HR=3.22; CI, 1.07-9.71; p =0.0145) and ΔMTV (HR=4.77; CI, 1.21-18.71; p =0.0005) are significant factors for OS.

In addition, in the same way in which EORTC proposes an early threshold of 15% after one cycle of chemotherapy, the ROC analysis to identify an early PERCIST threshold of 17% (PERCIST17%) rather than the conventional cut-off of 30% is used. In this way, PERCIST17% is a significant factor for OS (HR= 4.63; CI, 1.45-

14.75; p =0.0016), unlike the original PERCIST that is inadequate to discriminate between patients who respond from those who do not respond to treatment. Figure 23 shows Kaplan-Meier curves. ΔTLG and $PERCIST_{17\%}$ discriminations between responder and non-responder patients show a median OS of 492 days for responder patients and 238 (ΔTLG) and 194.5 ($PERCIST_{17\%}$) days for non-responder patients. ΔMTV reveals a median of 423 versus 188 days.

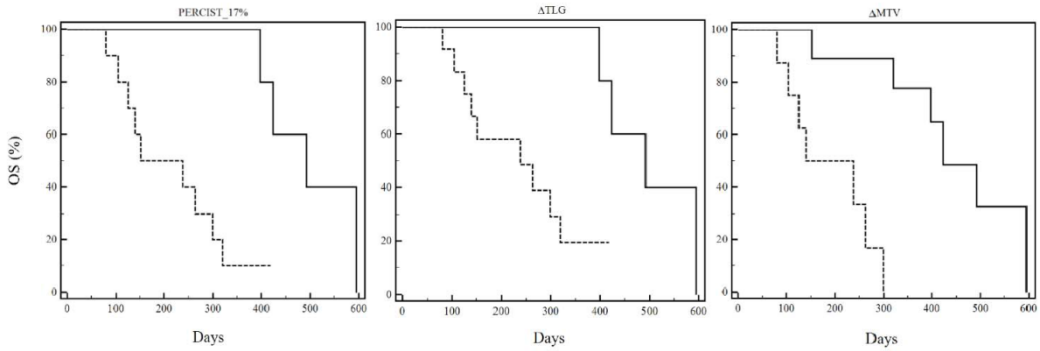

Figure 23 OS curves according to $PERCIST_{17\%}$, ΔTLG, ΔMTV using cut-off points obtained by ROC analysis to discriminate between patients who respond (continuous line) from those who do not respond to treatment (non-continuous line).

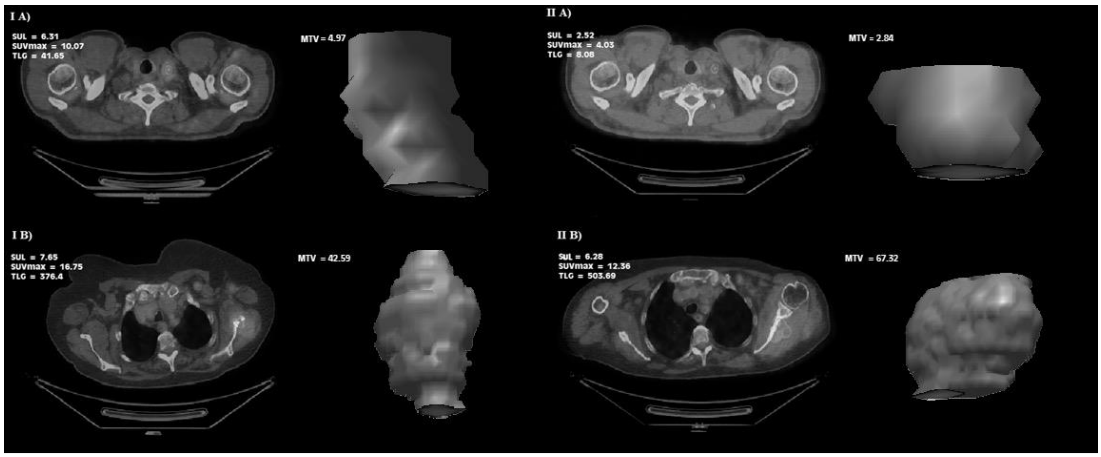

Figure 24 The highest 18F-FDG uptake lesion before treatment (I) and after treatment (II) with relative MTV in two patients with multiple metastases. Patient A is classified as a responder in all metabolic classifications showing 60% decrease in SUV and SUL, 80% decrease in TLG, and 42% decrease in MTV. Patient B is classified as a responder in $EORTC_{25\%}$ and $EORTC_{15\%}$ (-26%), and as a non-responder according to PERCIST (-18%), ΔTLG (+33%), and ΔMTV (+58%).

The 1-year OS are 100% and <25% for responder and non-responder patients, respectively, in agreement with RECIST, ΔTLG and $PERCIST_{17\%}$. Both $EORTC_{15\%}$

and ΔMTV have poorer OS than RECIST and ΔTLG (78% versus 12% and 78% versus 0%, respectively).

Representative PET images for two patients (A and B) affected by multiple metastases are presented in figure 24. The lesion with the highest FDG uptake is the same in PET_0 and PET_I in both patients. For the first patient, metabolic response is PMR according to EORTC (both $EORTC_{25\%}$ and $EORTC_{15\%}$), PERCIST, ΔTLG, and ΔMTV. The patient is classified as CR according to RECIST and is censored for OS after 498 days.

6.2.2.3 Gamma Knife Treatment

About Gamma Knife treatment response evaluation, results are shown in Table IV. Most patients show the same treatment response trend using SUV_{max}, MTV, and MTV variations in sequential scans. Differently, the third patient shows a minimal change in SUV_{max} (-1.00%) and great variation in the metabolic tumor mass and in TLG (-61.12% and -59.43%, respectively). Patients number 4, and 9 show a positive ΔSUV_{max} (+9.47% and +14.73%, respectively), and a negative ΔTLG, and ΔMTV (-16.89% and -8.50%, +49.98% and +39.74%, respectively).

The sixth patient study shows a positive ΔMTV (+24.85%), and a negative ΔSUV_{max} (-44.36%), and ΔTLG (-26.15%). At last, patients number 8 and 12 show a visual disappearance of metabolic lesions indicating a complete metabolic response (CMR) to treatment.

Patients	ΔSUV_{max}	ΔMTV	ΔTLG
#1	-30.10%	-62.20%	-66.67%
#2	-26.21%	-31.53%	-43.83%
#3	-1.00%	-61.12%	-59.43%
#4	9.47%	-16.89%	-8.50%
#5	-41.83%	-81.16%	-83.86%
#6	-44.36%	24.85%	-26.15%
#7	-62.72%	-20.22%	-64.93%
#8	CMR	CMR	CMR
#9	14.73%	-49.98%	-39.74%
#10	-39.64%	-13.27%	-10.21%
#11	-37.00%	-30.95%	-56.50%
#12	CMR	CMR	CMR

Table IV Metabolic response obtained on the segmented lesions for each patient.

6.2.3 Discussion

The early identification of responder and non-responder patients is mandatory to avoid inefficient treatments and to preserve the patient's quality of life. Unnecessary administration of expensive, potentially non-curative, and toxic drugs in subtypes of cancers that are refractory or resistant to specific treatment could be avoided by the choice of an appropriate strategy for the evaluation of therapy response.

For early response, after 48 h from the first Erlotinib administration, PET images were assessed according to EORTC, EORTC ratio, and PERCIST classifications. Kaplan-Meier analyses showed that EORTC was able to discriminate responder from non-responder patients, predicting DFS and OS (p-value<0.011), and that EORTC ratio and PERCIST were inadequate to predict DFS and OS (p-value>0.05). The difference between the PERCIST threshold (30%) and the EORTC threshold (15%) in differentiating responders from non-responders could explain this result in the early evaluation.

For treatment response, at 45 days from the first Erlotinib administration, both morphological (RECIST) and metabolic (EORTC, EORTC ratio and PERCIST) classifications were evaluated. Statistical analyses showed that RECIST and PERCIST were able to classify responders from non-responders (p-value<0.05). EORTC and EORTC ratio classifications were not suitable to discriminate responders from non-responders: this result could be affected by statistical variation of SUVmax due to the high noise in PET images.

