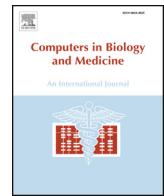




ELSEVIER

Contents lists available at ScienceDirect

## Computers in Biology and Medicine

journal homepage: [www.elsevier.com/locate/combiomed](http://www.elsevier.com/locate/combiomed)

# A smart and operator independent system to delineate tumours in Positron Emission Tomography scans

Albert Comelli<sup>a,b,c</sup>, Alessandro Stefano<sup>a,\*</sup>, Giorgio Russo<sup>a,d</sup>, Maria Gabriella Sabini<sup>d</sup>,  
Massimo Ippolito<sup>e</sup>, Samuel Bignardi<sup>c</sup>, Giovanni Petrucci<sup>b</sup>, Anthony Yezzi<sup>c</sup>

<sup>a</sup> Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Cefalù, PA, Italy

<sup>b</sup> Department of Industrial and Digital Innovation (DIID) – University of Palermo, PA, Italy

<sup>c</sup> Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332, USA

<sup>d</sup> Medical Physics Unit, Cannizzaro Hospital, Catania, Italy

<sup>e</sup> Nuclear Medicine Department, Cannizzaro Hospital, Catania, Italy

## ARTICLE INFO

## Keywords:

Cancer segmentation

Active contour algorithm

18F-fluoro-2-deoxy-*d*-glucose and 11C-labeled

methionine PET imaging

Biological target volume

## ABSTRACT

Positron Emission Tomography (PET) imaging has an enormous potential to improve radiation therapy treatment planning offering complementary functional information with respect to other anatomical imaging approaches.

The aim of this study is to develop an operator independent, reliable, and clinically feasible system for biological tumour volume delineation from PET images. Under this design hypothesis, we combine several known approaches in an original way to deploy a system with a high level of automation.

The proposed system automatically identifies the optimal region of interest around the tumour and performs a slice-by-slice marching local active contour segmentation. It automatically stops when a “cancer-free” slice is identified. User intervention is limited at drawing an initial rough contour around the cancer region. By design, the algorithm performs the segmentation minimizing any dependence from the initial input, so that the final result is extremely repeatable.

To assess the performances under different conditions, our system is evaluated on a dataset comprising five synthetic experiments and fifty oncological lesions located in different anatomical regions (i.e. lung, head and neck, and brain) using PET studies with 18F-fluoro-2-deoxy-*d*-glucose and 11C-labeled Methionine radio-tracers.

Results on synthetic lesions demonstrate enhanced performances when compared against the most common PET segmentation methods. In clinical cases, the proposed system produces accurate segmentations (average dice similarity coefficient:  $85.36 \pm 2.94\%$ ,  $85.98 \pm 3.40\%$ ,  $88.02 \pm 2.75\%$  in the lung, head and neck, and brain district, respectively) with high agreement with the gold standard (determination coefficient  $R^2 = 0.98$ ). We believe that the proposed system could be efficiently used in the everyday clinical routine as a medical decision tool, and to provide the clinicians with additional information, derived from PET, which can be of use in radiation therapy, treatment, and planning.

## 1. Introduction

Positron Emission Tomography (PET) is a non-invasive medical imaging technique which has the advantage over other anatomical imaging techniques, such as Computerized Tomography (CT) and Magnetic Resonance (MR) of providing direct information about patient's functional processes. Metabolic indicators, and in general parameters derived from PET imaging, might be predictive of patient therapy response to the pharmacological treatment of cancer [1] and are useful in obtaining an objective evaluation of the changes in the

patient condition [2,3]. As a matter of fact, metabolic parameters are often faster changing and more indicative of therapy effects than morphological changes [2]. PET imaging possesses an enormous potential to improve clinical cancer treatment decision making [4]. For this reason, PET imaging is being increasingly considered for the quantitative assessment of individual response to therapy and for clinical testing of novel cancer therapy protocols. In this context, the first parameter historically introduced, the Standardized Uptake Value (SUV) provides punctual information about the investigated tissue and has gained a central role in PET studies [5]. Unfortunately, SUV alone

\* Corresponding author.

E-mail address: [alessandro.stefano@ibfm.cnr.it](mailto:alessandro.stefano@ibfm.cnr.it) (A. Stefano).

<https://doi.org/10.1016/j.combiomed.2018.09.002>

Received 17 April 2018; Received in revised form 20 August 2018; Accepted 6 September 2018

0010-4825/© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

does not provide any information about lesion volume and consequently additional quantitative indices have been introduced, such as the biological tumour volume (BTV) and the tumour lesion glycolysis (TLG) [6,7]. TLG, in particular, is obtained as the product of the SUV with the BTV and provides both volumetric and metabolic information. Therefore, for efficient identification of the TGL, and to obtain accurate and reproducible PET parameters, it is of utmost importance to employ a reliable BTV segmentation strategy.

In addition, quantitative analysis of cancer tissues in PET is a crucial step towards precise radiation therapy treatment planning. To date, MR plays a key role in radiation therapy planning, providing several advantages over CT including high quality detailed images, and excellent soft-tissue contrast. Conversely, CT attenuation maps convey a fundamental source of geometric information which is not available in MR, where image distortion could be produced by static magnetic field non-uniformities [8]. Inclusion of PET in radiotherapy protocols can provide additional information to target the oncological lesion even more efficiently [9]. Accurate BTV definition is essential for escalating the radiation dose without increasing normal tissue injury, and may be used for enhanced dose delivery to the target, in turn improving the Radiation Treatment Planning (RTP) [10]. Nevertheless, the BTV segmentation in PET images is a precarious task both because of their typically low resolution and the relatively high level of noise [11,12]. Further, results strongly depend on the algorithm used [13]. The choice of a standard method for BTV contouring is still a challenging and debated issue, so that manual contouring is still typically adopted in clinical practice. Of course, manual contouring depends on the operator expertise and clinical specialization. It is also extremely time-expensive because dozens of slices must be inspected. To date, several automatic or semi-automatic PET segmentation methods have been proposed [13–15]. A brief overview of the state-of-the-art is reported in Section 1.1. In general, the level of performance imposed by the everyday clinical routine, makes properties such as repeatability (i.e. the result should be operator independent), and a real-time data processing workflow not only desirable, but necessary.

In the proposed study, we focus on the uptake of 18F-fluoro-2-deoxy-*D*-glucose (FDG) and 11C-labeled Methionine (MET) to establish an innovative segmentation system and assess its performances under different conditions in the clinical environment. The use of FDG for oncology imaging accounts for approximately 90% of all PET imaging procedures. When compared with its neighbouring normal tissue, a pathological mass presents an increased adsorption of FDG (which is an analogue of glucose). By this mechanism, a new characterization of the oncological diseases is possible, thereby opening new opportunities for a patient-customized approach to diagnosis and therapy. However, FDG is not an efficient radio-tracer when the body district of interest presents highly active metabolic activity and a different radio-tracer must be used. A typical example is the brain, where high uptake of glucose is a normal condition. It has been reported that the extent of tumour cell invasion in brain metastases can be detected by MET PET even more clearly than by CT or MRI [16,17]. For this reason, we considered MET as an alternative radio-tracer.

The present study tackles the volume reconstruction challenge using SUV. We combined several known approaches in an original way to devise a system with a high level of automation. Reconstruction is started by creating an optimal initial mask on an automatically identified slice. The mask is evolved into an optimal contour which is then propagated to the neighbouring layers using a slice-by-slice marching approach. Finally, volume reconstruction is automatically stopped when a suitable stopping condition is met. This latter feature represents a key point toward the automatic and operator independent BTV segmentation. To assess the performance of the present system and to investigate the reliability and repeatability of the results, we performed comparative tests with other methods. To do so, we used both body phantoms containing objects of known a-priori volume and shape, and

fifty tumours located at various anatomical districts (i.e. lung, head and neck, and brain) and different PET radio-tracers.

The article is organized as follows: Section 2 describes the proposed segmentation system and the framework used to assess the system performance. The dataset used for the system evaluation is described in Section 3, test results are shown in Section 4, while discussion and conclusions are provided in Sections 5 and 6, respectively.

### 1.1. Background

A huge number of PET segmentation methods are present in the literature. Among others, thresholding [18,19] and region growing (RG) [20] methods are the most widely adopted, especially because they are simple to implement. Unfortunately, they show a drop in performance when low contrast, and heterogeneous cancer regions are considered [14,21]. Indeed, segmentations of small or non-spherical tumours are often below the expectations [22]. The new adaptive RG algorithm [23] repeatedly applies a confidence connected RG algorithm with an increasing relaxing factor  $f$ . A maximum curvature strategy is used to automatically identify the optimal  $f$ -value. This algorithm, in the case of relatively homogeneous background, results robust to parameter settings and region of interest selection, without scanner, imaging protocol, or tumour dependencies.

Variational approaches based on gradient differences between healthy and cancer tissues are mathematically efficient but sensitive to noise and subject to numerical fluctuation [24,25]. Learning methods, such as artificial neural networks [26], and support vector machines [27] are efficient, but the training of such algorithms usually requires large and diversified databases. Fuzzy C-Means (FCM) [28] is extensively used in PET image delineation for the fuzzy nature of the lesion contours. The FCM reveals accurate for large targets of simple shape, while lesions of complex shape are not easily managed [29]. Affinity propagation [30] considers multi-focal radiotracer uptake patterns, but so far it has been proved a viable solution only in animal studies. Stochastic models, such as Gaussian mixture model, based on statistical differences in intensity distribution between foreground and background can be considered optimal for noisy images, provided that a proper noise model is used [31–33]. The latter approaches have been tested on simulated studies or on a few patient studies where the ground truth is defined manually by a nuclear medicine expert. In active contour (AC) algorithms, an initial contour around the target deforms and moves towards the target edges. This deformation is handled by minimizing what is termed as the energy function. Li et al. [34] used RG as a pre-processing step to optimize the AC's initial contour. Unfortunately, the result tended to overestimate the tumour volume [23]. Similarly, ACs can be found combined with anisotropic diffusion filtering, followed by a multi-resolution contourlet transform [35]. However, the latter approach main limitation is a heavy dependence on user-defined parameters. In order to improve delineation accuracy, histogram FCM clustering and textural information were used to constrain the AC [36]. The method however proved to suffer in presence of nearby high physiologic uptake and was susceptible to initial cropping area. Graph-based approaches yield efficient segmentation by using foreground and background seeds to locate different tissues [37]. However, while seed identification can be automated (e.g. Ref. [29]), normal anatomical structures (i.e. brain, heart, bladder, and kidneys) are prone to be mistakenly identified as initial target seeds, giving misleading guidance to the segmentation process [38,39].

Alternative studies exist which tackle the challenge of automatically discriminating between normal and pathological tissues in PET. Unfortunately, a full comparison of such algorithms is not possible as the relative studies often concern different body district and specific types of abnormality, e.g., lung tumours [40–42], oesophageal tumours [43], and nasopharyngeal tumours [44]. Studies on discrimination of pathological structures in whole-body PET have been conducted as

well, and some preliminary results have shown that different anatomical areas pose different challenges [45,46]. In addition, numerous PET-based radiomics studies have been proposed and the results of the relative analysis are highly dependent on the method used to derive the BTV [47,48].

