

An enhanced random walk algorithm for delineation of head and neck cancers in PET studies

Alessandro Stefano^{1,2}  · Salvatore Vitabile³ · Giorgio Russo^{1,4} · Massimo Ippolito⁵ · Maria Gabriella Sabini⁴ · Daniele Sardina⁴ · Orazio Gambino² · Roberto Pirrone² · Edoardo Ardizzone² · Maria Carla Gilardi¹

Abstract An algorithm for delineating complex head and neck cancers in positron emission tomography (PET) images is presented in this article. An enhanced random walk (RW) algorithm with automatic seed detection is proposed and used to make the segmentation process feasible in the event of inhomogeneous lesions with bifurcations. In addition, an adaptive probability threshold and a k-means based clustering technique have been integrated in the proposed enhanced RW algorithm. The new threshold is capable of following the intensity changes between adjacent slices along the whole cancer volume, leading to an operator-independent algorithm. Validation experiments were first conducted on phantom studies: High Dice similarity coefficients, high true positive volume fractions, and low Hausdorff distance confirm the accuracy of the proposed method. Subsequently, forty head and neck lesions were segmented in order to evaluate the clinical feasibility of the proposed approach against the most common segmentation algorithms. Experimental results show that the proposed algorithm is more accurate and robust than the most common algorithms in the literature. Finally, the proposed

method also shows real-time performance, addressing the physician's requirements in a radiotherapy environment.

Keywords Head and neck cancer segmentation · Random walks · PET imaging · Biological target volume

1 Introduction

Recent advances in radiotherapy have improved the clinical effectiveness of radiation treatment planning (RTP) delivering a high radiation dose to the target and maintaining a low radiation dose to nearby critical organs. However, hardware precision in the radiation dose delivery is greater than the software precision in target volume delineation. Accurate target volume definition is essential for escalating the radiation dose without increasing normal tissue injury especially in head and neck cancer (HNC). Computed tomography (CT) is considered the reference modality for target volume delineation in HNC, although cancers and the surrounding soft tissues show similar density. CT imaging may not show the viable extension of cancers and may not localize isolated positive lymph nodes [1, 2]. To improve these results and assist the radiation oncologist in RTP, positron emission tomography (PET) has been introduced to the radiotherapy field. PET is a non-invasive functional imaging technique giving complementary information with respect to anatomical imaging and providing an in vivo measurement of the cancer's biological processes. Moreover, metabolic changes are often faster and more indicative of the therapy effects than morphological changes, providing a more rapid method to detect the treatment response [3, 4]. Among several PET radiotracers derived from isotopes, the glucose analogue 18F-fluoro-2-deoxy-D-glucose (FDG) is widely used in the evaluation process of several

✉ Alessandro Stefano
alessandro.stefano@ibfm.cnr.it

¹ Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Cefalù, PA, Italy

² Department of Chemical, Management, Information Technology and Mechanical Engineering, University of Palermo, Palermo, Italy

³ Department of Biopathology and Medical Biotechnologies (DIBiMED), University of Palermo, Palermo, Italy

⁴ Medical Physics Unit, Cannizzaro Hospital, Catania, Italy

⁵ Nuclear Medicine Department, Cannizzaro Hospital, Catania, Italy

neoplastic pathologies. FDG uptake increases in high metabolic rate tissues, such as cancers. FDG PET is capable of identifying the location of many primary cancers and metastases, offering the opportunity to radically modify a patient's treatment or a RTP [5].

PET images have a lower spatial resolution than CT or MR (Magnetic Resonance) images. Anatomical imaging techniques are still needed to localize and characterize abnormal regions. PET high contrast images and anatomical high spatial resolution images can be fused in multimodal images: current generation PET/CT and, more recently, PET/MRI systems are able to differentiate between a normal and an abnormal tissue, even using metabolic information. The diagnostic accuracy of the combined systems has proven to be higher than each individual technique, with their complementary features. In radiotherapy, it is possible to delineate the biological tumor volume (BTV) on PET images inside or outside the anatomical gross tumor volume (GTV), defined by CT or MR images. As result, the BTV can be used for enhanced RTP in order to treat the cancer region more precisely [6]. On the other hand, PET segmentation is a critical task due to its lack of consistency in a cancer contour, its low image resolution, its relatively high level of noise, and the FDG uptake heterogeneity within a lesion. For the above reasons, the BTV has great size variability, since it depends on the algorithm used to delineate the PET images.