In Platinum-based therapy response study, the Kaplan-Meier analysis showed that RECIST, $EORTC_{15\%}$, ΔTLG, and ΔMTV proved to be a significant prognostic factor for predicting OS. Conversely, $EORTC_{15\%}$ and PERCIST classifications were inadequate to discriminate between patients who respond from those who do not respond to treatment. $EORTC_{15\%}$ classification identified 9 patients with a PMR, PERCIST classification identified only 4 patients with a PMR: if we evaluate the data with the PERCIST analysis, 5 patients, classified as non-responders, should change therapy, which, instead, according to $EORTC_{15\%}$ recommendations are classified as responders and could continue with the therapy that is working. The presence of an early threshold in $EORTC_{15\%}$ could explain this result. The 1-year OS was 100% for responder patients, in agreement with RECIST and ΔTLG analyses. EORTC15% had a poorer OS (78%) than RECIST and ΔTLG. In addition, in the second case study in figure 21, ΔTLG classified the patient as a non-responder unlike the RECIST and EORTC classifications that classified the patient as a responder (the patient died after 132 days): TLG has the advantage that it reflects both the metabolic information and the lesion size, whereas RECIST is exclusively

based on anatomical information and EORTC is mainly based on metabolic information.

Al last, MTV and TLG can reflect metabolic changes in sequential PET scans after Gamma Knife treatment throughout the entire tumor mass. These parameters should be more accurate methods of detecting global changes than a single-pixel value measurement such as SUV_{max}. The results showed that the patient number 3, 4, and 9 could be a demonstration of this theory. The sixth patient showed a positive ΔMTV, and a negative ΔTLG that is indicative of partial response to treatment. TLG combines the volumetric (MTV) and the semi-quantitative parameter (SUV) information: it probably provides a better evaluation of the prognosis compared to MTV. In conclusion, PET delineation in neurosurgical radiosurgery appears helpful for assessing the effects of therapy in brain metastases. The developed method could be used as a Medical Decision Support System to help clinicians in treatment response evaluation of oncological patients.

6.3 Multimodal Image Segmentation

The results reported in this section have been used to produce the following research paper:

- L. Rundo, A. Stefano, C. Militello, G. Russo, M. G. Sabini, C. D'Arrigo, F. Marletta, M. Ippolito, G. Mauri, S. Vitabile, and M. C. Gilardi. *A Fully Automatic Approach for Multimodal PET and MR Image Segmentation in Gamma Knife Treatment Planning*, submitted to Computer Methods and Programs in Biomedicine, Elsevier.

A total of 19 brain metastatic tumors originating from several primary cancers, undergone stereotactic neuro-radiosurgery, were retrospectively considered in the present study. All the patients (mean age ± standard deviation: 60 ± 9.80 years; median age: 57 years; age range: 48÷78 years) were treated with Leksell Gamma Knife® model C at Cannizzaro Hospital in Catania, Italy.

Both MRI and PET datasets were acquired for these subjects. PET and MR imaging data were obtained by two different scanners in two different times. MRI dataset, used for Gamma Knife treatment planning, was scanned a few days after MET-PET study. Representative instances of input brain MR and PET image pairs are shown in figure 25.

PET brain acquisitions without head fixation are performed using the on time of flight PET/CT Discovery 690 by General Electric Medical Systems (Milwaukee, WI, USA). Patients fast for 4 hours before PET exam and are intravenous injected with MET. The PET protocol begins 10 minutes after the injection. PET images consist into a matrix of 256x256 voxels of 1.1719x1.1719x3.27 mm3 voxel size.

All available MRI datasets are acquired on the Gyroscan Intera 1.5T MR Scanner (Philips Medical System Eindhoven, the Netherlands), before treatment, for the planning phase. MR images are "T1w FFE" (T1-weighted Fast Field Echo) contrast-enhanced sequences. MRI acquisition parameters are: TR: 25 ms, TE: 1.808-3.688 ms, matrix size: 256 × 256 pixels, slice thickness: 1.5 mm, slice sacing: 1.5 mm, pixel spacing 1.0 mm. Therefore, each size voxel is 1.0×1.0×1.50 mm3. Thanks to the Gadolinium-based contrast agent, brain lesions appear as enhanced hyper-intense zones. Sometimes a dark area might be present due to either edema or necrotic tissue.

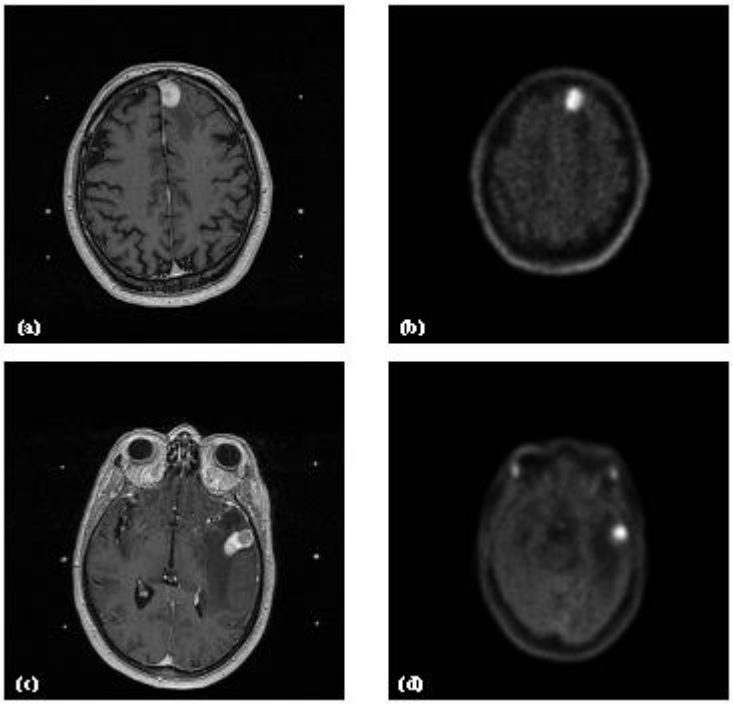

Figure 25 Examples of original input brain images concerning Gamma Knife patients: (a, c) MR slices and (b, d) the nearly respective corresponding PET slices.

6.3.1 Evaluation Metrics

Volume-based, overlap-based and spatial distance-based metrics are considered, in order to measure similarity and correlation concerning PET and MRI segmentation approaches. Assuming PET and MRI segmentations as two different processes, the evaluation metrics were computed according to the formulations presented in (Fenster & Chiu 2005). These formulations are suitably generalized to quantify the similarity and correlation between two different segmentation procedures on images defined on the same space (e.g. same image or fused images).

The aim is to demonstrate that PET and MR imaging convey different information. For this purpose, similarity and correlation measures were considered.

Let BTV and GTV be the results achieved by the PET and MRI 3D segmentation approaches, respectively.

The quantitative measure that best points out the similarity between PET and MRI volumes is the volumetric similarity (VS), namely the absolute difference between volumes divided by the volumes sum:

$$VS = 1 - \frac{\mathrm{abs}\left(|BTV| - |GTV|\right)}{|BTV| + |GTV|} = 1 - VD \tag{14}$$

where VD is the volumetric distance.

However, this volume-based metric does not take into account the intersection between PET and MRI segmentation at all. Therefore, overlap-based metrics must be considered. In particular, the DSC (see see section 6.1.3). Since medical images are usually characterized by noise and the HD is generally sensitive to outliers, a more stable distance formulation must be considered. The average distance (AVD), that is the HD averaged over all points in A and B, is defined by:

$$AVD = \max\{d(A,B), d(B,A)\} \tag{16}$$

where $d(A,B) = (1/N)\sum_{\mathbf{a}\in A} \min_{\mathbf{b}\in B} \|\mathbf{a} - \mathbf{b}\|$ is the directed average distance.

Another similarity measure is the Jaccard index (JI), which is the ratio between the cardinality of the intersection and the cardinality of the union calculated on the two segmentation results:

$$JI = \frac{|BTV \cap GTV|}{|BTV \cup GTV|} \times 100\% = \frac{DSC}{2 - DSC} \tag{15}$$

DSC is always larger than JI except at extrema $\{0,100\%\}$, where they take equal values. In both measures, higher values imply greater similarity between PET and MRI segmentations.

Lastly, in order to take into account the correlation of all samples belonging to two different points clouds, a variant of the Mahalanobis distance (MHD) is used:

$$MHD = \sqrt{(\mu_{BTV} - \mu_{GTV})^T \Sigma^{-1} (\mu_{BTV} - \mu_{GTV})} \quad (14)$$

where μ_{BTV} and μ_{GTV} are the means of the two 3D segmentations. Furthermore, Σ is the common covariance of the two sets and is given by $\Sigma = \frac{|BTV|\Sigma_{BTV} + |GTV|\Sigma_{GTV}}{|BTV| + |GTV|}$.

In the proposed joint segmentation approach, two different PET segmentation results are computed:

- BTV, by considering PET images alone (section 5.2.1), and
- BTV_{MRI}, by considering by considering PET/MRI co-segmentation (section 5.2.4.5).

Therefore, BTV and BTV_{MRI} can be used interchangeably in equations, depending on the usage context.

6.3.2 Results

Firstly, volumes measured on PET and MRI data are reported and, for a quantitative assessment, PET/MRI volume difference and centroid distance were calculated for each lesion.