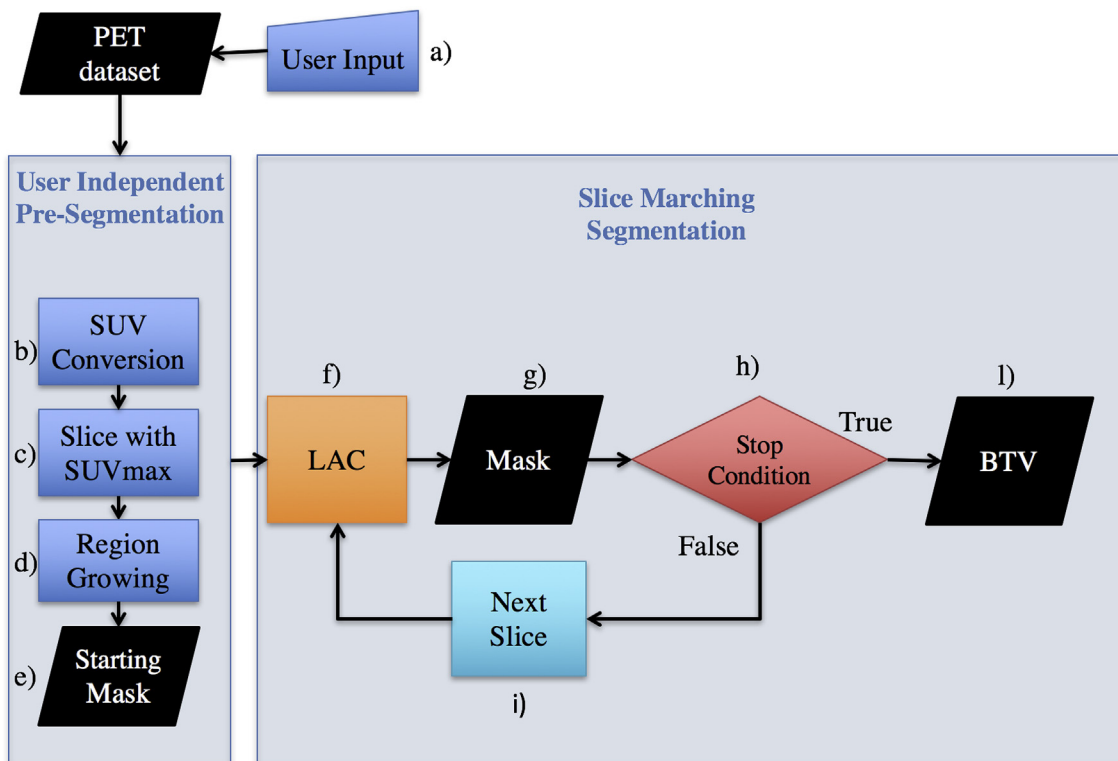
The combination of PET and CT or MRI in a single scanner is desirable, as it offers the advantage of generating complementary information. However, a one-to-one relationship between anatomical and functional images is not always appropriate [38]. When PET/CT or PET/MRI imaging is employed, BTV delineation should be obtained avoiding the use of the anatomical information provided by CT and MRI measurements or alternatively, the latter should be used very carefully [49]. For example, Yezzi et al. [50] proposed ACs to simultaneously segment and register features from multimodal images. Rundo et al. [9] combined two computer-assisted and operator-independent single modality methods, previously developed and validated obtaining a fully automatic multimodal PET/MRI segmentation method. Co-segmentation methods which assume no significant anatomical and functional changes between the images acquired using different modalities, generally yield a single target volume where the information of each modality is not optimally exploited, or at worse, lost.

## 2. Overview of the proposed system

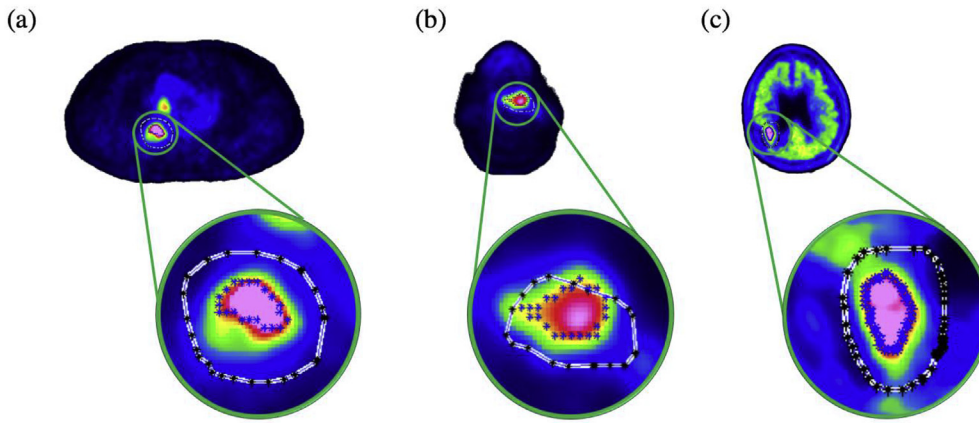
In order to overcome the general limitations of the above mentioned studies, we combined and adapted to PET imaging several existing technologies. Doing so, we created a new smart system (Fig. 1) whose main purpose is to deploy a segmentation strategy to be used in radiotherapy treatment planning, and outcome evaluation. The proposed

system is applicable to various types of lesion, different anatomical districts and can perform segmentation on PET studies based on different radiotracers. Our approach reduces at minimum the intra-observer and inter-observer dependencies with respect to manual delineation, without introducing any user-dependent parameter. The AC segmentation algorithm at the core of the system is applied using a slice-by-slice approach, starting from an initial, automatically individuated slice. Process is terminated when a suitable stopping condition is met. Consequently, the system presents a high level of automation. In the following we provide a high level description of our approach.

Briefly, the region containing the cancer must be localized by the operator to avoid healthy tissues having high radio-tracer uptake or critical conditions, i.e. when the lesion is located near the heart. The PET Digital Imaging and Communications in Medicine (DICOM) dataset is then converted into SUV images to normalize the voxel activity, taking into account functional aspects of the disease. Based on this minimal input, where the user is required to highlight a region on just a single PET slice, the algorithm performs all subsequent operations automatically. The first step consists of a pre-segmentation which computes a user independent region of interest (ROI). The algorithm automatically finds an optimal starting mask, which may lay on a slice different from the one initially highlighted by the operator. Once the independent ROI has been obtained, the relative initial mask is feed to the next step of the system, where the segmentation is performed using a Local region-based Active Contour (LAC) segmentation algorithm, appropriately modified to support metabolic images as explained in Section 2.2. The obtained LAC segmentation is propagated to the adjacent slices using a slice-by-slice marching approach. Propagation is



**Fig. 1.** The proposed segmentation system. a) to avoid any false positive, the region containing the lesion is localized by the operator; b) PET dataset is converted into SUV images to incorporate functional information; c) The PET slice containing maximum Standardized Uptake Value is found d) RG is used to identify the user independent ROI on the slice; e) initial ROI is sent to the next logical block of the system; f) segmentation is performed using the LAC algorithm; g) segmentation mask is propagated to the adjacent slices (slice-by-slice marching approach); h) the stop condition is evaluated (see Section 2.2); i) segmentation on the next slices is performed until stop condition is false; l) an operator independent BTV is finally obtained.



**Fig. 2.** Operator independent ROI generation on PET images. Figures a, b, and c show the pre-segmentation output (Fig. 1e) on three different body districts: lung (a), brain (b) and head and neck (c), respectively. The white contour with black asterisks represent the input provided by the user, while the blue asterisks contour shows the result of the RG algorithm after the slice containing the maximum SUV is identified. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

performed in parallel both upward and downward within the SUV volume to obtain the BTV, until a suitable stopping condition is met. In the following, a detailed discussion of the various steps is presented.

### 2.1. User independent pre-segmentation step

We defined the proposed system as semi-automatic because in total body FDG-PET examinations, an initial user input is still required to roughly identify the area containing the cancer. The motivation is to avoid false positives when cancer is situated near anatomical structures such as brain, heart, bladder, kidneys, and ureters where high FDG uptake is to be considered normal. However, user intervention is minimal and limited at drawing a rough contour around the cancer area, on just one PET slice (Fig. 1-a). No further intervention is needed. The input method (e.g. touch screen, stylus, etc.) depend on user's preference. In the present study input was performed by mouse.

Pre-processing the uptake data in PET examinations is mandatory for inter-patient and follow-up comparisons. Among PET metrics, SUV is the most widely used quantification parameter giving a measure of cellular metabolism. For this reason, we integrated SUV information in our system. The PET dataset is converted into weighted SUV unit (g/ml) [51] images (Fig. 1-b), as suggested in Ref. [38]. SUV is calculated as the ratio of tissue radioactivity concentration (RC) in kBq/ml and injected dose (ID) in MBq at the time of injection divided by body weight ( $M_p$ ) in kilograms, thereby taking into account essential functional aspects of the patient:

$$SUV_i = \frac{RC}{ID} * M_p \quad (1)$$

RC is calculated as the ratio between the image intensity and the image scale factor. ID is the product between actual activity and dose calibration factor [5,52].

Based on the region highlighted by the user ( $ROI_0$ ), the first logical block of our algorithm consists of an automatic pre-segmentation step which computes a user independent ROI. This goal is achieved by first identifying a volume of interest (VOI) which is formed by propagating  $ROI_0$  to adjacent slices.

Every time a new slice (index  $j$ ) is considered,  $ROI_0$  is propagated to obtain  $ROI_j$ , and the maximum SUV within  $ROI_j$  is detected ( $SUV_{max}^j$ ). If this value exceeds 42% [53] of  $SUV_{max}^0$  found within the initial slice, then the new slice is added to the volume. A new  $ROI_j$  is then computed in the slice  $j$  using the region growing (RG) method [20] to follow the evolution of the lesion along the SUV volume. In particular, the  $SUV_{max}^j$  voxel is used as target seed. Otherwise, the growing is stopped. In this way, a cylindroid (i.e. a cylinder with a non-circular base) is obtained.

The next pre-segmentation step is to locate the voxel of maximum SUV within this volume of interest,  $SUV_{max}^{VOI}$  (Fig. 1-c). Note that this maximum value may reside on a different slice than the one selected by the user, but it will belong to the same anatomical anomaly.

The rough segmentation of the newly identified slice using the RG algorithm (Fig. 1-d) represents the output of the pre-segmentation step (Fig. 1-e) which will feed the next logical block of the system (Fig. 1-f).

Note that in this pre-segmentation process, the region highlighted by the user does not need to be carefully drawn. A rough inclusion of the anomalous area is sufficient. Further, manual drawing does not need to be repeated on any other slice. As a result, the pre-segmentation process generates an output which is independent to local variations in the initial user input, and for that, thereby making the result extremely repeatable. Furthermore, the RG algorithm is only used to obtain a rough estimate of the contour encircling the highest radio-tracer uptake area within the slice containing the  $SUV_{max}^{VOI}$ . This contour is then feed into the next section of the system, the ‘‘Slice Marching Segmentation’’ as shown in Fig. 1, where the delineation is further refined through a more sophisticated segmentation algorithm (Fig. 1f), accompanied by an automated data-driven stopping condition (Fig. 1h).

Consequently, we can expect the final BTV (Fig 1i) to contain the cancer within its limits and the segmentation to account for extremely irregular lesion shapes. Three examples of ROI bounding regions delineated on PET images by the RG method are shown in Fig. 2 (a–c).

### 2.2. Slice Marching Segmentation step

The LAC algorithm blends benefits of purely local edge based active contours and fully global region based active contours. At each point along a prominent intensity edge of an object, nearby points inside and outside the object will be modelled well by the mean intensities within the local neighbourhoods on either side of the edge. This is the motivation behind the model proposed by Lankton et al. [54]; where the contour energy to be minimized is defined as:

$$E = \oint_C \left( \int_{R_{in}} \chi_l(x, s) (I(x) - u_l(s))^2 dx + \int_{R_{out}} \chi_l(x, s) (I(x) - v_l(s))^2 dx \right) ds \quad (2)$$

where  $R_{in}$  and  $R_{out}$  represent the regions inside and outside the curve  $C$ ; where  $s$  represents the arc-length parameter of  $C$ , where  $\chi$  represents the characteristic function of the ball of radius  $l$  (local neighbourhood) centred around a given curve point  $C(s)$ , where  $I$  represents the intensity function of the image to be segmented, and where  $u_l(s)$  and  $v_l(s)$  denote the local mean image intensities within the portions of the local neighbourhood  $\chi_l(x, s)$  inside and outside the curve respectively (within  $R_{in}$  and  $R_{out}$ ). These neighbourhoods are defined by the function  $\chi$ , the radius parameter  $l$ , and the position of the curve  $C$ . Note that the function  $\chi_l(x, s)$  evaluates to 1 in a local neighbourhood around each contour point  $C(s)$  and 0 elsewhere, thereby localizing the processing of image information. The shape of the contour  $C$  then divides each such local region into interior local points and exterior local points in accordance with the contour's role to segment the domain of  $I$ . The

resulting flow is more robust to initial curve placement and image noise like region-based flows described on [36]; and yet it also capable of capturing significant local structure and partitioning the image without making strong global assumptions about its makeup. To be successful, the present segmentation technique relies on the assumption that nearby points inside and outside the true edge of an object, will be modelled well by the mean intensities within these localized regions. As a result, the energy (2) is more global in nature than edge-based flows but with a “tunable” degree of locality defined by the neighbourhood radius  $l$ .

Actually, the LAC method for 3D MR datasets was applied by Lankton et al. [54] via independent segmentation of the 2D slices. A more powerful and coherent segmentation procedure could be performed on all slices simultaneously by evolving a single surface within the corresponding three-dimensional space. While we are currently investigating such 3D shape evolution for future evaluation and publication, the present work moves a step toward 3D data segmentation improving upon Lankton et al. by introducing a system to automatically govern the slice-by-slice strategy. Further improvements consist of introducing the SUV measurements (functional information), an optimal identification of the starting slice, the production of an initial mask for LAC segmentation, and a fully automatic stopping condition. The latter, in particular is a key feature in order to achieve automation.