In the literature, various approaches have been presented [7–12]. The choice of a standard method is a very challenging and open issue [13, 14], since accurate lesion segmentation in PET imaging is essential for an accurate quantification of prognosis assessment and therapy response. Visual delineation is usually widely used because it is easily applicable but potentially inaccurate for its window level settings dependence and its intra- and inter-operator variability. An objective, robust, fast, accurate, operator, and scanner-independent segmentation method is thus mandatory to properly use the information provided by molecular imaging.

In a previous publication, a segmentation method based on random walks on graphs (RW) had been used on phantom studies to assess its accuracy with excellent results [15]. Unfortunately, in clinical practice with real patients, our previous method often fails because cancers in PET imaging now have complex shapes, such as lesions with bifurcations, and, unlike phantom spheres, images show inhomogeneous uptake regions. Figure 1 shows a bifurcated PET/CT cancer in an oncological patient with HNC: The volume of a metabolic lesion may evolve splitting it into several sub-lesions, merging them in a single lesion, or both. Thus, a complex shape requires an accurate and efficient segmentation method capable of following the lesion in its whole volume and shape.

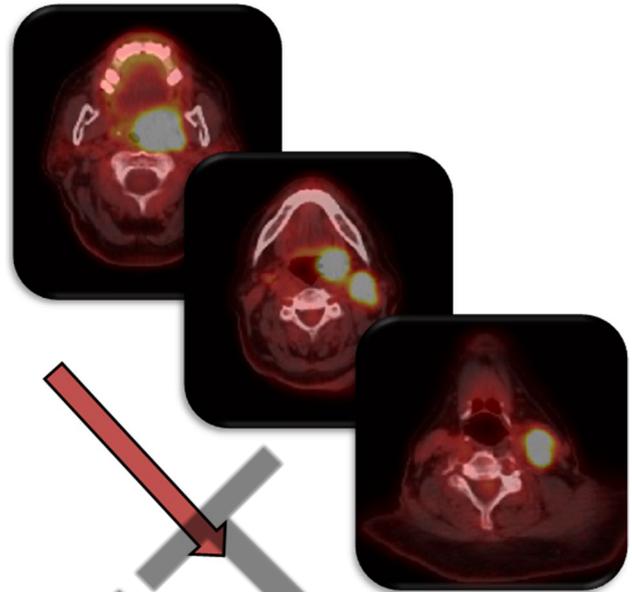


Fig. 1 A PET lesion that evolves splitting into two parts and merging in a single part along its volume

In this study, an enhanced RW algorithm is proposed. Unlike standard RW algorithm, the proposed enhanced algorithm includes a k-means clustering technique to select the target seeds along the whole cancer volume, resulting in a precise delineation of complex lesions. The achieved results are more accurate than the results produced by the standard RW algorithm. The above feature improves the BTV delineation accuracy in a RTP as well as the calculation of the total lesion glycolysis (TLG) and its fractional change, which is needed in the clinical practice for a treatment response evaluation [16]. Unlike the original RW method, an adaptive probability threshold has also been included to differentiate between a target and a background region. The adaptive probability threshold takes into account the intensity changes between adjacent PET slices along the metabolic volume, resulting in an operator-independent algorithm.

2 Methods

2.1 Phantom studies

National Electrical Manufacturers Association International Electrotechnical Commission (NEMA IEC) body phantom was used to estimate the accuracy of the PET segmentation algorithms. The phantom is composed of an elliptical cylinder ($D_1 = 24$ cm, $D_2 = 30$ cm, $h = 21$ cm) with six spheres of different diameters ($d_1 = 10$ mm, $d_2 = 13$ mm, $d_3 = 17$ mm, $d_4 = 22$ mm, $d_5 = 26$ mm, $d_6 = 37$ mm)

placed at 5.5 cm from the center of the phantom. Spheres and background were filled with FDG. Actual sphere and background radioactivity concentrations were measured using a dose calibrator system (Dose calibrator PET Dose, Comecer). Background radioactivity concentration ranged from 2 kBq/ml to 8 kBq/ml at the time of acquisition. The ratio between measured sphere and background radioactivity concentrations (S/B) ranged from 2 to 11 for 4 independent experiments. The proposed segmentation method was assessed by matching the sphere delineation with the ground truth in the CT images.