Figure 26 shows segmentation results obtained by the proposed multimodal PET/MRI segmentation method on the input image pairs reported in figure 25. In addition, the corresponding GTV and BTV_{MRI} volume rendering is reported.

Other three interesting examples obtained by the proposed multimodal segmentation method can be visually and qualitatively evaluated in Figure 27. It is worth to observe that the margin of enhancement of tumors in MRI datasets are not always correlated with high uptake regions in PET images. Segmentation results of PET images combined with GTV MRI masks are very stable (magenta boundaries in Figure 27), even if GTV is quite different. In figure 27c, PVE on PET image is reduced by using MRI binary masks.

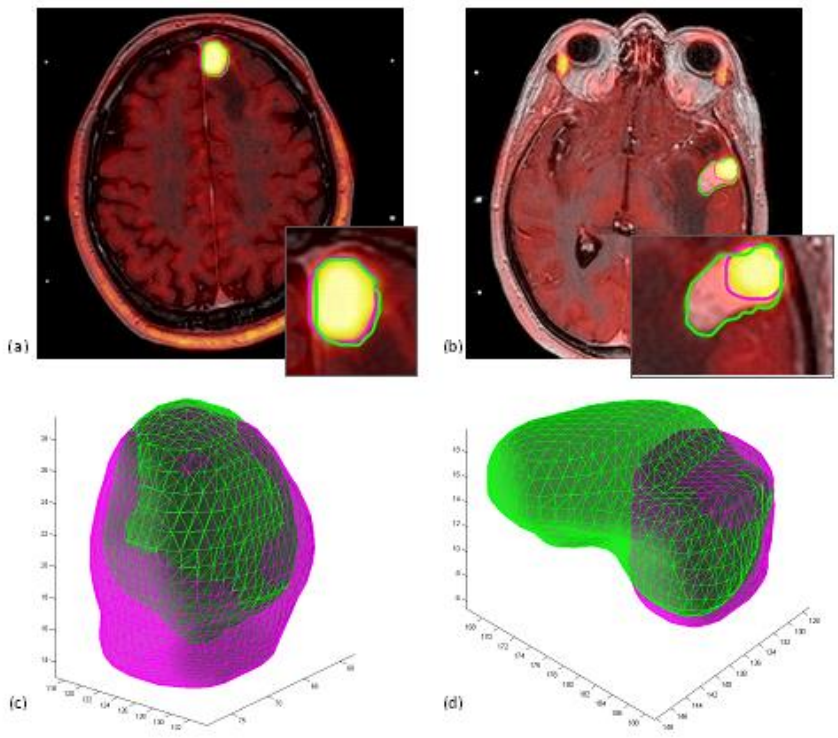

Figure 26 Results achieved by the proposed multimodal PET/MRI segmentation method on the input image pairs reported in Figure 25. The BTV_{MRI} (magenta contour) and GTV (green contour) are superimposed: (a) boundaries computed on PET and MR images are nearly overlaid; (b) the two ROIs are very different in this case. At the bottom right, a blow-up of the lesion regions is shown in both cases. 3D volume reconstruction of brain tumors in (a) and (b) are shown the lesions in (c) and (d), respectively.

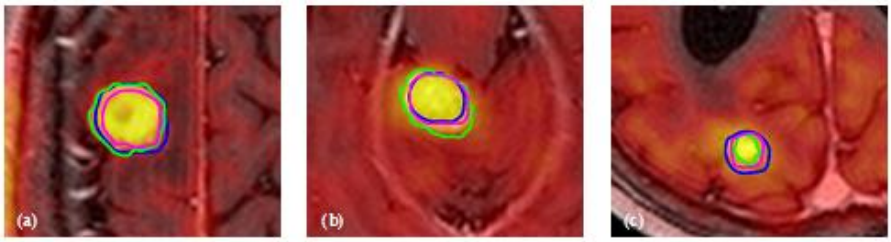

Figure 27 Graphical examples of co-segmentation results achieved by the proposed multimodal PET/MRI approach, where GTV (green contour), BTV (blue contour) and BTV_{MRI} (magenta contour) ROIs are overlaid on the corresponding fused PET/MRI slice. Three metastatic tumors in different anatomical regions of human brain are considered: (a) temporo-parietal, (b) limbic, and (c) occipital regions. The images are displayed with 2× zoom factor.

Volumes, calculated for each brain tumor imaged on MRI and PET modalities, are shown in Table V. Absolute average volume difference between BTV, BTV_{MRI} and GTV, were also calculated, in order to evaluate either overestimation or underestimation of PET functional imaging with respect anatomical MRI data, before and after GTV binary mask combination into PET images. In addition, Euclidean distance between PET and MRI volume centroids (centers of mass) was reported. Mean BTV and GTV values and their average absolute volume difference attest that tumor regions are imaged and represented differently on PET and MRI datasets. Furthermore, high standard deviation in BTV and GTV measurements proves that we are dealing with metastatic lesions characterized by various types and dimensions. On the other hand, BTV_{MRI} measurements follow a different trend because BTV_{MRI} segmentation is influenced by the combination of GTV masks and PET images.

Tumor	**GTV** [voxels]	**BTV** [voxels]	**BTV_{MRI}** [voxels]	**Absolute volume difference** [voxels]			**Volume centroid distance** [pixels]		
				BTV vs GTV	BTV_{MRI} vs GTV	BTV vs BTV_{MRI}	BTV vs GTV	BTV_{MRI} vs GTV	BTV vs BTV_{MRI}
#1	2267	3029	2380	33.61	4.98	27.27	1.805	1.067	1.457
#2	1546	1754	1297	13.45	16.11	35.24	1.523	0.740	0.839
#3	1945	1938	1561	0.36	19.74	24.15	1.250	0.949	0.398
#4	323	687	579	112.69	79.26	18.65	2.591	1.275	0.607
#5	346	484	417	39.88	20.52	16.07	1.446	0.845	0.246
#6	2293	2582	1996	12.60	12.95	29.36	0.974	0.971	0.098
#7	537	609	381	13.41	30.98	59.84	2.248	1.042	1.229
#8	1303	1010	1002	22.49	23.10	0.80	1.830	0.802	1.454
#9	286	122	159	57.34	44.41	23.27	2.238	1.681	0.399
#10	600	593	621	1.17	3.50	4.51	0.697	1.080	0.730
#11	2906	4052	2552	39.44	12.18	58.78	0.647	0.772	0.801
#12	210	226	138	7.62	34.29	63.77	3.118	2.720	0.447
#13	255	620	325	143.14	27.45	90.77	5.660	0.840	0.648
#14	2271	1202	1094	47.07	51.83	9.87	5.753	5.551	0.718
#15	1634	1129	1158	30.91	29.13	2.50	1.162	0.768	0.518
#16	2589	2491	1984	3.79	23.37	25.55	5.941	5.614	0.431
#17	1404	869	828	38.11	41.03	4.95	0.972	1.099	0.238
#18	2008	2398	1697	19.42	15.49	41.31	1.414	1.211	0.224
#19	7307	2238	1581	69.37	78.36	41.56	7.141	6.769	0.885
μ	1685.79	1475.42	1144.74	37.15	29.93	30.43	2.548	1.884	0.651
σ	1627.99	1081.92	749.63	37.59	21.37	24.22	2.018	1.890	0.395

Table V Brain tumor volume measurement for each metastatic tumor using MR and PET imaging. Total average and standard deviation calculated on 19 brain lesions are also reported.

Although we slightly integrate MRI ROI masks with PET datasets during BTV_{MRI} computation, this conservative choice frequently reduces radioactivity spill-over and spill-out effects, by exploiting higher MRI spatial resolution. This is evident in the examples shown in Figure 27. Moreover, BTV_{MRI} measurement standard deviation is lower than BTV and GTV ones. This means that multimodal PET/MRI segmentation yields more stable results with respect to single modality segmentations, since both metabolic and morphological imaging are considered. These experimental evidences are corroborated by the achieved absolute volume difference values, where BTV_{MRI} has about the same difference with GTV and BTV. Volume centroid distances reflect the aforementioned situation, too. To show the statistical distribution of GTV, BTV, and BTVMRI measurements, boxplots are reported in figure 28.

Boxplots confirm that BTV_{MRI} has a more uniform distribution than BTV on tumor volume variability range, by taking into account GTV contribution. Moreover, BTV has a positive skewed distribution characterized by outliers above the upper whisker. Accordingly, BTV_{MRI} samples are characterized by approximately the same median of BTV values, but BTV_{MRI} statistical distribution tends to GTV distribution shape. The above observations, based on numerical experimental results, are qualitatively supported by the graphical multimodal PET/MRI segmentation results in figure 29.