We now describe these new improvements integrated into the LAC algorithm.

In step f) (see Fig. 1), an initial mask is obtained from user independent smart pre-segmentation (seen Section 2.1). To incorporate SUV in the LAC algorithm, the energy (2) for the PET image segmentation approach is adapted as:

$$E = \oint_C \left( \int_{R_{in}} \chi_l(x, s) (SUV(x) - u_l(s))^2 dx + \int_{R_{out}} \chi_l(x, s) (SUV(x) - v_l(s))^2 dx \right) ds \quad (3)$$

where  $SUV$  represents the intensity function of the standardized uptake value to be segmented, and where  $u_l(s)$  and  $v_l(s)$  denote the local mean  $SUV$  intensities within the portions of the local neighbourhood  $\chi_l(x, s)$  inside and outside the curve respectively (within  $R_{in}$  and  $R_{out}$ ). These neighbourhoods are defined by the function  $\chi$ , the radius parameter  $l$  (in our study  $l = 3$  has been determined to provide the best performance using trial and error strategy, see appendix), and the position of the curve  $C$ . Note that the function  $\chi_l(x, s)$  evaluates to 1 in a local neighbourhood around each contour point  $C(s)$  and 0 elsewhere, thereby localizing the processing of SUV image information. The shape of the contour  $C$  then divides each such local region into interior local points and exterior local points in accordance with the contour's segmentation of the  $SUV$ . The local means are specified in terms  $S_{I_l}(s)$ ,  $S_{E_l}(s)$ ,  $A_{I_l}(s)$ , and  $A_{E_l}(s)$  which represent the local sums of  $SUV$  intensities and the areas of their respective portions of the local neighbourhood  $\chi_l(x, s)$  inside and outside the curve (within  $R_{in}$  and  $R_{out}$ ). More precisely, the local interior region may be expressed as  $R_{in} \cap \chi_l(x, s)$  and local exterior region as  $R_{out} \cap \chi_l(x, s)$ .

$$u_l(s) = \frac{S_{I_l}(s)}{A_{I_l}(s)} \quad , \quad v_l(s) = \frac{S_{E_l}(s)}{A_{E_l}(s)} \quad (4)$$

$$S_{I_l}(s) = \int_{R_{in}} \chi_l(x, s) SUV(x) dx \quad , \quad S_{E_l}(s) = \int_{R_{out}} \chi_l(x, s) SUV(x) dx \quad (5)$$

$$A_{I_l}(p) = \int_{R_{in}} \chi_l(x, s) dx \quad , \quad A_{E_l}(p) = \int_{R_{out}} \chi_l(x, s) dx \quad (6)$$

$$\chi_l(x, s) = \begin{cases} 1 & \text{when } x \in l - \text{Ball}(C(s)); \\ 0 & \text{otherwise;} \end{cases} \quad (7)$$

After the segmentation step f) is achieved on the slice containing the  $SUV_{max}^{VOI}$ , the resulting segmented mask is used to initiate parallel segmentations on the neighbouring slice above and below. Subsequently, for all the other slices in both directions, we similarly use the segmentation results of the previous slices as the initial mask inputs as shown in step g) Fig. 1. The LAC method is inherently capable of locally widening or tightening where necessary when the contour is propagated from slice to slice. Since, this behaviour is driven by the image properties rather than by an inherent knowledge of whether the cancer is present, a stopping criterion is necessary to prevent the LAC algorithm from misbehaving or even diverging when it reaches a slice where the cancer is absent (i.e. when there is nothing to be segmented).

Therefore, we devised a fully automatic stopping condition as shown in step h) Fig. 1. For the slice under consideration, at each point on the cancer edge, nearby points inside and outside the cancer must have a different local mean SUV. If the cancer is present, a positive difference between background and foreground intensity must occur, and consequently the algorithm can safely proceed with the next neighbouring slice (Fig. 1-i). When the system encounters a slice where the local mean  $v_l(s)$  on  $R_{out}$  is greater or equal to the local mean SUV  $u_l(s)$  on  $R_{in}$ , which is the opposite of what is expected, the slice is classified as cancer-free and the slice-to-slice segmentation propagation is halted (Fig. 1-l) in that direction. The approach can be mathematically summarized as:

$$\text{Stop Condition} = \begin{cases} \text{segmentation process is stopped} & \text{when } v_l(s) \\ \geq u_l(s) \text{ is true;} \\ \text{segmentation process is continues} & \\ \text{otherwise;} \end{cases} \quad (8)$$

In this way, slice after slice, the BTV is obtained and the segmentation process is automatically stopped, thereby avoiding the need for any user intervention. In the following section, the criteria to judge the quality of the segmentation results is outlined.

### 2.3. Framework for performance evaluation

A framework for the evaluation of the proposed segmentation system is presented. Overlap-based and spatial distance-based metrics, according to the formulations presented in Refs. [55–57], are considered to determine the accuracy achieved by the proposed computer-assisted segmentation system (assessed volume)  $s_{Seg}$  against the gold-standard (reference volume)  $s_{ref}$  (in our study, manual segmentations performed by three experts are used to define a consolidated reference as described in the next Section 3.4).

The overlap and difference between the two contours were measured according to true positive (TP), false positive (FP), true negative (TN) and false negative (FN) voxels. In particular, we calculated the mean, standard deviation, 95% confidence interval (CI) and coefficient of variation (CV) of sensitivity, positive predictive value (PPV), dice similarity coefficient (DSC), Hausdorff distance (HD), and Mahalanobis distance (MHD).

Sensitivity, also called the true positive volume fraction (TPVF), is the fraction of the total amount of tissue in the proposed segmentation system  $s_{Seg}$  which overlaps with the reference volume  $s_{ref}$  [55]. A perfect segmentation algorithm would be 100% sensitive (segmenting all voxels from the target voxels) and 100% specific (not segmenting any from the background voxels). In particular, TPVF [58] is a crucial measure in radiotherapy in order to obtain optimal RTP avoiding cancer recurrence.

Specificity, also called the true negative volume fraction (TNVF), is the fraction of tissue defined in the reference volume  $s_{ref}$  that is missed by the segmentation system  $s_{seg}$ . Considering that the number of true negatives depends on the space volume  $I$ , the specificity makes little sense and only the sensitivity conveys useful information. The specificity can be replaced with the positive predictive value (PPV) [57].

PPV, also called precision, is the fraction of the total amount of tissue in the reference volume  $s_{ref}$  which overlaps with the segmentation system  $s_{seg}$ .

DSC value [55] measures the spatial overlap between the reference volume  $s_{ref}$  and the segmentation system  $s_{seg}$ : a DSC value equal to 100% indicates a perfect match between two volumetric segmentations, while a DSC whose value is 0% indicates no overlap:

$$DSC = \frac{2|s_{seg} \cap s_{ref}|}{|s_{seg}| + |s_{ref}|} = \frac{2 \times TP}{2 \times TP + FP + FN} \times 100\% \quad (9)$$

Nevertheless, overlap-based metrics are highly dependent on the segmentation size. For this reason, distance-based metrics are highly recommended when the boundary segmentation is critical, such as in BTV delineation for RTP. In particular, HD is used to measure the most mismatched boundary voxels between automatic and manual BTV: a small median of HD means an accurate segmentation, while a large median of HD means no accuracy.

The HD [56,59] between two finite point sets A and B is defined by:

$$HD = \max\{h(A, B), h(B, A)\} \quad (10)$$

The two distances  $h(A, B)$  and  $h(B, A)$  are termed as forward and backward Hausdorff distances of A to B.  $h(A, B)$  identifies the point  $a \in A$  that is farthest from any point  $b \in B$  and it measures the distance from  $a$  to its nearest neighbour in B.  $h(B, A)$  identifies the point  $b \in B$  that is farthest from any point  $a \in A$  and it measures the distance from  $b$  to its nearest neighbour in A.

Each point of A (or B) must be within distance  $h(A, B)$  of some point of B, and there is at least one of A that is exactly distance  $h(A, B)$  from the nearest point of B considered as the most mismatched point. Symmetric consideration for  $h(B, A)$ , with respect to B instead of A.

$h(a, B)$  is defined as  $\min_{b \in B} h(a, b)$  where  $h(a, b)$  is the Euclidean distance, and, similarly,  $h(b, A)$  is defined as  $\min_{a \in A} h(b, a)$ . Consequently,  $h(A, B)$  is defined as the  $\max_{a \in A} h(a, B)$ , and  $h(B, A)$  is defined as the  $\max_{b \in B} h(b, A)$ . Then,

$$h(A, B) = \max_{a \in A} \left\{ \min_{b \in B} \sqrt{\sum_{k=1}^n (a_k - b_k)^2} \right\} \quad (11)$$

where the Euclidean distance is defined on n-dimensional Euclidean space. Similarly for  $h(B, A)$ . Thus, HD measures the degree of mismatch between A and B by measuring the distance of the point of A that is farthest from any point of B, and vice versa.

Since medical images are usually characterized by noise and the HD is generally sensitive to outliers, the HD averaged over all points in A and B has been considered [56]:

$$HD = \max\{d(A, B), d(B, A)\} \quad (12)$$

where  $d(A, B) = (1/N) \sum_{a \in A} \min_{b \in B} \|a - b\|$ .

A variant of the MHD has been considered to take into account the correlation of all points belonging to two different points clouds [56]. The MHD between the sets of voxels contained in  $s_{seg}$  e  $s_{ref}$  is given by:

$$MHD = \sqrt{(\mu_{Seg} - \mu_{Ref})^T S^{-1} (\mu_{Seg} - \mu_{Ref})} \quad (13)$$

where  $\mu_{seg}$  and  $\mu_{ref}$  are the means of the two segmentations. S is the common covariance of the two sets and is given by

$$S = \frac{n_1 S_{Seg} + n_2 S_{Ref}}{n_1 + n_2} \quad (14)$$

where  $S_{seg}$ ,  $S_{ref}$  are the covariance matrices of the voxel sets and  $n_1$ ,  $n_2$  are the numbers of voxels in each set. In phantom experiments, shape

modifications and volume translations are unlikely to be observed. In these studies, sensitivity, PPV, DSC and HD measurements can be considered more than sufficient for performance assessment, as reported in Ref. [55]. For this reason, MHD has been considered in only patient studies.

Finally, to perform statistical test between the proposed system and human segmentations, the combination of sensitivity and PPV, as recommended in Ref. [15], has been used. Further, to facilitate the evaluation and ranking of the results, three new accuracy scores have been defined:

- score =  $0.5 \times \text{sensitivity} + 0.5 \times \text{PPV}$ ;
- score radiotherapy planning (RT) =  $0.6 \times \text{sensitivity} + 0.4 \times \text{PPV}$ ;
- score follow-up (FU) =  $0.4 \times \text{sensitivity} + 0.6 \times \text{PPV}$ .

In radiotherapy planning, the aim is to reduce the risk of missing the target, even if it means delivering higher dose to the surrounding healthy tissues and organs-at-risk. For this reason, sensitivity could be considered more important than PPV. For therapy follow-up the aim is to obtain consistent volume measurements in sequential PET scans and to avoid including background/nearby tissues. As a result, PPV could be considered more important than sensitivity.

The performance of the proposed system is compared to other state of the art PET image segmentation methods. In particular, the fixed thresholding (42% of the  $SUV_{max}$ ) method (T42%) [53], the RG method [20], the FCM clustering method [28], the enhanced RW method such as described in Ref. [38], and the original LAC method [54] have been used for comparison. The software package used to provide manual gold-standard and proposed segmentation BTVs and evaluation task has been implemented in the Matlab R2016a simulation environment (MathWorks, Natick, MA, USA), running on an iMac computer with a 3.5 GHz Intel Core i7 processor, 16 GB 1600 MHz DDR3 memory, and OS X El Capitan.