2.2 Clinical studies

The clinical feasibility of the proposed segmentation algorithm was assessed in oncological PET studies. Eighteen patients affected by HNC and subjected to diagnostic PET/CT scan before radiotherapy treatments were selected. The institutional Medical Ethics Review Board approved the study protocol, and all subjects signed a written informed consent form. Patients fasted for 12 h before the PET examination, and the FDG was administered. The PET/CT oncological protocol began 60 min after the FDG administration. 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. Two nuclear medicine experts of diagnostic and staging purposes reported these studies. The active tumor volume was manually defined by drawing a 2D outline slice by slice: BTV included the cancer volume with an intense tracer uptake with respect to background FDG activity level. The study is not a clinical trial but an observational study that did not influence the management of oncological patients.

2.3 Data acquisition

The acquisition of phantom experiments and clinical studies was performed using the Discovery 690 scanner with time of flight by General Electric Medical Systems. For each bed position, the PET image volume consisted of $256 \times 256 \times 47$ voxels of $2.73 \times 2.73 \times 3.27$ mm³ size, while the CT volume consisted of $512 \times 512 \times 47$ voxels of $1.36 \times 1.36 \times 3.75$ mm³ size. 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). Phantoms were acquired in two bed positions. PET images were reconstructed by a 3D ordered subset expectation maximization (OSEM) algorithm.

2.4 The random walk algorithm

Graph-based methods are used to perform segmentation of images. The graph cut algorithm [17] is a computationally

expensive algorithm, and it has no exact solution. In addition, it may return very small regions for images with low contrast or which are noisy. To solve this problem, known as “small cut”, many seed points must be placed. The RW algorithm was developed by Grady [18] and was extended for image segmentation [19]. Despite both of them being graph-based methods, they are actually quite different. Rather than considering the segmentation as a max-flow/min-cut problem, RW treats the segmentation as the solution of a linear system with an exact solution. In addition, RW is less susceptible to “small cut” behaviors than graph cut ones and is more efficient in terms of handling ambiguities among object boundaries. The PET image is converted into a graph where some voxels are known and others are unknown. The aim is to assign a label to unknown nodes. This is done by finding the minimum cost/energy among all possible scenarios in the graph to provide an optimal segmentation. The RW problem has the same solution as the combinatorial Dirichlet problem [18]. A threshold of 50 % is chosen to discriminate the foreground from the background so that a voxel binary mask can be created:

- target node value = 1 if its probability ≥ 50 %
- background node value = 0 if its probability < 50 %.

This threshold implies that any voxel 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 imposed by a Gaussian-like function:

$$w_{ij} = \exp\left(-\beta(g_i - g_j)^2\right) \quad (1)$$

where both g_i and g_j are the intensities of the voxels i and j , respectively; β is a free parameter depending on the user.

2.4.1 The random walk algorithm in PET imaging

The Gaussian weighting function for the PET image has been defined as:

$$w_{ij} = \exp\left(-\beta(\text{SUV}_i - \text{SUV}_j)^2\right) \quad (2)$$

to incorporate metabolic information in the RW algorithm. The SUV is the standardized uptake value (SUV), the most common semi-quantitative parameter used to estimate FDG accumulation within a lesion in clinical practice. The SUV 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} value in accordance with Eq. 2. Due to the partial volume effect (PVE), the separation of target and background voxels is very difficult. To reduce this

effect, voxels of similar intensity have been grouped with the probability likelihood for each cluster (target and background) to the original RW algorithm as proposed in [20]. Fuzzy c-means algorithm is used to identify the target and background clusters.

2.5 The enhanced version of the random walk algorithm

The RW method is very sensitive to the choice of β factor in the Gaussian-like weighting function in Eq. 2. β influences how quickly the probability decreases with increasing intensity differences: A high β value reduces the weight of walker, which weakens the connection between the adjacent voxels and underestimates the foreground volume. Vice versa, a low β value increases the weight, which overestimates the target volume. In [21], the authors improved RW robustness in PET imaging, but they did not deal with the dependence of this parameter. In [20], the authors take into account this limitation using the Euclidean distance between adjacent nodes. Nevertheless, the authors do not consider the issue of complex lesions, which may split into several parts or merge into a single part or both of these, like HNC in which the Euclidean distance might not be an optimal solution. The Euclidean distance does not provide any information regarding the number of hot areas in the bifurcated lesion, like the one shown in Fig. 1. The proposed approach provides an enhanced version of the original RW method to automatically detect foreground/background seeds including k-means clustering to make the BTV delineation process feasible in complex and heterogeneous lesions, like bifurcated ones. K-means is implicitly based on Euclidean distances and, in addition, is able to follow the evolution of the target in the whole volume identifying centroids of hot regions.