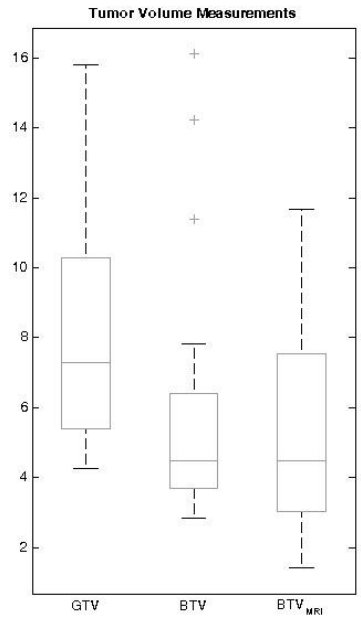

Figure 28 Boxplots of tumor volume measurements: GTV, BTV and BTV_{MRI}. Whisker value is 1.5 in all cases.

To measure the degree of dependence between GTV, BTV, and BTV_{MRI}, both Pearson's correlation coefficient and Spearman's rank correlation coefficient were calculated for each comparison (Table VI). Spearman's correlation yields different information with respect to Pearson's correlation, since Spearman's is less sensitive to strong outliers that are in the tails of both samples. In fact, outlier importance is limited to the value of its rank.

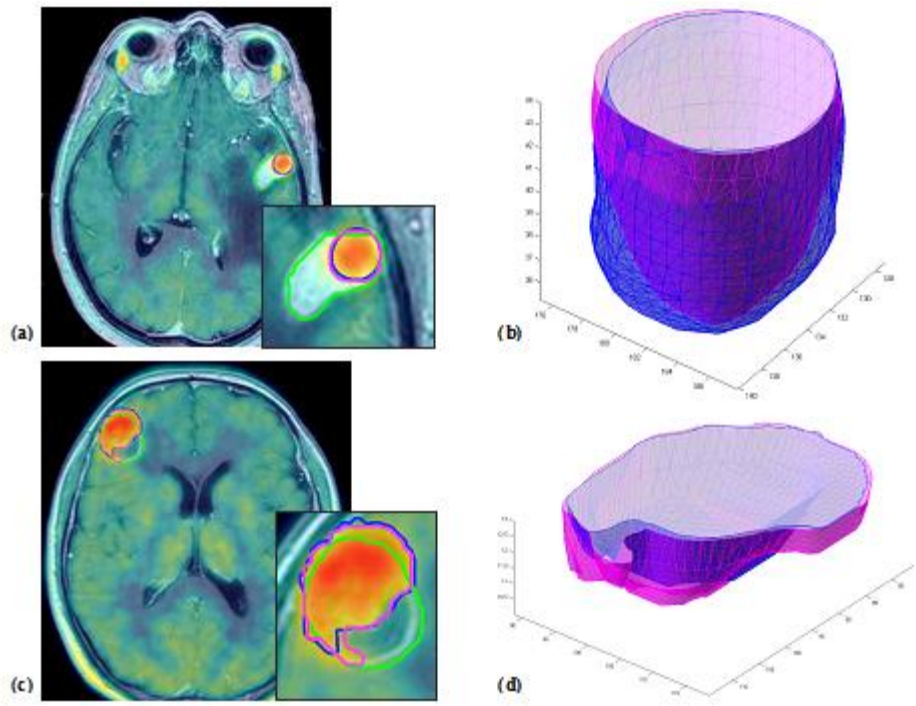

Figure 29 Two examples of segmentation results obtained by the single modality PET method (BTV) versus those obtained by the multimodal PET/MRI approach after the PET data updating based on the results of the MRI brain lesion segmentation (BTV_{MRI}). (a, c) GTV (green contour), BTV (blue contour), and BTV_{MRI} (magenta contour) ROIs overlaid on the corresponding fused PET/MRI slice; (b, d) 3D representation of BTV and BTV_{MRI} volume brain tumors.

As expected, GTV revealed greater correlation to BTV_{MRI}, especially according to Spearman. In addition, BTV and BTV_{MRI} are significantly correlated, using both Pearson and Spearman correlation coefficients. Accordingly, scatter diagrams in figure 30 represent graphically these numerical results.

The comparisons of GTV, BTV, and BTV_{MRI} for the available brain lesions (statistical samples) are also visualized using Bland-Altman plots in Figure 31. Bland-Altman plots showed the greatest differences between GTV and BTV, confirming the trend observed previously.

Table VII shows tumor volume-based and overlap-based metrics for PET and MRI segmentation methods. Both volume-based and overlap-based metrics demonstrated that GTV was, on average, more similar to BTV_{MRI} than BTV. The highest mean values and the lowest standard deviation of the evaluation measures were obtained in BTV_{MRI} and BTV comparisons. To provide a graphical representation of the statistical distribution of results, corresponding boxplots of volume-based and overlap-based evaluation metrics are also reported in Figure 32.

Table VIII reports the values of spatial distance-based metrics achieved by the proposed multimodal PET/MRI segmentation approach in the experimental trials. Distance-based metrics values agreed with achieved volume-based and overlap-based measurements, showing the same trend. To visualize dispersion and variability of achieved results, corresponding boxplots of spatial distance-based evaluation metrics are also reported in Figure 33.

Statistical comparison	**Pearson's correlation coefficient**		**Spearman's rank correlation coefficient**	
	ρ_P coefficient	Significance level p	ρ_S coefficient	Significance level p
BTV versus GTV	0.625	0.0042	0.898	< 0.0001
BTV_{MRI} versus GTV	0.627	0.0041	0.921	< 0.0001
BTV versus BTV_{MRI}	0.977	< 0.0001	0.975	< 0.0001

Table VI Correlation measure between GTV, BTV, and BTVMRI, using Pearson's correlation coefficient and Spearman's rank correlation coefficient.

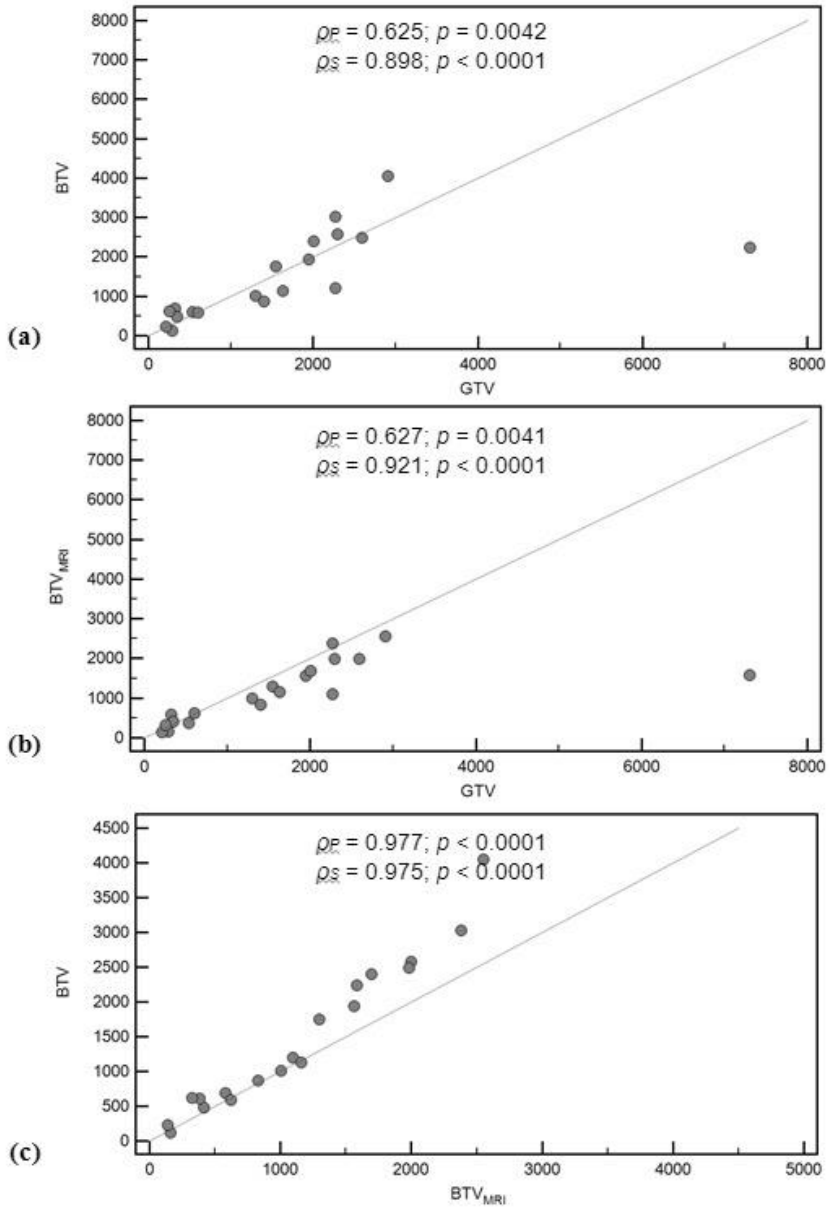

Figure 30 Scatter diagrams of tumor volume measurements: (a) BTV versus GTV; (b) BTV_{MRI} versus GTV; (c) BTV versus BTV_{MRI}. Red dashed line is the line of equality. In the upper part of each sub-figure, Pearson and Spearman correlation coefficients with the corresponding significance level are also reported.