### 3. Data

#### 3.1. Phantom studies

National Electrical Manufacturers Association International Electrotechnical Commission (NEMA IEC) phantom is used for preliminary performance testing [18,29,38]. The phantom is composed of an elliptical cylinder ( $D1 = 24$  cm,  $D2 = 30$  cm,  $h = 21$  cm) with six different spherical lesions of size 10, 13, 17, 22, 28, and 37 mm in diameter placed at 5.5 cm from the centre of the phantom. Both body phantom and spheres are filled with a known amount of radioactive tracer to simulate oncological lesions. It could be argued that cancer is often inhomogeneous and irregularly shaped in contrast with spherical targets and that tracer-filled spheres suffer from cold-body effects when compared to the real case (the human body) [15,60]. However, the aim of this test is to evaluate the efficiency of the system and fairly compare its performance with other methods under identical conditions where results may be quantitatively and reliably evaluated. For this purpose, using a well-defined sharp-edged target is preferable because the only uncertainty on the exact boundary location is introduced by the point spread function of the measuring system. The ratio between sphere concentrations and background radioactivity concentrations ranged from 1.5:1 to 8:1 for five independent experiments: 1.5:1 for the phantom ‘‘I’’, 3:1 for the phantom ‘‘II’’, 5:1 for the phantom ‘‘III’’, 7:1 for the phantom ‘‘IV’’, and 8:1 for the phantom ‘‘V’’. Performance results are calculated considering small spheres (sphere diameter less than or equal to 17 mm) and large spheres (sphere diameter greater than 17 mm). This choice was motivated by the fact that large biases are introduced [61] when the lesion size is smaller than 2–3 times the Full Width at Half Maximum (FWHM) of the PET point spread function. The PET/CT acquisition protocol is described in Section 3.3.

### 3.2. Clinical studies

In the present study, we retrospectively considered 50 cases acquired at the Nuclear Medicine Department of the Cannizzaro Hospital (Catania, Italy). The dataset contains 50 PET examinations of patients with lung, head & neck, and brain cancers that had been referred for a diagnostic PET/CT scan before radiotherapy treatments. Segmentations were performed off-line and the results had no influence on the treatment protocol, nor on the patient management. No sensitive patient information was accessed. As such, the institutional Hospital Medical Ethics Review Board approved this study protocol and all subjects involved were properly informed and released their written consent. In FDG PET studies, patients fasted for 12 h before the PET examination, and successively were intravenously injected with FDG. The PET/CT oncological protocol began 60 min after the injection. Patients breathed normally during the PET and CT examinations, and scanning was executed from the top of the skull to the middle of the thigh with the arms along the body. In MET PET studies, for only brain acquisitions, patients fasted for 4 h before the PET examination. The PET/CT protocol began 10 min after the MET injection.

### 3.3. PET/CT acquisition protocol

The acquisitions for both phantom experiments and clinical studies were performed within the same Nuclear Medicine Department using the same equipment, a Discovery 690 PET/CT scanner with time of flight (General Electric Medical Systems, Milwaukee, WI, USA). The phantom and patient protocols included a SCOUT scan at 40 mA, a CT scan at 140 keV and 150 mA (10 s), and 3D PET scans (2.5 min per bed position). PET images were reconstructed using a 3D ordered subset expectation maximization (OSEM) algorithm. All imaging data were encoded in the 16-bit DICOM format.

Each PET slice consists of  $256 \times 256$  voxels with a grid spacing of  $2.73 \text{ mm}^3$  and thickness of  $3.27 \text{ mm}^3$ . Consequently, the size of each voxel is  $2.73 \times 2.73 \times 3.27 \text{ mm}^3$ . Thanks to the injected PET radiotracer (FDG or MET), tumours appears as hyper-intense regions.

The non-diagnostic CT scan is performed for attenuation correction and anatomic localization of the tumour contextually to PET image acquisition. The CT slice consists of  $512 \times 512$  voxels with size  $1.36 \times 1.36 \times 3.75 \text{ mm}^3$ .

### 3.4. Gold standard

Phantoms offer the advantage of known target boundaries. Consequently, we used the match between segmented CT images and the known position of the spherical shapes to quantitatively evaluate the performance of the proposed segmentation system.

In patient studies, knowledge of the ground truth would require exact knowledge of the pathology region. In PET imaging, histopathology analysis provides the only valid ground truth for quantitative assessment [57]. Since in radiotherapy the histopathology analysis is unavailable, the actual gold-standard is impossible to retrieve. To assess the clinical effectiveness and feasibility of a proposed segmentation method it is therefore standard practice to refer to manual delineations performed by expert professionals as a substitute for ground truth [15]. The BTV delineation is actually a critical task performed by expert physicians to determine which areas to include or exclude in the planned target volume. Yet, manual delineation carries a certain amount of subjectivity and is often influenced by the clinical specialization of the operator. For example, radio-therapists will, on average, draw larger boundaries than oncologists. For this reason, we used as gold-standard the segmentations performed by three experts with high clinical and PET imaging insight and different expertise (the chief nuclear medicine physician –M.I. author-, the chief radiotherapy physician –M.S. author- and an expert radiotherapy physician –G.R. author-).

**Table 1**

Sensitivity, PPV, DSC, HD rates obtained over NEMA IEC phantoms (II–V). Phantoms have the following signal contrast: (II) 3:1, (III) 5:1, (IV) 7:1, and (V) 8:1. Three different spherical lesions of size 10, 13, and 17 mm in diameter are considered to assess the performance of the proposed segmentation system. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

NEMA IEC Phantom	Sphere Diameter	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]
II	10 mm	66.71%	76.22%	71.15%	1.00
III		67.62%	76.71%	71.96%	1.47
IV		75.81%	73.50%	74.63%	1.00
V		71.10%	82.16%	76.22%	1.11
II	13 mm	66.73%	92.31%	77.40%	1.00
III		75.04%	81.81%	78.38%	1.08
IV		66.01%	97.17%	78.60%	1.00
V		70.03%	91.30%	79.20%	1.00
II	17 mm	66.00%	97.16%	78.61%	1.21
III		69.95%	96.33%	81.06%	1.28
IV		73.43%	92.00%	81.60%	1.20
V		71.61%	94.07%	81.35%	1.07
<i>Mean</i>		70.00%	87.56%	77.51%	1.12
<i>± std</i>		± 3.49%	± 8.87%	± 3.46%	± 0.15
<i>± CI</i>		± 1.97%	± 5.02%	± 1.95%	± 0.08
<i>± CV</i>		± 4.98%	± 10.13%	± 4.46%	± 0.13

**Table 2**

Sensitivity, PPV, DSC, HD rates obtained over NEMA IEC phantoms (I–V). Phantoms have the following signal contrast: (I) 1.5:1, (II) 3:1, (III) 5:1, (IV) 7:1, and (V) 8:1. Three different spherical lesions of size 22, 26, and 37 mm in diameter are considered to assess the performance of the proposed segmentation system. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

NEMA IEC Phantom	Sphere Diameter	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]
I	22 mm	72.03%	72.51%	80.52%	1.77
II		91.20%	88.05%	89.60%	1.17
III		92.01%	85.81%	88.82%	1.06
IV		90.52%	86.42%	88.41%	1.07
V		96.33%	87.90%	91.90%	1.00
I	26 mm	89.22%	81.32%	84.87%	1.71
II		92.91%	90.90%	91.90%	1.18
III		91.87%	88.98%	90.31%	1.05
IV		94.40%	90.20%	92.20%	1.00
V		99.80%	85.21%	91.91%	1.05
I	37 mm	89.04%	90.90%	89.90%	1.60
II		99.90%	88.22%	93.72%	1.19
III		99.90%	88.63%	93.90%	1.05
IV		99.41%	91.50%	95.35%	0.97
V		99.82%	93.81%	96.70%	0.89
<i>Mean</i>		93.22%	87.36%	90.67%	1.18
<i>± std</i>		± 7.18%	± 5.09%	± 4.07%	± 0.28
<i>± CI</i>		± 3.64%	± 2.58%	± 2.06%	± 0.14
<i>(95%)</i>					
<i>± CV</i>		± 7.71%	± 5.83%	± 4.49%	± 0.23

In order to investigate the inter-observer variance, a simultaneous ground truth estimation tool was employed (as suggested by Ref. [62]), and the results from three observers were combined to define a consolidated reference.

## 4. Results

### 4.1. Clinical testing and results on phantoms

Performance results have been divided in two tables considering small spheres (sphere diameters: 10, 13, and 17 mm) and large spheres

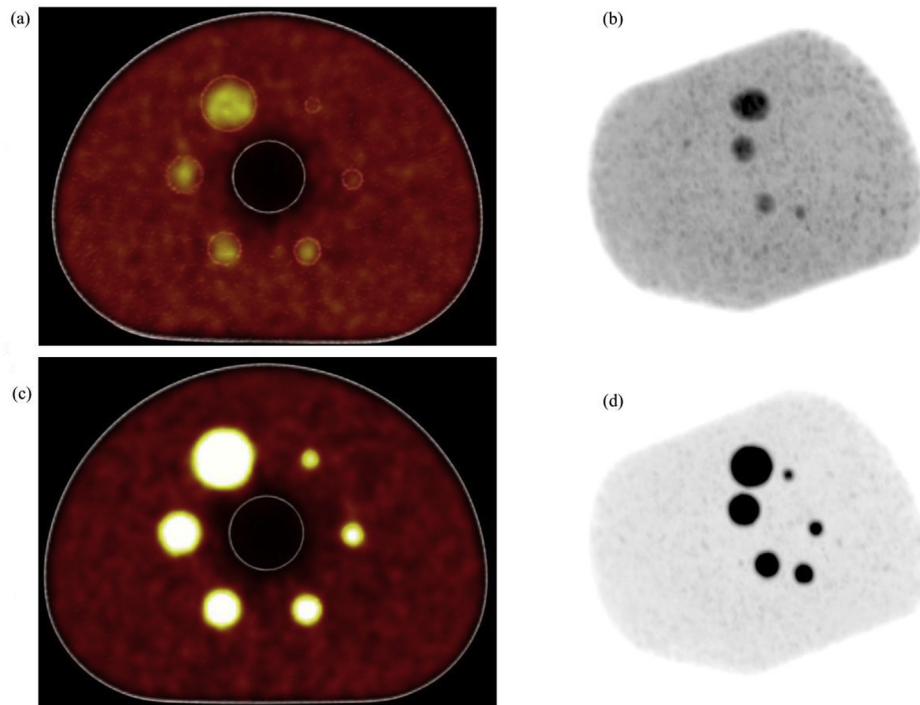


Fig. 3. PET/CT images of I and V phantoms. Phantoms have the following signal contrasts: (a) 1.5:1, (c) 8:1. PET volume rendering of phantom I (b), and V (d).

(sphere diameters: 22, 28, and 37 mm) as explained in Section 3.1. Tables 1 and 2 show sensitivity, PPV, DSC and HD values in five independent phantom experiments carried with different ratios between sphere concentration and background radioactivity concentration: 1.5:1 for the phantom “I”, 3:1 for the phantom “II”, 5:1 for the phantom “III”, 7:1 for the phantom “IV”, and 8:1 for the phantom “V”.

Table 1 shows the volumetric accuracy results for the smaller spheres, while Table 2 shows the volumetric accuracy results for the larger spheres. In Table 1, the results for the phantom “I” are missing because the ratio between sphere and background concentrations is too low (the target radioactivity concentration is only more than one-and-a-half times the background radioactivity concentration), and no PET region with high uptake concentration is visible for spheres with diameter < 22 mm. No significant difference between background and target is observed. For the same reason, the accuracy improved for all spheres, regardless of their volume, when the ratio between sphere concentration and background radioactivity concentration was increased. Phantom images with different signal contrast (1.5:1 and 8:1, respectively) are shown in Fig. 3.

In general, due to the PVE, the separation of small targets from the background region is very challenging, and the difficulty increases in critical conditions (i.e. low signal contrast). The volumes of smaller spheres are underestimated (mean difference between segmented and actual volumes =  $-19.12 \pm 10.69\%$ ) with more false negatives than false positives. As expected, large errors occur in lesions less than 2 cm in diameter [11].

Table 2 shows the results for all spheres with a diameter greater than 17 mm. In all conditions, excluding the phantom ‘I’ for the aforementioned reason, a DSC rate above 90% and a sensitivity rate greater than 90% are observed. These performances are generally accepted as excellent. A slight oversizing is observed (mean difference between segmented and actual volumes =  $6.69 \pm 5.23\%$ ), nevertheless, larger margins can help to prevent the extension of tumour infiltration in radiotherapy treatments [63].

In addition, high sensitivity, PPV and DSC, and low HD confirm the accuracy of the proposed segmentation system even when compared against the most common segmentation algorithms used in BT

extraction (RW, original LAC, RG, FCM, and T42%). Fig. 4 reports the quantitative comparison between the semi-automatic segmentation and the gold-standard considering all phantom experiments. Despite limitations due to the use of phantoms, results show that the proposed segmentation system out-performs the other algorithms tested for comparison.