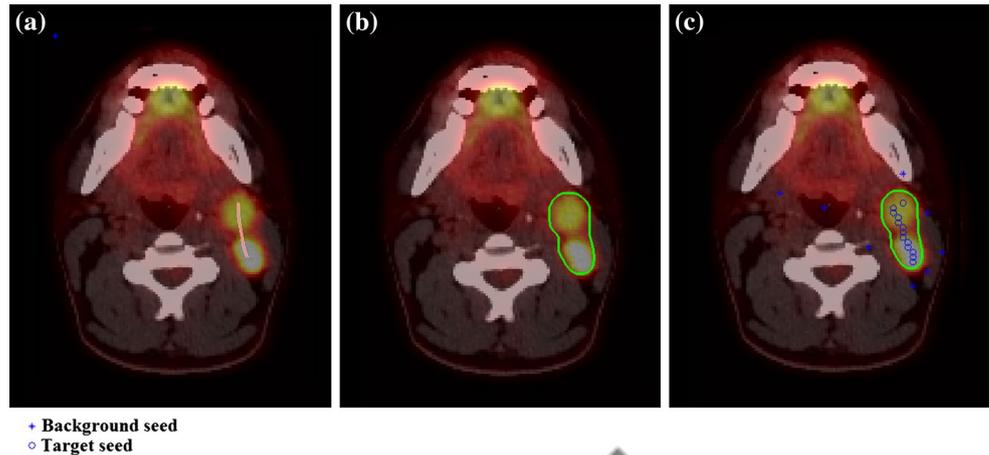
In addition, the proposed approach automatically computes the probability threshold for each slice to discriminate target voxels from background ones to obtain the final cancer segmentation. In the original RW, the final binary delineation is obtained using a fixed threshold value of 50 %. In the proposed method, the adaptive probability threshold of each slice is computed separately, taking into account the intensity gradient and contrast changes of the metabolic lesion over the volume. Finally, in phantom studies emulating clinical conditions, the RW with $\beta = 1$ provides higher Dice similarity coefficients (DSC) than other β values (0.5, 0.7, 0.9, and 2). Under the proposed context, the β factor is set to 1, as also proposed in [20], and the weights between nodes, necessary for the walk moving on the graph, are based just on SUVs (Eq. 2). The proposed algorithm retains all the properties of the original RW method and, in addition, uses an adaptive parameter to have a more robust performance.

2.5.1 The K-RW algorithm: the RW algorithm with K-means

An automatic user-friendly method to detect background and foreground seed points is proposed. The user draws a line on the coronal PET image along the target, and the axial slice (slice_{max}) with maximum SUV (SUV_{max}) is automatically identified and showed to the user that draws a new line along the lesion. This approach allows the cancer to be properly delineated, excluding false positives (normal structures like the brain, heart, bladder, and kidneys that normally have high FDG uptake). The algorithm can be broken down into two main steps: the pre-segmentation step to automatically detect the RW seeds and the segmentation step to delineate the final metabolic cancer. The initial target seeds are the voxels corresponding to the line drawn by the user (target seed line). The delineation method is achieved by the following steps:

1. The target seeds with a SUV less than 30 % (optimal threshold identified in phantom experiments, see “3.1 Trials and Results on Phantoms” section) of the SUV_{max} are removed to avoid any necrotic or background area.
2. The 8-neighborhood (north, south, west, east, and the 4 diagonal directions) of the voxel with SUV_{max} are explored to detect background seeds. The neighbor with a value less than 30 % of the SUV_{max} is identified. Those 8 voxels are marked as background seeds.
3. The RW delineation performs a “rough” 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 centers within the pre-segmented lesion. In a complex volume, a lesion can be divided into two or more areas with different hot peaks. This algorithm follows the evolution of the target in the whole volume, identifying centroids of hot regions. In the event of a homogenous target (such as a sphere in phantom studies), the algorithm returns a single centroid. The k value is automatically inferred; a more accurate description is proposed in the next Sect.
5. The centroids (one or more) and the voxels within the pre-segmented lesion with a SUV greater than 90 % (optimal threshold identified in phantom experiments, see “3.1 Trials and Results on Phantoms” section) of SUV_{max} are identified as new target seeds.
6. The RW algorithm performs the segmentation by utilizing the seeds identified in step 2 (background seeds) and 5 (target seeds).