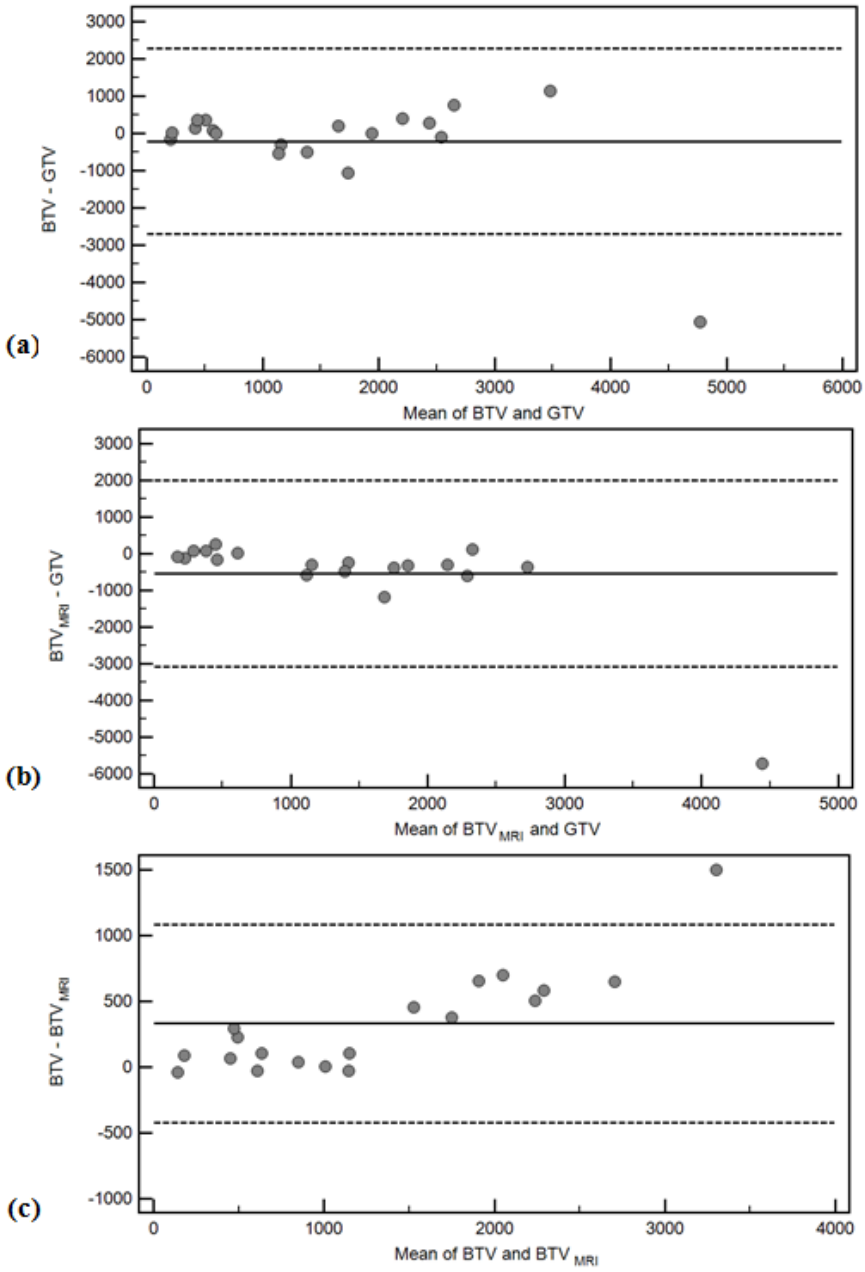

Figure 31 Bland-Altman plots of tumor volume measurements: (a) BTV versus GTV; (b) BTV_{MRI} versus GTV; (c) BTV versus BTV_{MRI}.

Tumor	Volumetric Similarity [%]			Dice Similarity Coefficient [%]			Jaccard Index [%]		
	BTV vs GTV	BTV_{MRI} vs GTV	BTV vs BTV_{MRI}	BTV vs GTV	BTV_{MRI} vs GTV	BTV vs BTV_{MRI}	BTV vs GTV	BTV_{MRI} vs GTV	BTV vs BTV_{MRI}
#1	47.03	65.14	61.51	63.97	78.89	76.17	47.03	65.14	61.51
#2	67.43	68.22	67.00	80.55	81.11	80.24	67.43	68.22	67.00
#3	68.53	66.40	72.28	81.33	79.81	83.91	68.53	66.40	72.28
#4	25.62	39.20	65.71	40.79	56.32	79.30	25.62	39.20	65.71
#5	38.33	42.09	58.35	55.42	59.24	73.70	38.33	42.09	58.35
#6	61.85	57.80	76.62	76.43	73.26	86.76	61.85	57.80	76.62
#7	41.13	64.55	45.16	58.29	78.46	62.22	41.13	64.55	45.16
#8	51.87	72.27	67.81	68.31	83.90	80.82	51.87	72.27	67.81
#9	24.39	34.44	72.39	39.22	51.24	83.99	24.39	34.44	72.39
#10	56.97	58.98	72.93	72.59	74.20	84.35	56.97	58.98	72.93
#11	48.96	64.89	51.68	65.74	78.71	68.14	48.96	64.89	51.68
#12	30.54	34.88	52.94	46.79	51.72	69.23	30.54	34.88	52.94
#13	35.66	62.46	51.93	52.57	76.90	68.36	35.66	62.46	51.93
#14	36.41	37.57	77.57	53.38	54.62	87.37	36.41	37.57	77.57
#15	62.15	65.40	81.51	76.66	79.08	89.81	62.15	65.40	81.51
#16	51.28	54.55	76.95	67.80	70.59	86.97	51.28	54.55	76.95
#17	48.76	50.40	75.13	65.55	67.03	85.80	48.76	50.40	75.13
#18	63.00	65.99	70.06	77.30	79.51	82.39	63.00	65.99	70.06
#19	19.66	18.81	64.26	32.85	31.66	78.24	19.66	18.81	64.26
μ	46.29	53.90	66.41	61.87	68.75	79.36	46.29	53.90	66.41
σ	15.02	15.02	10.34	14.64	14.07	7.79	15.02	15.02	10.34

Table VII Achieved tumor volume-based and overlap-based metrics for GTV, BTV, and BTV_{MRI} segmentations. Total average and standard deviation calculated on 19 brain lesions are also reported.

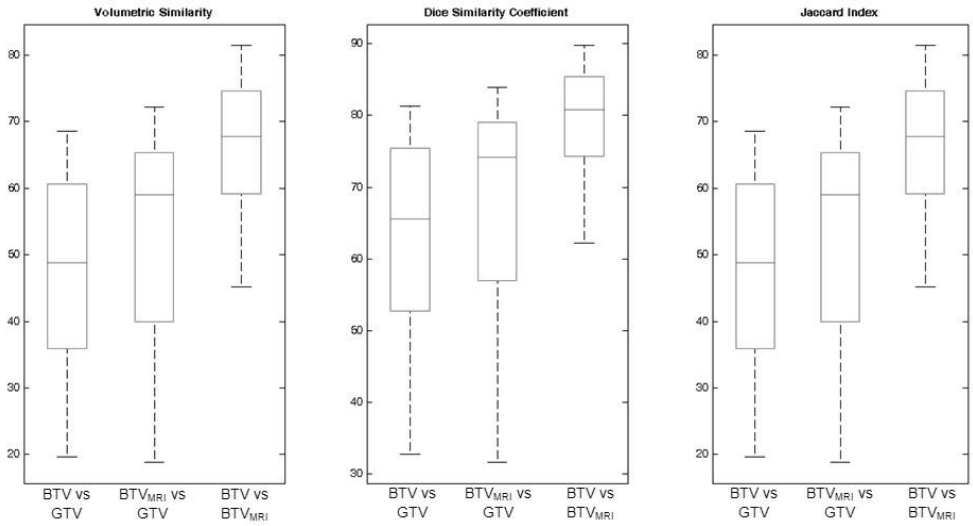

Figure 32 Boxplots of tumor volume-based and overlap-based metrics calculated for GTV, BTV and BTV_{MRI} segmentations: **(a)** Volumetric similarity; **(b)** Dice similarity coefficient; **(c)** Jaccard index.