#### 4.2. Clinical testing and results on patients

The evaluation of the segmentation system presented in this study was performed retrospectively (i.e. data were acquired before beginning of the radio-therapy and only at later time employed for the purposes of this study), on 50 tumours: in detail, 10 patients with lung cancer (FDG PET), 25 patients with head & neck cancers (FDG PET), and 15 patients with brain metastases (MET PET).

Performance was evaluated against a “ground truth” manual delineation provided by three expert operators [62]. Automatically segmented BTVs showed high agreement with the ground truth delineations (the determination coefficient  $R^2 = 0.98$ , see Fig. 5).

To assess the performance level represented by automatically and manual segmentations provided by three expert operators, we used STAPLE algorithm [62]. Results are provided in Table 3 as “mean  $\pm$  standard deviation ( $\pm$  CI) [ $\pm$  CV] of the combination of sensitivity and PPV (see Section 2.3).

Tables 4–6 show sensitivity, PPV, DSC, HD, and MHD results for the three lesion datasets (lung, head & neck, and brain). Since in the Cannizzaro Hospital dataset the segmented BTVs are greater than 2.5 ml (lesions with a sphere-equivalent diameter greater than 17 mm), no distinction between lesion volumes is reported as for phantom sphere results in the previous section. In addition, most segmented BTV ( $\sim 90\%$ ) are homogenous (one single peak is visually detected on the histogram of lesion voxel values). Fig. 6 reports the quantitative comparison between the operator independent segmentation and the gold-standard.

In addition, in Table 7 the proposed segmentation system is compared with the original LAC and RW methods (both methods obtained similar results and outperformed T42%, RG and FCM methods in



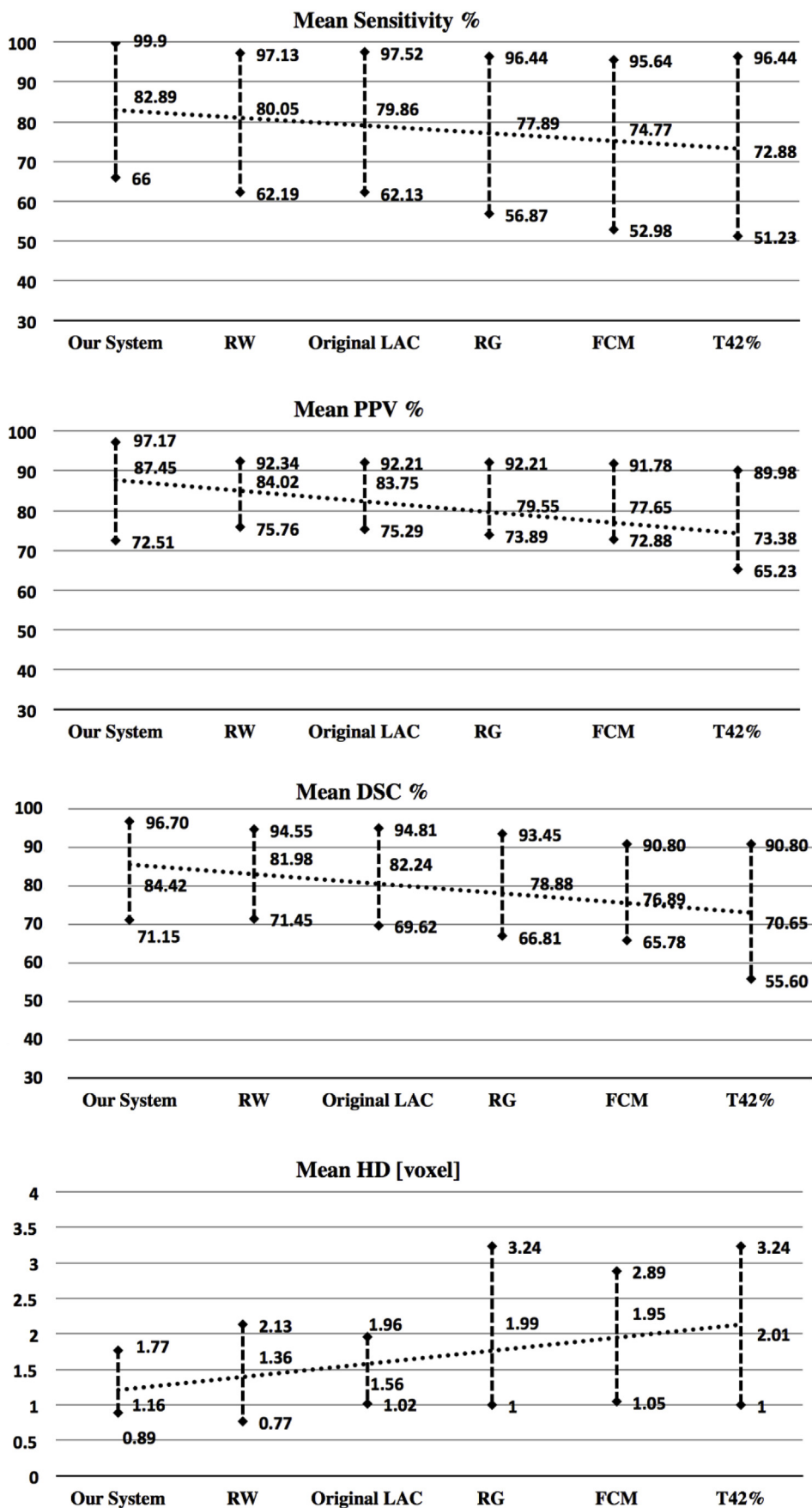
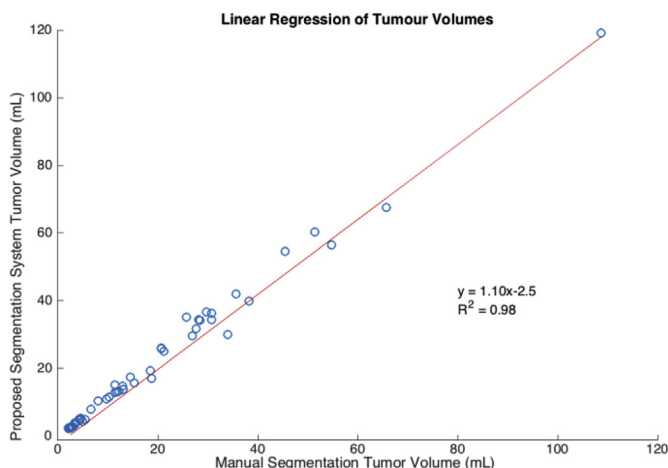


Fig. 4. Sensitivity, PPV, DSC, and HD comparisons (mean and range) of the proposed system to the ones commonly used in the literature in delineation of PET images.



**Fig. 5.** Linear regression of manually and automatically segmented tumour volumes. The determination coefficient  $R^2 = 0.98$  demonstrates high correlation between them. The first-order coefficient 1.10 indicates that segmentation slightly overestimates tumour volume. The constant offset of  $-2.5$  indicates that automatic segmentation may not identify small lesions ( $< 2.5$  ml).

phantom experiments, for this reason, T42%, RG and FCM methods were not considered in patient studies). The proposed segmentation yielded better performance in minimizing the difference between manual and automated BTVs than the other two reference methods. As a final test, in order to confirm that the system is indeed user-independent, a sub-dataset of 15 lesions (5 lung cancers, 5 head & neck cancers, and 5 brain metastases) was independently segmented using the proposed system by five different operators (the three experts mentioned above and two non-specialists). The result consisted of five identical BTVs. Without the RG pre-segmentation described in Section 2.1, BTVs changed less than 6% among users using manual ROIs drawn in the slice with the  $SUV_{max}^{VOI}$  by the five users in the above mentioned sub-dataset. Changing more significant (up to 50%) occurred if the users chose different PET slices among them to contour the lesion.

**5. Discussion**

PET quantification and segmentation are crucial for the accurate interpretation of clinical PET images and are of foremost importance to obtain operator-independent evaluations and results. The repeatability of the BTV delineation may be ensured only by using computer-assisted methods. For this reason, lesion delineation must be obtained using algorithms able to yield an accurate segmentation with consequent accurate calculation of parameters such as SUV, BTV, and TLG, minimizing operator-dependence and increasing result repeatability.

The key novelty of the proposed approach is the integration of different methodologies that have been appropriately combined and adapted to the PET imaging to obtain a new, smart system for the delineation of an operator independent BTV to be used for RTP or therapy evaluation purposes. The system achieves a high degree of automation without the need for any user-defined parameter.

First, we integrate the body-weighted SUV [51] in our system to pre-process the uptake data of PET images. The SUV normalizes the voxel

**Table 3**  
Performance level of the three manual segmentations and of the proposed automatic segmentation.

	Score	Score RT	Score FU
Expert 1	95.76% ± 0.23( ± 0.27)[ ± 0.25]	95.17% ± 0.31( ± 0.35)[ ± 0.33]	96.36% ± 0.41( ± 0.46)[ ± 0.42]
Expert 2	87.41% ± 0.88( ± 1.00)[ ± 1.01]	84.89% ± 0.95( ± 1.07)[ ± 1.12]	89.92% ± 0.82( ± 0.93)[ ± 0.91]
Expert 3	89.84% ± 1.85( ± 2.09)[ ± 2.06]	87.94% ± 2.42( ± 2.74)[ ± 2.71]	91.73% ± 1.28( ± 1.45)[ ± 1.40]
Our System	90.68% ± 1.34( ± 1.52)[ ± 1.48]	90.80% ± 1.35( ± 1.53)[ ± 1.49]	90.57% ± 1.34( ± 1.51)[ ± 1.47]

**Table 4**  
Mean Sensitivities, PPVs, DSCs, HDs, and MHDs for 10 lung cancer studies using FDG-PET examinations are reported. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

Lung Cancer	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]	MHD [voxels]
#1	91.60%	81.70%	86.40%	2.49	1.60
#2	97.60%	82.80%	89.60%	1.98	1.39
#3	98.50%	71.60%	82.90%	2.59	2.09
#4	85.00%	85.00%	85.00%	1.08	0.83
#5	98.90%	83.20%	90.40%	1.48	0.72
#6	88.80%	77.20%	82.60%	1.69	0.84
#7	91.00%	73.60%	81.40%	1.14	0.48
#8	96.70%	78.40%	86.60%	1.25	0.91
#9	94.70%	76.30%	84.50%	2.60	0.67
#10	93.50%	76.60%	84.20%	2.38	0.60
<b>Mean</b>	<b>93.63%</b>	<b>78.64%</b>	<b>85.36%</b>	<b>1.87</b>	<b>1.01</b>
± std	± 4.55%	± 4.41%	± 2.94%	± 0.62	± 0.51
± CI (95%)	± 2.82%	± 2.73%	± 1.82%	± 0.38	± 0.32
± CV	± 4.86%	± 5.60%	± 3.44%	± 0.33	± 0.51

**Table 5**  
Mean Sensitivities, PPVs, DSCs, HDs, and MHDs for 25 head & neck cancer studies using FDG-PET examinations are reported. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

Neck & head Cancer	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]	MHD [voxels]
#1	98.70%	81.30%	89.10%	1.46	0.43
#2	83.30%	89.70%	86.40%	1.00	0.69
#3	89.80%	76.20%	82.40%	1.07	0.74
#4	85.50%	94.60%	89.80%	0.85	0.59
#5	93.20%	74.80%	83.00%	1.17	0.36
#6	96.20%	81.30%	88.10%	1.42	1.41
#7	93.90%	89.90%	91.80%	1.19	1.18
#8	99.50%	71.80%	83.40%	2.79	0.60
#9	96.90%	78.80%	86.90%	2.46	1.37
#10	97.50%	77.10%	86.10%	2.59	0.72
#11	73.10%	90.50%	80.90%	1.00	0.71
#12	99.60%	73.70%	84.70%	1.40	0.69
#13	93.80%	73.80%	82.60%	1.03	0.94
#14	91.40%	86.50%	88.90%	1.00	0.78
#15	88.80%	74.60%	81.10%	2.78	2.28
#16	77.60%	84.90%	81.10%	1.07	1.01
#17	89.80%	86.30%	88.00%	1.14	0.75
#18	91.90%	87.00%	89.40%	0.90	0.81
#19	91.40%	80.00%	85.30%	1.37	1.27
#20	96.30%	83.90%	89.70%	0.90	1.16
#21	97.30%	80.70%	88.20%	1.88	1.98
#22	80.80%	95.10%	87.40%	0.94	1.07
#23	78.70%	92.50%	85.10%	0.93	1.00
#24	93.30%	70.00%	80.00%	1.33	1.52
#25	96.60%	84.30%	90.00%	0.99	0.58
<b>Mean</b>	<b>91.00%</b>	<b>82.37%</b>	<b>85.98</b>	<b>1.39</b>	<b>0.99</b>
± std	± 7.33%	± 7.30%	± 3.40%	± 0.61	± 0.46
± CI (95%)	± 2.87%	± 2.86%	± 1.33%	± 0.24	± 0.18
± CV	± 8.05%	± 8.86%	± 3.95%	± 0.44	± 0.47

**Table 6**

Mean Sensitivities, PPVs, DSCs, HDs, and MHDs for 15 brain cancer studies using MET-PET examinations are reported. Mean value, standard deviation (std), 95% confidence interval (CI) and coefficient of variation (CV) are reported in the last rows.