Fig. 2 Pre-segmentation and segmentation steps in the slice_{max} are shown in (a–c). The user draws a red line on the lesion (a). Then, the RW algorithm performs a rough pre-segmentation step [green region of interest in (b)] to automatically detect the seeds used to perform the final segmentation step (c)



7. For the first slice (slice_{max}), the user can manually change the probability threshold, set by default to 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 the slice_{max} remains fixed for the whole volume.

This process is repeated for all the slices 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 subsequent slices until no segmentation or abnormal increment of target seeds is verified. An overview of the proposed seed localization method is outlined in Fig. 2.

2.5.1.1 K-means clustering for target seed detection The input of the k-means algorithm is the matrix containing the pre-segmented lesion in step A). It returns the coordinates of the k centroids of hot regions, where k is the number of clusters with which the data are segmented.

The k value is automatically inferred:

- $k = 2$ for the target seed line containing voxels with SUV greater than 30 % of the SUV_{max}. This check permits the exclusion of necrotic or background area. The two clusters represent cancer and background regions.
- $k = n+1$ for the start target seed line containing seeds with SUV less than 30 % of the SUV_{max}. These voxels belong to the background or necrotic region: The seed line passes from one “hot” region to another “hot” region (Fig. 3). $k = n + 1$ 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).



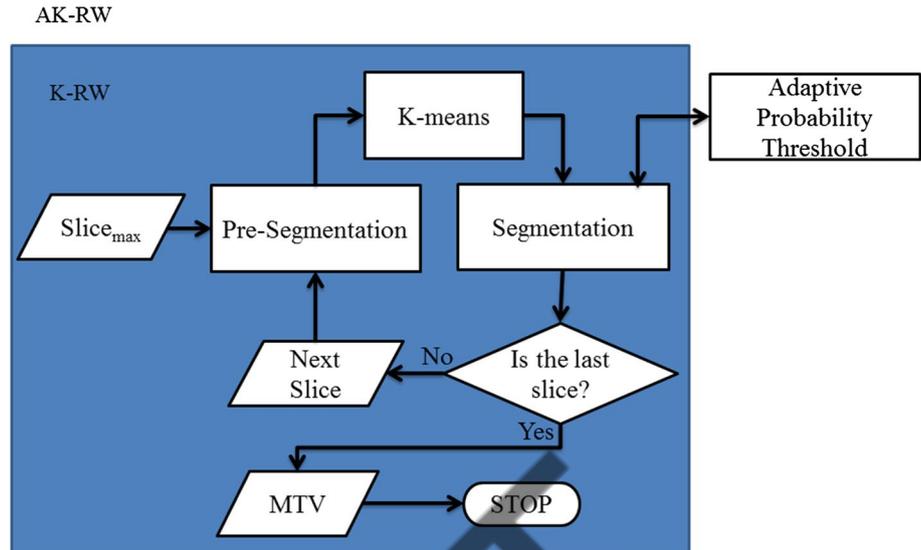
Fig. 3 Target seed line: Voxels with SUV greater than 30 % of the SUV_{mean} belong to the lesion (green segments). Voxels below the threshold belong to the background (blue segment). In this case, $k = 3$ (2 “hot” regions and the background region)

2.5.2 The AK-RW algorithm: the K-RW algorithm with adaptive probability threshold

A further extension of the K-RW method to automatically change the probability threshold for each slice to discriminate foreground from background voxels is proposed to take into account the intensity gradient and contrast changes of the lesion over the whole target volume.

The algorithm is the same as the previous one except for step 7 where the probability threshold is automatically inferred by the system for each slice (the probability threshold changes during volume delineation). The AK-RW method flowchart is shown in Fig. 4.

Fig. 4 AK-RW method flow-chart



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

1. Calculating the mean (M) of the probability values inside a large pre-segmented lesion obtained by using a default probability threshold of 80 % (optimal threshold identified in phantom experiments, see “3.1 Trials and Results on Phantoms” section).

Identification of two groups of voxels:

- Voxels with a probability $< M$.
- Voxels with a probability $\geq M$.
- Calculating the probability means (P1 and P2) of the two groups.
- The adaptive threshold value is then calculated as $P = \frac{1}{2} (P1 + P2)$.