Tumor	Hausdorff Distance [pixels]			Average Distance [pixels]			Mahalanobis Distance [pixels]		
	BTV vs GTV	BTV_{MRI} vs GTV	BTV vs BTV_{MRI}	BTV vs GTV	BTV_{MRI} vs GTV	BTV vs BTV_{MRI}	BTV vs GTV	BTV_{MRI} vs GTV	BTV vs BTV_{MRI}
#1	8.544	6.403	7.616	2.669	1.377	1.756	0.412	0.264	0.339
#2	5.831	4.000	5.831	1.343	1.163	1.259	0.340	0.178	0.195
#3	8.485	7.810	4.472	1.400	1.355	1.145	0.436	0.312	0.138
#4	5.831	3.606	5.000	1.090	0.667	0.778	3.497	3.257	0.277
#5	5.385	3.000	3.606	1.295	0.798	0.742	1.857	1.814	0.107
#6	5.099	4.123	3.000	1.121	1.243	0.940	1.208	1.247	0.029
#7	9.220	4.000	5.831	2.233	0.804	1.332	0.786	0.441	0.421
#8	10.630	3.162	10.630	1.312	0.943	0.815	0.610	0.239	0.524
#9	4.472	4.472	1.414	0.727	0.969	0.204	1.392	0.994	0.377
#10	7.810	6.403	3.162	1.368	1.369	0.689	0.285	0.379	0.286
#11	12.000	6.083	11.662	2.630	1.535	2.262	0.734	0.180	0.214
#12	4.472	5.388	2.236	1.898	1.556	1.167	1.496	1.389	0.249
#13	7.280	2.828	7.280	2.345	0.809	1.711	0.530	0.404	0.296
#14	15.621	14.213	3.606	3.618	3.281	0.810	1.115	1.045	0.233
#15	4.243	3.606	2.236	1.250	1.107	0.643	0.286	0.219	0.147
#16	11.705	11.402	7.616	3.503	3.175	1.906	0.788	0.754	0.094
#17	5.385	4.243	2.236	2.136	1.939	0.769	0.267	0.319	0.087
#18	7.071	5.385	4.472	1.782	1.420	1.356	0.374	0.336	0.062
#19	15.811	16.125	10.000	5.820	5.811	1.486	1.446	1.360	0.195
μ	8.15	6.12	5.36	2.081	1.648	1.146	0.940	0.796	0.225
σ	3.57	3.79	3.04	1.211	1.235	0.515	0.787	0.778	0.131

Table VIII Achieved tumor spatial distance-based metrics for GTV, BTV, and BTV_{MRI} segmentations. Total average and standard deviation calculated on 19 brain lesions are also reported.

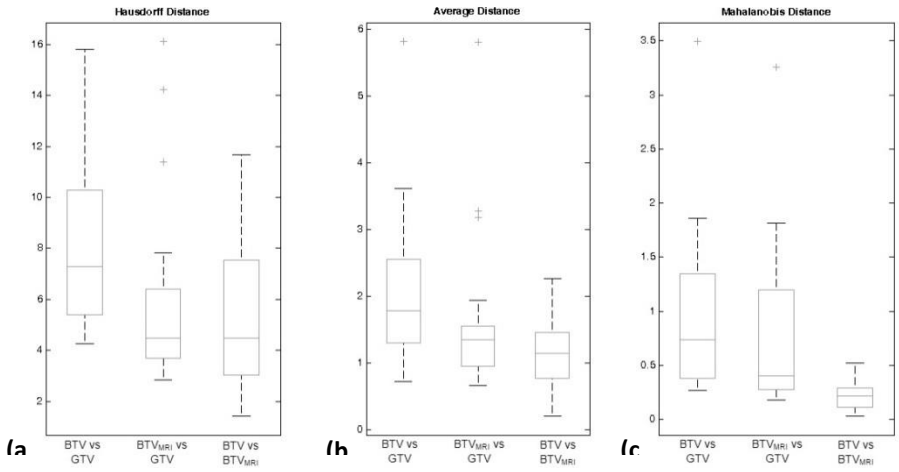

Figure 33 Boxplots of tumor spatial distance-based metrics calculated for GTV, BTV and BTVMRI segmentations: (a) Hausdorff Distance; (b) Average distance; (c) Mahalanobis distance.

6.3.3 Discussion

The proposed fully automated multimodal segmentation method introduces several novelties and advantages. Firstly, two computer-assisted single modality segmentation approaches for PET and MR images are used. These single modality segmentation methods are certainly more effective and reliable on PET and MRI datasets, with respect to approaches that unify information into a single graph. In these cases, no significant anatomical and functional changes have to be assumed between the scans to directly construct the hyper-graph (Bagci et al. 2013).

Secondly, IURs detected on PET images are not blindly propagated to MRI segmentation algorithm, but properly modified according to MRI data using the developed ROI bounding region generation method. This ad hoc MRI-driven method, based on LSFs, finds adaptively a valid bounding region in order to eliminate user intervention, required by the MRI brain tumor segmentation method in (Militello et al. 2015). Hence, two existing and efficient single modality processing pipelines are combined in a smart fashion. In particular, PET and MRI segmentation results are mutually exploited each other.

Lastly, this fully automated approach is very robust and reliable, also due to the brain anatomical district imaging. We focused on the application domain regarding brain tumors that underwent Leksell Gamma Knife stereotactic radiosurgery. In fact, the fully automatic identification of target IURs is feasible because MET-PET datasets for stereotactic neuro-radiosurgery treatment planning include only the brain area avoiding the possible presence of false positives in other anatomic regions. Differently, in total body FDG-PET examinations user interaction to manually identify the target lesion is always needed (normal structures such as brain, heart, bladder, kidneys, and ureters normally have high FDG uptake). Accordingly, these co-segmentation software implementations platform must provide manual seeding facilities in addition to automated multimodal segmentation algorithm.

Although an one-to-one relationship between two different structural images of the same abnormal region could be reasonable, the assumption of identical lesion contours in both functional and morphologic images is often infeasible in many anatomical regions (Bagci et al. 2013). For instance, depending on the metabolic characteristics of the cells within a lung tumor, it may not take up a radiotracer in all its volume, or uptake region may be larger than the anatomical boundary of the tumor due to cellular activation in nearby tissues. Further issues are introduced by combining more than two structural modalities, such as MRI and CT, with metabolic PET imaging.

In this study, PET and MRI datasets regarding the same subject were acquired at two different times (MRI is scanned a few days after MET-PET) by two different dedicated PET/CT and MRI scanners, respectively. This represents a non trivial problem in co-segmentation domain, because an efficient and accurate multimodal co-registration is required. Instead, state of the art about co-segmentation methods process mostly simultaneous PET/CT or PET/MRI acquisitions. Of course, the potential of hybrid imaging (PET/CT and PET/MRI systems) overcomes image registration issues by giving multimodal images contextually.

Achieved experimental results showed that GTV and BTV segmentations are statistically correlated but have not higher degree of similarity. In fact, GTV and BTV measurements as well as evaluation metrics values (volume-based, overlap-based, and spatial distance-based metrics) corroborate that MRI and MET-PET convey different but complementary imaging information. The output of the proposed multimodal PET/MRI segmentation method combines PET and MR imaging and is similar to both BTV and GTV. Moreover, BTV_{MRI} segmentation reduces radioactive spill-over and spill-out effects between tumor and surrounding tissues, taking advantage of the higher spatial resolution of MRI.

Finally, PET imaging should be usually consider by clinicians in order to determine a CTV that takes into account the “active” part of the cancer in addition to anatomical tumor boundaries. According to these findings, it is appropriate to include PET images in stereotactic neuro-radiosurgery treatment planning.

Chapter 7

Discussions and Conclusions

PET segmentation in radiotherapy is a critical task due to the lack of consistency in tumor contour, low image resolution, and relatively high level of noise and heterogeneity of FDG uptake within a lesion. Nevertheless, accurate lesion segmentation in PET imaging is essential for an accurate quantification of prognosis assessment and therapy response. Qualitative visual interpretation is the most commonly used method. The manual segmentation method is dependent on the experience of the nuclear physician limiting the measurement accuracy. In addition, it is time-consuming and impaired by inter- and intra-observer variability that is due to subjectivity and sensitivity to the display window level settings. To reduce these issues, several automatic methods have been presented but few clinical studies are available and there is no consensus for proper BTV determination.