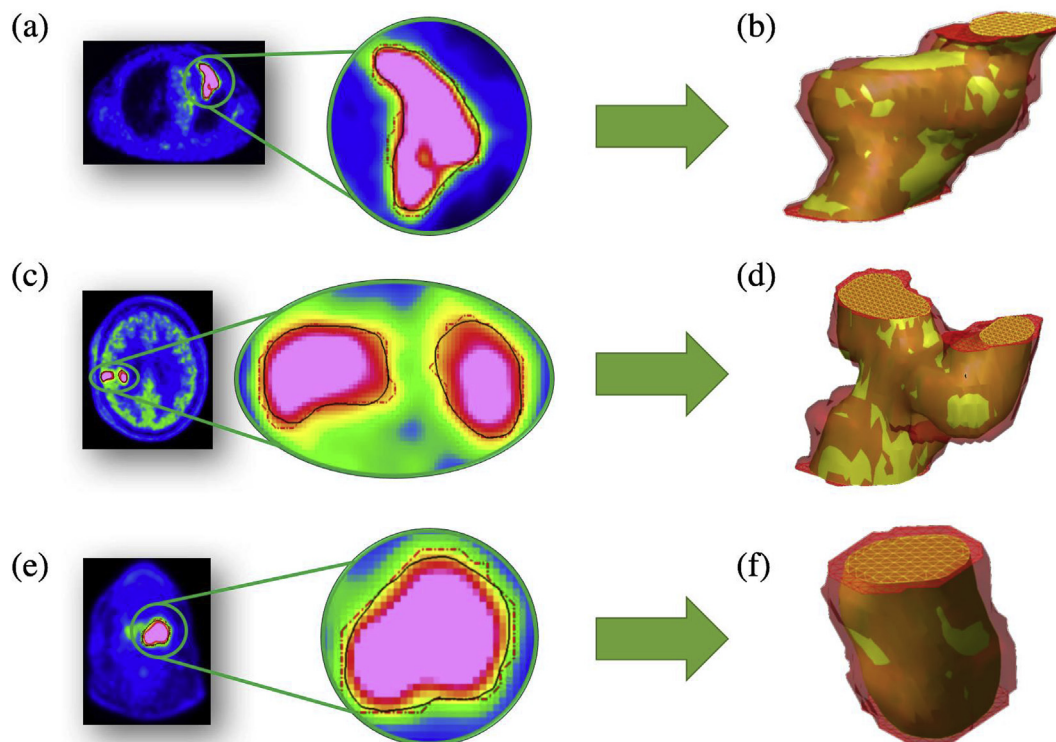
Brain Cancer	Sensitivity [%]	PPV [%]	DSC [%]	HD [voxels]	MHD [voxels]
#1	93.50%	89.90%	91.70%	0.62	1.06
#2	90.10%	78.60%	83.90%	1.33	1.14
#3	88.40%	82.80%	85.50%	1.54	0.97
#4	76.75%	99.18%	86.33%	0.58	0.61
#5	86.47%	95.89%	90.83%	0.58	0.56
#6	96.20%	75.60%	84.70%	1.91	1.18
#7	91.80%	79.50%	85.20%	2.37	1.22
#8	94.50%	84.70%	89.40%	0.79	0.85
#9	86.80%	83.30%	85.00%	1.24	1.05
#10	91.20%	88.70%	90.00%	1.40	0.93
#11	91.80%	82.90%	87.10%	1.54	0.71
#12	85.20%	97.00%	90.70%	1.38	0.59
#13	93.20%	90.30%	91.70%	1.54	1.07
#14	93.00%	84.70%	88.70%	1.81	1.72
#15	93.60%	85.80%	89.50%	0.65	0.57
<i>Mean</i>	90.17	86.60%	88.02	1.28	0.95
$\pm$ <i>std</i>	$\pm$ 4.89%	$\pm$ 6.89%	$\pm$ 2.75%	$\pm$ 0.54	$\pm$ 0.32
$\pm$ <i>CI</i> (95%)	$\pm$ 2.47%	$\pm$ 3.49%	$\pm$ 1.39%	$\pm$ 0.28	$\pm$ 0.16
$\pm$ <i>CV</i>	$\pm$ 5.42%	$\pm$ 7.95%	$\pm$ 3.12%	$\pm$ 0.42	$\pm$ 0.33

activity considering acquisition time, administered activity, and patient's weight. Others PET parameters (i.e.,  $SUV_{lbm}$  or  $SUL_{peak}$  [2]) could be used rather than body-weighted SUV as well with minor changes to the system that would still lead a convergent, fully

repeatable, and user independent segmentation. We use the body-weighted SUV because it is the most common parameter used in our nuclear medicine department.

Second, while the system still exploits an initial ROI provided by the user to avoid false positive or critical conditions, an automatic pre-segmentation process is used to compute a robust user independent ROI in the vicinity of the user input. This is then automatically fed as input into the detailed segmentation algorithm. Third, segmentation is performed using a slice-by-slice marching approach based on the LAC method appropriately modified to support PET images. And as final feature a fully automatic stop condition is provided. Due to the high level of automation for the whole process, the final results become completely independent of variation in the initial user input, as confirmed by the inter operator segmentation test in Section 4.2.

The main reason to require the user to provide an initial ROI is because the distribution of FDG is not limited to malignant tissue. FDG enters the cells according to glucose transport mechanism. For this reason, in order to avoid healthy but active structures where a high radiotracer uptake is to be considered normal, the target lesions must be highlighted by the operator. The region growing is used to “expand” the initial user-provided manual segmentation (performed on one slice) to a three-dimensional volume enclosing the anomalies of interest. We chose region growing because it is able to identify a contour encircling the tumour area starting from a single seed point (the voxel with  $SUV_{max}^j$ ) always obtaining the same starting contour (Fig. 1e). The use of alternative segmentation methods for the computation of the user independent ROI would lead to a slight BTv variation. However, the identification of the hottest slice (the slice with  $SUV_{max}^{VOI}$ ) in which to compute the user independent ROI is mandatory to avoid a great variability of the final segmentations (see Section 4.2). The proposed pre-segmentation process used to identify the hottest slice and,



**Fig. 6.** Three results achieved by the proposed segmentation method on PET images (lung cancer #2, head and neck cancer #25, and brain cancer #13) are reported in figure (a, c, e), respectively. At the bottom right or left, the lesion regions are zoomed. The proposed segmentations (red contours) and the gold standards (black contours) are superimposed. The three-dimensional reconstructions of the tumours are shown in figure (b, d, e): manual (yellow) and proposed (red) BTVs are rendered with transparent surfaces to emphasize volume intersections. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

**Table 7**  
Sensitivities, PPVs, DSCs and HDs for cancer studies using the proposed system, original LAC and RW methods.

Cancer	Sensitivity [Mean $\pm$ std]	PPV [Mean $\pm$ std]	DSC [Mean $\pm$ std]	HD [Mean $\pm$ std]
<b>Our System</b>				
Lung	93.63 $\pm$ 4.55%	78.64 $\pm$ 4.41%	85.36 $\pm$ 2.94%	1.87 $\pm$ 0.62
Head & Neck	91.00 $\pm$ 7.33%	82.37 $\pm$ 7.30%	85.98 $\pm$ 3.40%	1.39 $\pm$ 0.61
Brain	90.17 $\pm$ 4.89%	86.60 $\pm$ 6.89%	88.02 $\pm$ 2.75%	1.28 $\pm$ 0.54
<b>Original LAC</b>				
Lung	92.68 $\pm$ 5.67%	71.92 $\pm$ 12.36%	80.33 $\pm$ 9.51%	3.29 $\pm$ 0.74
Head & Neck	80.32 $\pm$ 10.82%	83.88 $\pm$ 12.06%	78.73 $\pm$ 6.68%	2.81 $\pm$ 0.49
Brain	91.16 $\pm$ 6.16%	77.38 $\pm$ 13.89%	83.55 $\pm$ 9.03%	2.88 $\pm$ 0.54
<b>RW</b>				
Lung	92.55 $\pm$ 5.66%	74.72 $\pm$ 7.44%	82.03 $\pm$ 6.52%	2.29 $\pm$ 0.77
Head & Neck	79.03 $\pm$ 7.29%	84.58 $\pm$ 9.57%	82.59 $\pm$ 4.95%	1.41 $\pm$ 0.40
Brain	88.12 $\pm$ 8.36%	86.68 $\pm$ 3.43%	86.75 $\pm$ 5.14%	1.31 $\pm$ 0.64

consequently, to compute the user independent ROI is crucial to obtain repeatable results.

Further, in the case of brain lesions, as reported in Ref. [16], a fully automatic method can be implemented by taking advantage of the great sensitivity and specificity of MET radiotracers in the discrimination between healthy versus cancer. Hence initial user input, which we require in FDG-PET studies, could be completely avoided in brain studies.

Performance measurements of the proposed segmentation system have been obtained by phantom experiments consisting of hot spheres in a warm background. While such phantoms don't replicate all the properties of real tissue, they nevertheless represent a useful tool to assess common reference performances across different algorithms. One of the main drawbacks of this validation is the plastic wall which separates the spheres from the background. Indeed, PET image quantification should be carried out in phantoms with inactive background to reduce the cold wall effect [60]. Nevertheless, this condition is too different from actual clinical conditions; the model proposed by Ref. [60] may not be useful to test segmentation algorithms [64]. To ease the problem, we insert hot spheres in a warm background. Another drawback of the body phantom validation comes from the use of homogeneous spheres. However, in our patient studies, almost all the tumours are characterized by a fairly homogeneous distribution of uptake. A single peak is generally detected on the histogram of segmented lesions (see Section 4.2). In this context, therefore, segmentation performance using phantom experiments can be considered a suitable tool to judge the reliability of the method.

In particular, the proposed segmentation system showed high degree of similarity (DSC and sensitivity greater than 90% for the spheres with a diameter greater than 17 mm) and provided better results in minimizing the difference between actual and automated BTVs than the other state-of-the-art methods.

Nevertheless, when lesion sizes are smaller than 2–3 times the FWHM of the point spread function of the PET image resolution reconstructed by the PET imaging system, the under-estimation of metabolic activity due to PVE cannot be assumed to be negligible. The separation of small oncological lesions from the background region is still very challenging [65].

Concerning patient studies, knowledge of the ground truth is impossible to obtain while patients are still undergoing radiotherapy, and unfortunately, the only valid ground truth for quantitative segmentation evaluation is obtained by histological investigations, in general performed after that long radio-therapeutic cycles altered the cancer mass morphology. Although histological specimen is used in some studies [34,35], the approach may be problematic because irregular contractions can occur during tissue fixation. Consequently, manual delineation of three experts was used as a gold-standard. Nevertheless, the manual BTV delineation is a challenging task even for expert physicians in deciding which regions to include or exclude in the RTP. It is different between operators (for example, radiotherapy planning experts tend to draw larger contours than nuclear medicine physicians and that

obviously has a strong impact on the resulting surrogate of truth). For this reason, to assess the applicability of the proposed algorithms in a clinical environment and variable conditions (radiotracers and body districts), manual segmentations were used to produce a consolidated reference (as suggested in Ref. [62]).

Fifty patients that had been referred for a diagnostic PET/CT scan before radiotherapy treatments have been considered. Patients underwent PET examinations with FDG or MET radiotracers because different body districts have been considered (10 lung cancers, 25 head & neck cancers and 15 brain metastases). Results show that the proposed segmentation approach can be considered clinically feasible, since it has been integrated in the current clinical practice. In addition, automatically segmented tumour volumes showed high agreement with the manual segmentations ( $R2 = 0.98$ ). Considering that the proposed system has been implemented in the Matlab R2016a environment and comfortably runs on a standard iMac machine, a qualified IT staff could integrate the system in the PET workstations, as a built-in PET tool for use in medical environment, with a minimum effort. This solution could significantly improve imaging workflow allowing clinicians to include BTV information into the RTP, so obtaining a therapy customized on the patient. In addition, while in our study the user initial input was provided by mouse, user input experience could be easily enhanced using alternative input methods (e.g. stylus, touch-screen, etc.), without affecting the result.

Finally, the proposed approach is different with respect to joint segmentation approaches (PET-CT-MR). The assumption of a “ground truth” joint volume, defined on fused multimodal imaging data is sometimes misleading. The assumption that a one-to-one correspondence between metabolic and anatomical regions exists is unrealistic [38]. Lesions may present smaller uptake regions compared to anatomical region. In the same way, the PET lesion may show additional area compared to lesion boundaries in CT or MR images. Since different imaging modalities could convey different information, multimodal segmentation may actually compromise the quality of segmentation producing disagreement between anatomical and metabolic boundaries. As a result, the tumour volumes defined on PET and on CT or MRI could be highly different [9]. For this reason, multimodal studies must be combined in a smart fashion with a customization/adaptation of the segmentation method according to the specific clinical scenario. For example, in the clinical case shown in Ref. [66], semi-automatic BTV radically changed the RTP because uptake is found outside the anatomical volume in an involved lymph node which is not visible in CT images. However, the extraction of anatomical structures by CT examination can still convey useful information to locate health tissues (i.e. brain, heart, bladder, and kidneys) so to avoid ambiguities and false positives and removing the confounding contribution of FDG-avid normal tissues from the analysis. In addition, further investigations will be carried out to assess the prognostic usefulness and long-term clinical impact of the proposed system, comparing the extracted BTV with clinical outcomes, progression-free survival and overall survival.

## 6. Conclusions

In the proposed study, several methodologies have been combined to obtain a system for the smart segmentation of PET volumes. The key advantages of the presented system are that:

- segmentation is performed after converting the PET data in SUVs, which convey patient-specific functional information.
- based on minimal user input (i.e. a rough manual delineation on just one PET slice), a pre-segmentation step identifies the most relevant PET slice and provides an initial properly localized contouring to be used for a subsequent segmentation using LAC.
- a more detailed segmentation process based on LAC and performed following a slice-by-slice based approach provides the flexibility necessary to segment cancer with very irregular shapes.
- an automatic stopping condition is implemented.

As a result, the presented algorithm produces cancer segmentations which are completely independent from variations in the input provided by the user. Such an input is minimal and, since subsequent processing steps are fully automatized, no further interaction is required. Inter-operator variations in the initial input still result in identical segmentations.

## Appendix

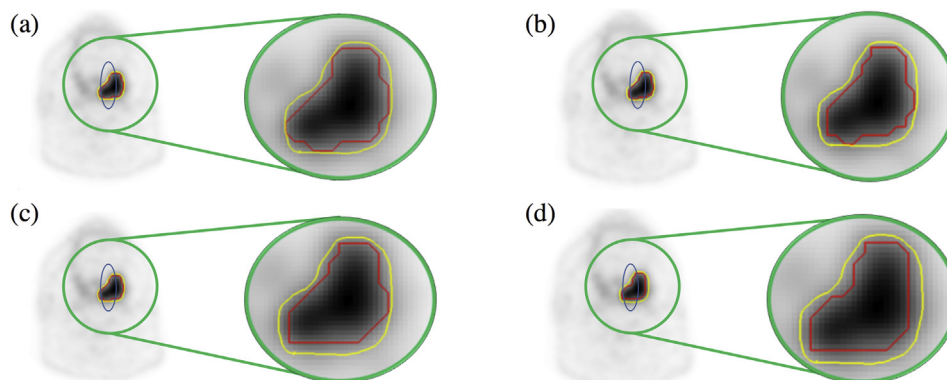
In Eq. (3), the radius parameter  $l$  has been determined to provide the best performance in our dataset by trial and error strategy.

When using energies based on local features, the radius defining the local neighbourhood (number of pixels comprised at its edge) defines how sensitive to local features the segmentation is meant to be. By tuning this parameter, one can choose the degree to which local and global behaviour are blend. A small radius behaves as an edge detector based on the statistics of the pixels immediately adjacent to one being investigated. Vice versa, if the radius is let to grow to includes the entire image, then the local region statistics will coincide exactly with the global statistics, and the behaviour would be the same as performing a global region-based flow.

One aspect that must be considered for the choice of the radius is the order of magnitude of the tumour size and the amount of surrounding noise. When attempting to capture tumours that are very small with nearby noise, as is the case in our study, a small localization radius is advisable.

In our preliminary experiments, this parameter ranged from 3 (very local, small compared to tumour size) to 9 (more global, size large when compared to the features of the tumour edge). To our opinion the selected range covered well enough the values of interest of this parameter with respect to the data at hand. After our testing, the best value empirically determined was “3”.

Fig. A1 shows an example in the head/neck area segmented using four different local radii (3, 5, 7, and 9).



**Fig. A1.** The figure shows the same initial contour on the PET image (blue line) and resulting segmentations using localizing radius 3 (a), 5 (b), 7 (c), and 9 (d) respectively (gold standard and resulting segmentations are shown as yellow and red lines in each image, respectively).

On the one hand (and as expected), the result suggests a small degradation of the result as the radius increases (i.e. as the energy becomes more global). On the other hand, segmentations were only slightly influenced by the radius.

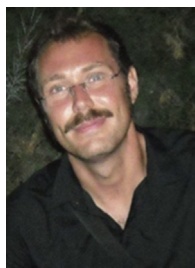
We therefore selected the most suitable value for the data at hand. To practical purposes this aspect does not impact the automatism of the system.

## References

- [1] P. Alongi, F. Caobelli, R. Gentile, et al., Recurrent bladder carcinoma: clinical and prognostic role of 18 F-FDG PET/CT, *Eur. J. Nucl. Med. Mol. Imag.* (2016), <https://doi.org/10.1007/s00259-016-3500-8>.
- [2] R.L. Wahl, H. Jacene, Y. Kasamon, M.A. Lodge, From RECIST to PERCIST: evolving Considerations for PET response criteria in solid tumors, *J. Nucl. Med.* 50 (Suppl 1) (2009), <https://doi.org/10.2967/jnumed.108.057307> 122S–50S.
- [3] A. Stefano, S. Vitabile, G. Russo, et al., An automatic method for metabolic evaluation of gamma knife treatments, *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Springer, Cham, 2015, pp.