To date, CT and MR imaging are considered to be the standard for target volume delineation in radiotherapy. On the other hand, CT and MR imaging does not show the biological features of tumors. For this reason, PET has been introduced in the radiotherapy field to assist radiation oncologists in clinical routine. Despite the fact that many studies use the information of co-registered PET and CT images to identify features for distinguishing a lesion from the background and, consequently, for PET image segmentation, I believe that BTV extraction is independent from anatomical segmentation being inappropriate to consider a matching between anatomical and metabolic regions, with special reference in HNC district. I believe that a correct delineation of BTV must be obtained without incorporating anatomical information or incorporating them with great attention. For example, MRI is crucial for the illustration of the anatomy of brain tumors: for this reason, a study to determine simultaneously tumor contours on PET and RM images is a topic of this thesis.

In addition, to identify oncological lesions in whole-body PET scans, a completely automatic detection method cannot be implemented since healthy organs such as brain, heart, bladder, and kidneys normally have a high FDG uptake, and, consequently, user interaction is mandatory. Vice versa, this is feasible in brain MET-PET studies.

In this thesis, the semi-automatic segmentation algorithm based on the RW formulation of Grady (L. Grady 2006) has been enhanced.

In HNC studies, the key strategies include k-means clustering to obtain refined target seed locations within pre-segmented lesions and a strategy to adaptively determine the appropriate threshold to be applied on the probabilistic output of RW in order to obtain the final tumor segmentation. The RW algorithm with a k-means algorithm embedded to identify centroids of hot regions is proposed to obtain an appropriate segmentation in complex lesions. In addition, the user can manually change the probability, rather the fixed one of 50% of the original RW method, to discriminate target from background voxels in order to select the value that optimizes the segmentation. We call this method "K-RW". Second, an extension of the previous method to adaptively determine the probability threshold to discriminate lesion from background voxels is implemented. For this reason, we call this method "Adaptive K-RW (AK-RW)". These methods segment PET images from SUV and they are very fast in a clinical environment if compared with the time needed for manual segmentation. Naturally, the AK-RW method is slightly less accurate than the supervised K-RW method, but it is expected because the selection of probability threshold is done automatically. AK-RW obtains good segmentation results besides having the benefit of requiring a low user interaction effort and low levels of user's specialist knowledge being independent of the choice of the probability threshold value to discriminate target from background regions. The development of user independent techniques able to perform good segmentation is crucial in a clinical environment.

The accuracy of the proposed methods is optimal in phantom studies: high DSC and TPVF values and low HD and FPVF values confirm the robustness and the accuracy of the two methods. A DSC rate greater than 90% is almost always observed in the larger spheres. The clinical results underline that the proposed methods are superior to other methods minimizing the difference between manual and automatic lesion segmentations and, especially, also delineating complex volumes unlike other state-of-the-art methods. The methods commonly used in the literature have shown an acceptable delineation under specific conditions especially when the uptake concentration is homogeneous. In many patient studies, the lesions have complex shapes, such as bifurcated lesions, and nonhomogeneous uptake. In these cases, these methods fail in lesion delineation. The proposed methods take into account these issues providing a better segmentation according to the accuracy required in radiotherapy field to prevent potential disease progression.

A drop in the accuracy can occur for small lesions; this is compatible with the large errors in the volume estimation reported for small tumor volume. The PVE for

smaller objects is one of the most important factors impacting the quality and the quantitative accuracy in PET imaging (Soret et al. 2007). The images are blurred due to the limited spatial resolution of PET scanners and small lesions appear larger. For this reason, a PVE correction method has been developed (Stefano et al. 2014; Gallivanone et al. 2011).

The aim of the brain cancer study is to make the proposed segmentation method fully automatic and user independent identifying automatically the foreground and background seeds. The fully automatic identification of RW seeds is feasible because the MET PET datasets for Gamma-Knife treatments include only the brain area avoiding the possible presence of false positives in other anatomic regions. In addition, the adaptive probability threshold for each patient ranges from 25.28% up to 70.50% with a mean of 41.49±10.68% indicating how this factor is different both from patient to patient and from the fixed value of 50% of the original RW method.

This result highlights the importance of adopting the proposed enhanced version of the existing algorithm of Grady (Grady, 2006) to adaptively determine the appropriate threshold to be applied on the probabilistic output of RW method.

The developed methods have been used to assess the metabolic response in three different clinical studies. The early stratification of responder and non-responder patients is mandatory to avoid non-curative treatments preserving the patient's quality of life. PET imaging is a useful technique to discriminate between patients who respond from those who do not respond to treatment: in particular, TLG has the advantage that it reflects both the metabolic information (SUV) and the lesion size (BTV).

At last, a fully automated multimodal segmentation method for Gamma Knife treatment is proposed. In this way, the BTV can be used to modify the GTV in order to treat the brain lesions more precisely combining the complementary information of tissues from both anatomical and functional domains: MRI GTV is utilized to combine the superior contrast of PET images with the higher spatial resolution of MR images. The proposed multimodal method exploits two PET and MRI segmentation methods, which have been properly combined together and improved. In particular, BTV and GTV segmentation results are mutually exploited each other in order to influence and refine CTV delineation.

In conclusion, the proposed methods are very powerful in terms of PET image and PET/MRI image segmentation, and may be used daily as a Medical Decision Support Systems to enhance the current methodology performed by healthcare operators in radiotherapy treatments.

References

Ambrosini, R.D., Wang, P. & O'Dell, W.G., 2010. Computer-aided detection of metastatic brain tumors using automated three-dimensional template matching. *Journal of magnetic resonance imaging : JMRI*, 31, pp.85–93.

Bagci, U. et al., 2011. A Graph-Theoretic Approach for Segmentation of PET Images. *2011 Annual International Conference of the Ieee Engineering in Medicine and Biology Society (Embc)*, 2011, pp.8479–8482. Available at: <Go to ISI>://MEDLINE:22256316.

Bagci, U. et al., 2013. Joint segmentation of anatomical and functional images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images. *Medical Image Analysis*, 17(8), pp.929–945.

Bezdek, J.C., Ehrlich, R. & Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3), pp.191–203. Available at: http://www.sciencedirect.com/science/article/pii/0098300484900207.

Boykov, Y., Veksler, O. & Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11), pp.1222–1239.

Ciernik, I.F. et al., 2003. Radiation treatment planning with an integrated positron emission and computer tomography (PET/CT): a feasibility study. *International Journal of Radiation Oncology*Biology*Physics*, 57(3), pp.853–863.

Collignon, A. et al., 1995. Automated multi-modality image registration based on information theory. *Information processing in medical imaging*, 3(6), pp.263–274.

Drever, L. et al., 2006. A local contrast based approach to threshold segmentation for PET target volume delineation. *Medical Physics*, 33(6), p.1583.

Drever, L. et al., 2007. Iterative threshold segmentation for PET target volume delineation. *Med Phys*, 34(4), pp.1253–1265. Available at: http://link.aip.org/link/?MPH/34/1253/1.

Eisenhauer, E.A. et al., 2009. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer*, 45(2), pp.228–247.

Fenster, A. & Chiu, B., 2005. Evaluation of Segmentation algorithms for Medical Imaging. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society*. pp. 7186–9.

Ford, E.C. et al., 2006. Tumor delineation using PET in head and neck cancers: threshold contouring and lesion volumes. *Med Phys*, 33(11), pp.4280–4288.

Friston, K.J., 2006. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*,

Gallivanone, F. et al., 2011. PVE correction in PET-CT whole-body oncological studies from PVE-affected images images. *IEEE Transactions on Nuclear*

Science, 58(3 PART 1), pp.736–747.
Geets, X. et al., 2007. A gradient-based method for segmenting FDG-PET images: methodology and validation. *European Journal of Nuclear Medicine and Molecular Imaging*, 34, pp.1427–1438.
Grosu, A.L. et al., 2005. Reirradiation of recurrent high-grade gliomas using amino acid PET (SPECT)/CT/MRI image fusion to determine gross tumor volume for stereotactic fractionated radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 63(2), pp.511–519.
Guido, A. et al., 2009. Combined 18F-FDG-PET/CT Imaging in Radiotherapy Target Delineation for Head-and-Neck Cancer. *International Journal of Radiation Oncology Biology Physics*, 73(3), pp.759–763.
Han, D. et al., 2011. Globally optimal tumor segmentation in PET-CT images: A graph-based co-segmentation method. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp. 245–256.
Hatt, M. et al., 2009. A Fuzzy Locally Adaptive Bayesian Segmentation Approach for Volume Determination in PET. *Ieee Transactions on Medical Imaging*, 28, pp.881–893.
Ippolito, M. et al., 2015. Evaluation of Platinum-based Therapy Response in Non-Small Cell Lung Cancer, Eur J Nucl Med Mol Imaging (2015) 42: S714-S715.
Ippolito, M. et al., 2015. Using anatomic and metabolic imaging in Gamma Knife treatments, accepted to World Molecular Imaging Congress (WMIC) 2015.
Kao, C.-H. et al., 2010. F-18-FDG PET/CT-based gross tumor volume definition for radiotherapy in head and neck Cancer: a correlation study between suitable uptake value threshold and tumor parameters. *Radiation Oncology*, 5.
L. Grady, 2006. Random Walks for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11), pp.1768–1783.
Larson, S.M. et al., 1999. Tumor treatment response based on visual and quantitative changes in global tumor glycolysis using PET-FDG imaging. The visual response score and the change in total lesion glycolysis. *Clinical Positron Imaging (Netherlands)*, 2(3), pp.159–171.
Lauve, A. et al., 2004. *Simultaneous integrated boost intensity-modulated radiotherapy for locally advanced head-and-neck squamous cell carcinomas: II--clinical results.*,
Leibfarth, S. et al., 2015. Automatic delineation of tumor volumes by co-segmentation of combined PET/MR data. *Physics in Medicine and Biology*, 60(14), pp.5399–5412.
Leksell, L., 1983. Occasional review Stereotactic radiosurgery. *Journal of neurology, Neurosurgery, and Psychiatry*, 46(April), pp.797–803.
Levivier, M. et al., 2002. The integration of metabolic imaging in stereotactic procedures including radiosurgery: a review. *Journal of neurosurgery*, 97(5 Suppl), pp.542–550.
Li, C. et al., 2010. Distance regularized level set evolution and its application to