- 579–589.
- [4] E. Allegra, V. Saita, M. De Natale, et al., Use of PET/CT to detect local and regional laryngeal cancer recurrence after surgery, *Rep. Med. Imag.* 10 (2017) 31–36, <https://doi.org/10.2147/RM.S124764>.
- [5] J.W. Fletcher, P.E. Kinahan, PET/CT standardized uptake values (SUVs) in clinical practice and assessing response to therapy, *NIH Public Access* 31 (2010) 496–505, <https://doi.org/10.1053/j.sult.2010.10.001.PET/CT>.
- [6] S.M. Larson, Y. Erdi, T. Akhurst, et al., 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, *Clin. Positron Imaging* 2 (1999) 159–171, [https://doi.org/10.1016/S1095-0397\(99\)00016-3](https://doi.org/10.1016/S1095-0397(99)00016-3).
- [7] A. Stefano, N. Porcino, G. Banna, et al., Metabolic response assessment in non-small cell lung cancer patients after platinum-based therapy: a preliminary analysis, *Curr. Med. Imag. Rev.* 11 (2015) 218–227.
- [8] D. Wang, D.M. Doddrell, Geometric distortion in structural magnetic resonance imaging, *Curr. Med. Imag. Rev.* 1 (2005) 49–60, <https://doi.org/10.2174/1573405052953029>.
- [9] L. Rundo, A. Stefano, C. Militello, et al., A fully automatic approach for multimodal PET and MR image segmentation in gamma knife treatment planning, *Comput. Meth. Progr. Biomed.* 144 (2017) 77–96, <https://doi.org/10.1016/j.cmpb.2017.03.011>.
- [10] M. Niyazi, S. Landrock, A. Elsner, et al., Automated biological target volume delineation for radiotherapy treatment planning using FDG-PET/CT, *Radiat. Oncol.* 8 (2013) 180, <https://doi.org/10.1186/1748-717X-8-180>.
- [11] M. Soret, S.L. Bacharach, Buvat II, Partial-volume effect in PET tumor imaging, *J. Nucl. Med.* 48 (2007) 932–945, <https://doi.org/10.2967/jnumed.106.035774>.
- [12] R. Boellaard, M.J. O'Doherty, W.A. Weber, et al., FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0, *Eur. J. Nucl. Med. Mol. Imag.* 37 (2010) 181–200, <https://doi.org/10.1007/s00259-009-1297-4>.
- [13] H. Zaidi, I. El Naqa, PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques, *Eur. J. Nucl. Med. Mol. Imag.* 37 (2010) 2165–2187, <https://doi.org/10.1007/s00259-010-1423-3>.
- [14] B. Foster, U. Bagci, A. Mansoor, et al., A review on segmentation of positron emission tomography images, *Comput. Biol. Med.* 50 (2014) 76–96, <https://doi.org/10.1016/j.compbiomed.2014.04.014>.
- [15] M. Hatt, B. Laurent, A. Ouahabi, et al., The first MICCAI challenge on PET tumor segmentation, *Med. Image Anal.* 44 (2018) 177–195, <https://doi.org/10.1016/j.media.2017.12.007>.
- [16] A. Stefano, S. Vitabile, G. Russo, et al., A fully automatic method for biological target volume segmentation of brain metastases, *Int. J. Imag. Syst. Technol.* 26 (2016) 29–37, <https://doi.org/10.1002/ima.22154>.
- [17] T. Nariai, Y. Tanaka, H. Wakimoto, et al., Usefulness of 1-[methyl-<sup>11</sup>C] methionine—positron emission tomography as a biological monitoring tool in the treatment of glioma, *J. Neurosurg.* 103 (2005) 498–507, <https://doi.org/10.3171/jns.2005.103.3.0498>.
- [18] E. Prieto, P. Lecumberri, M. Pagola, et al., Twelve automated thresholding methods for segmentation of PET images: a phantom study, *Phys. Med. Biol.* 57 (2012) 3963–3980, <https://doi.org/10.1088/0031-9155/57/12/3963>.
- [19] D.A.X. Schinagel, P.N. Span, F.J.A. Van Den Hoogen, et al., Pathology-based validation of FDG PET segmentation tools for volume assessment of lymph node metastases from head and neck cancer, *Eur. J. Nucl. Med. Mol. Imag.* 40 (2013) 1828–1835, <https://doi.org/10.1007/s00259-013-2513-9>.
- [20] E. Day, J. Betler, D. Parda, et al., A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients, *Med. Phys.* 36 (2009) 4349–4358, <https://doi.org/10.1118/1.3213099>.
- [21] C.-H. Kao, T.-C. Hsieh, C.-Y. Yu, et al., 18F-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, *Radiat. Oncol.* 5 (2010) 76, <https://doi.org/10.1186/1748-717X-5-76>.
- [22] J.A. Lee, Segmentation of positron emission tomography images: some recommendations for target delineation in radiation oncology, *Radiother. Oncol.* 96 (2010) 302–307, <https://doi.org/10.1016/j.radonc.2010.07.003>.
- [23] S. Tan, L. Li, W. Choi, et al., Adaptive region-growing with maximum curvature strategy for tumor segmentation in 18F-FDG PET, *Phys. Med. Biol.* (2017), <https://doi.org/10.1088/1361-6560/aa6e20>.
- [24] M. Wanet, J.A. Lee, B. Weynand, et al., 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, *Radiother. Oncol.* 98 (2011) 117–125, <https://doi.org/10.1016/j.radonc.2010.10.006>.
- [25] X. Geets, J.A. Lee, A. Bol, et al., A gradient-based method for segmenting FDG-PET images: methodology and validation, *Eur. J. Nucl. Med. Mol. Imag.* 34 (2007) 1427–1438, <https://doi.org/10.1007/s00259-006-0363-4>.
- [26] M.S. Sharif, M. Abbod, A. Amira, H. Zaidi, Artificial neural network-based system for PET volume segmentation, *Int. J. Biomed. Imag.* (2010), <https://doi.org/10.1155/2010/105610>.
- [27] Aliaksei Kerhet, Cormac Small, Quon Harvey, Terence Riauka, Russell Greiner, W.R. Alexander McEwan, Segmentation of lung tumours in positron emission tomography scans: a machine learning approach, *Artificial Intelligence in Medicine. AIME 2009. Lecture Notes in Computer Science*, 5651 2009, pp. 146–155 [https://link.springer.com/chapter/10.1007/978-3-642-02976-9\\_19#citeas](https://link.springer.com/chapter/10.1007/978-3-642-02976-9_19#citeas).
- [28] S. Belhassen, H. Zaidi, A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET, *Med. Phys.* 37 (2010) 1309–1324, <https://doi.org/10.1118/1.3301610>.
- [29] U. Bagci, J.K. Udupa, N. Mendhiratta, et al., Joint segmentation of anatomical and functional images: applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images, *Med. Image Anal.* 17 (2013) 929–945, <https://doi.org/10.1016/j.media.2013.05.004>.
- [30] B. Foster, U. Bagci, B. Luna, et al., Robust segmentation and accurate target definition for positron emission tomography images using affinity propagation, 2013 IEEE 10TH Int. Symp. Biomed. Imaging, 2013.
- [31] M. Aristophanous, B.C. Penney, M.K. Martel, C.A. Pelizzari, A Gaussian mixture model for definition of lung tumor volumes in positron emission tomography, *Med. Phys.* 34 (2007) 4223–4235, <https://doi.org/10.1118/1.2791035>.
- [32] M. Hatt, C.C. le Rest, A. Turzo, et al., A fuzzy locally adaptive bayesian segmentation approach for volume determination in PET, *IEEE Trans. Med. Imag.* 28 (2009) 881–893, <https://doi.org/10.1109/tmi.2008.2012036>.
- [33] M. Hatt, C. Cheze le Rest, P. Descourt, et al., Accurate automatic delineation of heterogeneous functional volumes in positron emission tomography for oncology applications, *Int. J. Radiat. Oncol. Biol. Phys.* 77 (2010) 301–308, <https://doi.org/10.1016/j.ijrobp.2009.08.018>.
- [34] H. Li, W.L. Thorstad, K.J. Biehl, et al., A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours, *Med. Phys.* 35 (2008) 3711–3721, <https://doi.org/10.1118/1.2956713>.
- [35] M. Abdoli, R.A.J.O. Dierckx, H. Zaidi, Contourlet-based active contour model for PET image segmentation, *Med. Phys.* (2013), <https://doi.org/10.1118/1.4816296>.
- [36] M. Zhuang, R.A.J.O. Dierckx, H. Zaidi, Generic and robust method for automatic segmentation of PET images using an active contour model, *Med. Phys.* 43 (2016) 4483–4494, <https://doi.org/10.1118/1.4954844>.
- [37] D.P. Onoma, S. Ruan, S. Thureau, et al., Segmentation of heterogeneous or small FDG PET positive tissue based on a 3D-locally adaptive random walk algorithm, *Comput. Med. Imag. Graph.* 38 (2014) 753–763, <https://doi.org/10.1016/j.compmedimag.2014.09.007>.
- [38] A. Stefano, S. Vitabile, G. Russo, et al., An enhanced random walk algorithm for delineation of head and neck cancers in PET studies, *Med. Biol. Eng. Comput.* 55 (2017) 897–908, <https://doi.org/10.1007/s11517-016-1571-0>.
- [39] W. Ju, D. Xiang, B. Zhang, et al., Random walk and graph cut for Co-segmentation of lung tumor on PET-CT images, *IEEE Trans. Image Process.* 24 (2015) 5854–5867, <https://doi.org/10.1109/TIP.2015.2488902>.
- [40] C. Ballangan, X. Wang, M. Fulham, et al., Lung tumor segmentation in PET images using graph cuts, *Comput. Meth. Progr. Biomed.* 109 (2013) 260–268, <https://doi.org/10.1016/j.cmpb.2012.10.009>.
- [41] X.-Y. Wang, Y.-F. Zhao, Y. Liu, et al., Comparison of different automated lesion delineation methods for metabolic tumor volume of 18F-FDG PET/CT in patients with stage I lung adenocarcinoma, *Medicine (Baltim.)* (2017), <https://doi.org/10.1097/MD.0000000000009365>.
- [42] M. Hatt, B. Laurent, H. Fayad, et al., Tumour functional sphericity from PET images: prognostic value in NSCLC and impact of delineation method, *Eur. J. Nucl. Med. Mol. Imag.* (2018), <https://doi.org/10.1007/s00259-017-3865-3>.
- [43] C. Parkinson, K. Foley, P. Whybra, et al., Evaluation of prognostic models developed using standardised image features from different PET automated segmentation methods, *EJNMMI Res.* (2018), <https://doi.org/10.1186/s13550-018-0379-3>.
- [44] B. Wu, P.-L. Khong, T. Chan, Automatic detection and classification of nasopharyngeal carcinoma on PET/CT with support vector machine, *Int. J. Comput. Assist. Radiol. Surg.* 7 (2012) 635–646, <https://doi.org/10.1007/s11548-011-0669-y>.
- [45] L. Bi, J. Kim, L. Wen, et al., Automated thresholded region classification using a robust feature selection method for PET-CT, 2015 IEEE 12th Int. Symp. Biomed. Imaging, IEEE, 2015, pp. 1435–1438.
- [46] A. Comelli, A. Stefano, V. Benfante, G. Russo, Normal and abnormal tissue classification in PET oncological studies, *Pattern Recogn. Image Anal.* 28 (2018) 121–128, <https://doi.org/10.1134/S1054661818010054>.
- [47] M. Hatt, F. Tixier, L. Pierce, et al., Characterization of PET/CT images using texture analysis: the past, the present... any future? *Eur. J. Nucl. Med. Mol. Imag.* (2017), <https://doi.org/10.1007/s00259-016-3427-0>.
- [48] M.L. Belli, M. Mori, S. Broggi, et al., Quantifying the robustness of [18F]FDG-PET/CT radiomic features with respect to tumor delineation in head and neck and pancreatic cancer patients, *Phys. Med.* 49 (2018) 105–111, <https://doi.org/10.1016/j.ejomp.2018.05.013>.
- [49] A. Sbei, K. ElBedoui, W. Barhoumi, et al., Hybrid PET/MRI co-segmentation based on joint fuzzy connectedness and graph cut, *Comput. Meth. Progr. Biomed.* 149 (2017) 29–41, <https://doi.org/10.1016/j.cmpb.2017.07.006>.
- [50] A. Yezzi, L. Zollei, T. Kapur, A variational framework for joint segmentation and registration, *Proc. IEEE Work. Math. Methods Biomed. Image Anal., IEEE Comput. Soc.* 2001, pp. 44–51 MMBIA.
- [51] M.M. Graham, L.M. Peterson, R.M. Hayward, Comparison of simplified quantitative analyses of FDG uptake, *Nucl. Med. Biol.* 27 (2000) 647–655, [https://doi.org/10.1016/S0969-8051\(00\)00143-8](https://doi.org/10.1016/S0969-8051(00)00143-8).
- [52] J. Feuardent, M. Soret, O. de Dreuille, et al., Reliability of SUV estimates in FDG PET as a function of acquisition and processing protocols, *Nucl Sci Symp Conf Rec* 2003 IEEE 4, vol. 4, 2003, pp. 2877–2881, <https://doi.org/10.1109/NSSMIC.2003.1352484>.
- [53] A.-S. Dewalle-Vignion, N. Yeni, G. Petyt, et al., Evaluation of PET volume segmentation methods, *Nucl. Med. Commun.* 33 (2012) 34–42, <https://doi.org/10.1097/MNM.0b013e32834d736f>.
- [54] S. Lankton, D. Nain, A. Yezzi, A. Tannenbaum, J. Hsieh, M.J. Flynn (Eds.), *Hybrid Geodesic Region-based Curve Evolutions for Image Segmentation*, International Society for Optics and Photonics, 2007, p. 65104U.
- [55] J.K. Udupa, V.R. Leblanc, Y. Zhuge, et al., A framework for evaluating image segmentation algorithms, *Comput. Med. Imag. Graph.* 30 (2006) 75–87, <https://doi.org/10.1016/j.compmedimag.2005.12.001>.
- [56] A.A. Taha, A. Hanbury, Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool, *BMC Med. Imag.* 15 (2015) 29, <https://doi.org/10.1186/s12880-015-0068-x>.

- [57] M. Hatt, J.A. Lee, C.R. Schmidlein, et al., Classification and evaluation strategies of auto-segmentation approaches for PET: report of AAPM task group No.211, *Med. Phys.* 44 (2017).
- [58] Y. Liu, S. Liu, M.S. Nacif, et al., A framework to measure myocardial extracellular volume fraction using dual-phase low dose CT images, *Med. Phys.* (2013), <https://doi.org/10.1118/1.4819936>.
- [59] N. Aspert, D. Santa-Cruz, T. Ebrahimi, MESH: measuring errors between surfaces using the Hausdorff distance, *Proc - 2002 IEEE Int Conf Multimed Expo, ICME 2002, 2002*, <https://doi.org/10.1109/ICME.2002.1035879>.
- [60] F. Hofheinz, S. Dittich, C. Pöttsch, J van den Hoff, Effects of cold sphere walls in PET phantom measurements on the volume reproducing threshold, *Phys. Med. Biol.* 55 (2010) 1099–1113, <https://doi.org/10.1088/0031-9155/55/4/013>.
- [61] F. Gallivanone, A. Stefano, E. Grosso, et al., PVE correction in PET-CT whole-body oncological studies from PVE-affected images, *IEEE Trans. Nucl. Sci.* 58 (2011) 736–747.
- [62] S.K. Warfield, K.H. Zou, W.M. Wells, Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation, *IEEE Trans. Med. Imag.* 23 (2004) 903–921, <https://doi.org/10.1109/TMI.2004.828354>.
- [63] L. Moghaddasi, E. Bezak, L.G. Marcu, Current challenges in clinical target volume definition: tumour margins and microscopic extensions, *Acta Oncol. (Madr.)* 51 (2012) 984–995, <https://doi.org/10.3109/0284186X.2012.720381>.
- [64] B. Berthon, C. Marshall, a Edwards, et al., Influence of cold walls on PET image quantification and volume segmentation: a phantom study, *Med. Phys.* 40 (2013) 082505, <https://doi.org/10.1118/1.4813302>.
- [65] A. Stefano, F. Gallivanone, C.L. Messa, et al., Metabolic impact of Partial Volume Correction of [18F]FDG PET-CT oncological studies on the assessment of tumor response to treatment, *Q. J. Nucl. Med. Mol. Imaging* 58 (2014) 413–423.
- [66] A. Stefano, S. Vitabile, G. Russo, et al., A graph-based method for PET image segmentation in radiotherapy planning: a pilot study, in: A. Petrosino (Ed.), *Image Anal. Process*, Springer-Verlag Berlin, 2013, pp. 711–720.



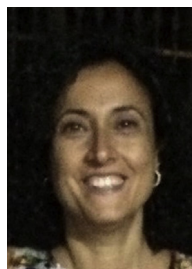
**Albert Comellir** is a Ph.D. student in Computer Engineering at the University of Palermo. He received combined BSc's/MSc's Degree in Computer Science in 2013. His research interests include medical image processing and medical data analysis.



**Alessandro Stefano** is a Research Scientist with the IBFM-CNR. He received the Laurea degree (summa cum laude) and the Ph.D. in computer engineering in 2005 and 2016, respectively. His research interests include medical image processing.



**Giorgio Russo** received the Laurea (summa cum laude) degree in physics (2003), and the medical physics qualification (2007). His research interests include the development of elaboration methods to biomedical imaging.



**Maria Gabriella Sabini** received the Laurea degree in physics in 1997, and the medical physics qualification in 1999. He is currently the head of Medical Physics Unit at Cannizzaro Hospital.



**Massimo Ippolito** is the head of Department of Nuclear Medicine at Cannizzaro Hospital. He received the Laurea degree (summa cum laude) in medicine and Surgery at the University of Catania in 1992.



**Samuel Bignardi** obtained both his M.S. degree in Physics, and his Ph.D. degree in Earth Sciences (Geophysics). Since 2017 he has been working as a Research Scientist in the School of Electrical and Computer Engineering at the Georgia Institute of Technology. His research interests include geophysics, applied physics, active surfaces, medical imaging, and numerical computing.



**Giovanni Petrucci** is Full Professor in Machine Design at University of Palermo from November 2011 to date. The main research topics are automated methods for the optical techniques of stress analysis, methods of random and multiaxial fatigue analysis of materials and mechanical components, and methods for the experimental evaluation of residual stresses.



**Anthony Yezzi** holds the position of Julian Hightower Chair Professor within the School of Electrical and Computer Engineering at Georgia Institute of Technology where he directs the Laboratory for Computational Computer Vision. He has over twenty years of research experience in shape optimization via geometric partial differential equations. He has consulted for a number of companies including GE, 3M, MZA, Philips, Picker, and VTI.