image segmentation. *IEEE Transactions on Image Processing*, 19(12), pp.3243–3254.
Li, H. et al., 2008. A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours. *Medical Physics*, 35, pp.3711–3721.
Maes, F. et al., 1997. Multimodality image registration by maximization of mutual information. *IEEE transactions on medical imaging*, 16(2), pp.187–198.
Militello, C. et al., 2015. Gamma Knife treatment planning: MR brain tumor segmentation and volume measurement based on unsupervised Fuzzy C-Means clustering. *International Journal of Imaging Systems and Technology*, 25(3), pp.213–225.
El Naqa, I. et al., 2007. Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning. *Medical physics*, 34(2007), pp.4738–4749.
Nariai, T. et al., 2005. *Usefulness of L-[methyl-11C] methionine-positron emission tomography as a biological monitoring tool in the treatment of glioma.*,
Onoma, D.P. et al., 2014. Segmentation of heterogeneous or small FDG PET positive tissue based on a 3D-locally adaptive random walk algorithm. *Computerized Medical Imaging and Graphics*, 38(8), pp.753–763.
Paulino, A.C. et al., 2005. Comparison of CT- and FDG-PET-defined gross tumor volume in intensity-modulated radiotherapy for head-and-neck cancer. *International Journal of Radiation Oncology Biology Physics*, 61, pp.1385–1392.
Pisciotta, P. et al., 2016. Using anatomic and metabolic imaging in sterotactic radio neuro-surgery treatments, accepted to IX Congresso Nazionale AIFM (Associazione Italiana Fisica Medica), Perugia 2016.
Pluim, J.P.W., Maintz, J.B.A. & Viergever, M.A., 2003. Mutual-information-based registration of medical images: a survey. *IEEE transactions on medical imaging*, 22(8), pp.986–1004.
Potesil, V., Huang, X. & Zhou, X.S., 2007. <title>Automated tumor delineation using joint PET/CT information</title>. In *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*. p. 65142Y–65142Y–8.
Rundo, L. et al., 2016. A Fully Automatic Approach for Multimodal PET and MR Image Segmentation in Gamma Knife Treatment Planning, submitted to Computer Methods and Programs in Biomedicine, Elsevier.
Schinagl, D.A.X. et al., 2007. Comparison of five segmentation tools for 18 F-FLUORO-DEOXYGLUCOSE-POSITRON emission tomography-based target volume definition in head and neck cancer. *International Journal of Radiation Oncology Biology Physics*, 69, pp.1282–1289.
Schwartz, D.L. et al., 2005. FDG-PET/CT imaging for preradiotherapy staging of head-and-neck squamous cell carcinoma. *International journal of radiation oncology, biology, physics*, 61(1), pp.129–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15629603.
Song, Q. et al., 2013. Optimal Co-segmentation of tumor in PET-CT images with

context information. *IEEE Transactions on Medical Imaging*, 32(9), pp.1685–1697.
Soret, M., Bacharach, S.L. & Buvat, I.I., 2007. Partial-volume effect in PET tumor imaging. *Journal of Nuclear Medicine*, 48(6), pp.932–945. Available at: <Go to ISI>://WOS:000247054800018.
Spieth, M.E. & Kasner, D.L., 2002. A Tabulated Summary of the FDG PET Literature. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 43(3), pp.439–441.
Stefano, A. et al., 2014. Metabolic Impact of Partial Volume Correction of [18F] FDG PET-CT oncological studies on the assessment of tumor response to treatment. QJ Nucl Med Mol Imag (2014) 58(4):413-23.
Stefano, A. et al., 2013. A graph-based method for biological target volume segmentation. Oral Communication at the VIII Congresso Nazionale AIFM (Associazione Italiana Fisica Medica), Torino 2013.
Stefano, A. et al., 2013. A Graph-Based Method for PET Image Segmentation in Radiotherapy Planning: A Pilot Study. A. Petrosino (Ed.): ICIAP 2013, Part II, LNCS 8157, pp. 711–720, 2013.
Stefano, A. et al., 2016. Head and Neck Cancer Tumor Volume Segmentation in PET images Using an Enhanced Random Walk Algorithm, Under Review at Medical & Biological Engineering & Computing.
Stefano, A. et al., 2016. Biological Target Volume Segmentation for Radiotherapy Treatment Planning, accepted to IX Congresso Nazionale AIFM (Associazione Italiana Fisica Medica), Perugia 2016.
Stefano, A. et al., 2014. Evaluation of erlotinib treatment response in non-small lung cancer using metabolic and anatomic criteria. QJ Nucl Med Mol Imag, EPUB ahead of print.
Stefano, A. et al., 2015. Metabolic Response Assessment in Non-Small Cell Lung Cancer Patients after Platinum-Based Therapy: A Preliminary Analysis. Current Medical Imaging Reviews, vol. 11, n. 4, pp. 218-227, 2015.
Stefano, A. et al., 2015. An Automatic Method for Metabolic Evaluation of Gamma Knife Treatments, A. Petrosino (Ed.): ICIAP 2015, Part II, LNCS 9279, pp. 579–589, 2015.
Stefano, A. et al., 2015. Biological Target Volume Segmentation for Gamma Knife Treatment, accepted to International Conference on Clinical PET-CT and Molecular Imaging (IPET) 2015.
Stefano, A. et al., 2015. An operator independent method for lesion segmentation to evaluate metabolic response in MET-PET studies, accepted to World Molecular Imaging Congress (WMIC) 2015.
Stefano, A. et al., 2015. An Automatic Method for PET Delineation of Cerebral Tumors, Eur J Nucl Med Mol Imaging (2015) 42: S710-S711.
Stefano, A. et al., 2016, A Fully Automatic Method for Biological Target Volume Segmentation of Brain Metastases, International Journal of Imaging Systems and Technology, in press.

Studholme, C., Hill, D.L.G. & Hawkes, D.J., 1999. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1), pp.71–86.
Udupa, J.K. et al., 2006. A framework for evaluating image segmentation algorithms. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 30(2), pp.75–87.
Wahl, R.L. et al., 2009. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 50 Suppl 1, p.122S–50S.
Wanet, M. et al., 2011. Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: A comparison with threshold-based approaches, CT and surgical specimens. *Radiotherapy and Oncology*, 98(1), pp.117–125.
Xia, Y. et al., 2012. Dual-modality brain PET-CT image segmentation based on adaptive use of functional and anatomical information. *Computerized Medical Imaging and Graphics*, 36(1), pp.47–53.
Yezzi, 2001. A variational framework for joint segmentation and registration. *Mathematical Methods in Biomedical Image Analysis, 2001. MMBIA 2001. IEEE Workshop on*, pp.44–51.
Young, H. et al., 1999. Measurement of clinical and subclinical tumour response using [18F]- fluorodeoxyglucose and positron emission tomography: Review and 1999 EORTC recommendations. *European Journal of Cancer*, 35(13), pp.1773–1782.
Yu, H. et al., 2009. Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning. *IEEE transactions on medical imaging*, 28(3), pp.374–83.
Zaidi, H. et al., 2002. Fuzzy clustering-based segmented attenuation correction in whole-body PET imaging. *Physics in medicine and biology*, 47(7), pp.1143–1160.
Zaidi, H. & El Naqa, I., 2010. PET-guided delineation of radiation therapy treatment volumes: A survey of image segmentation techniques. *European Journal of Nuclear Medicine and Molecular Imaging*, 37(11), pp.2165–2187.