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

2.6 Evaluation metrics

The segmentation performance of the proposed methods is evaluated by making a volumetric comparison with manual BTM segmentation using the DSC, median Hausdorff distance (HD), and true positive and false positive volume fractions (TPVF and FPVF). The DSC measures the spatial overlap between the manual and the automated segmentation volume: A DSC value equal to one indicates a perfect

match between the two volumetric segmentations, while a DSC whose value is zero indicates no overlap. HD is a shape dissimilarity metric measuring the most mismatched boundary voxels between automatic and manual BTM: A small median of HD values means an accurate segmentation, while a large median of HD values means no accuracy. TPVF concerns the fraction of the total amount of tissue inside the target lesion (sensitivity), and FPVF denotes the amount of tissue falsely identified (specificity = $100 - FPVF$) [22]. 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). The average time employed to delineate targets is recorded to evaluate algorithm performance.

The inter-operator agreements between the two radiation oncologist segmentations are analyzed by DSC overlap ratios. The patient studies are used to assess the applicability of the proposed algorithms in a clinical environment.

2.7 Comparison against other methods

The performance of the proposed methods (K-RW and AK-RW) is compared to the PET image segmentation methods commonly used in the clinical environment. In particular, K-RW and AK-RW algorithms, the original RW method, thresholding method (40 %), and region growing method [23, 24] have been implemented. For this purpose, a software package to provide a segmentation task tool has been implemented on the MATLAB R2014 simulation environment, running on a general purpose PC with a 3.00 GHz Intel R Core™ i5-2320 processor, 4 GB memory, and 64-bit Windows 7 Professional OS.

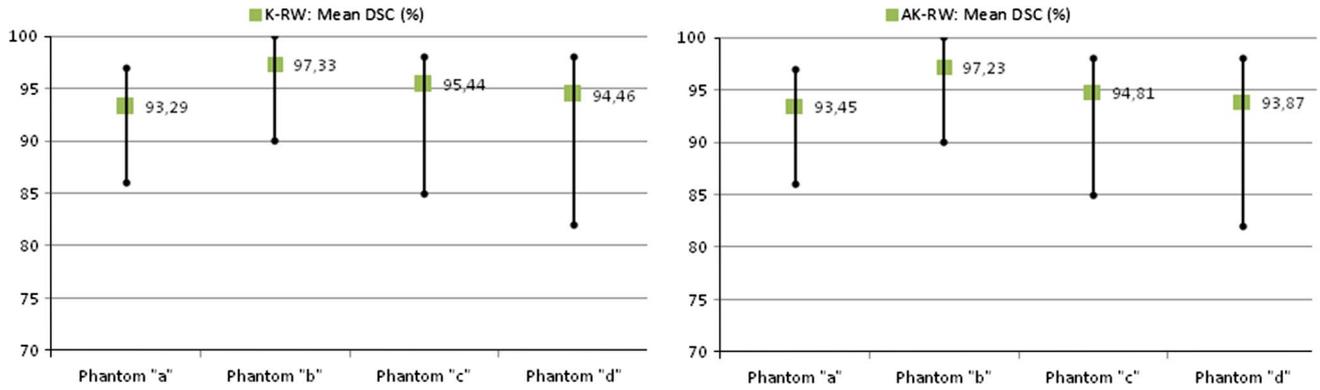


Fig. 5 Mean DSC obtained from NEMA IEC body phantoms. Phantoms have a measured S/B of 2–3 for phantom “a”, 3–5 for phantom “b”, 5–6 for phantom “c”, and 6–7 for phantom “d”

3 Results

3.1 Trials and results on phantoms

The delineation method accuracy was assessed using spheres of known volumes.

In Eq. 2, $\beta = 1$ provided the highest DSCs among all the tested β -values (0.5, 0.7, 0.9, 1, and 2).

30 % and 90 % are the SUV_{max} threshold values required to identify the background and target seeds (see Sect. 2.5.1) that minimize the differences between CT and PET volumes using DSC measure. The threshold values ranged from 10 % to 40 % and from 70 % to 95 % for background and target seeds, respectively, with a step size of 5 % in both cases. In the same way, the adaptive probability threshold value of 80 % (see Sect. 2.5.2) produced the highest DSC (threshold range: 10 % ÷ 95 %, step size of 5 %).

Figure 5 shows the mean DSC values in four independent phantom experiments carried out at different S/B: 2–3 for the phantom “a”, 3–5 for the phantom “b”, 5–6 for the phantom “c”, and 6–7 for the phantom “d”.

The K-RW method showed a DSC range from 83.51 % (phantom “d”) up to 99.86 % (phantom “a”) for the spheres with a diameter less than 17 mm ($DSC = 92.95 \pm 5.90$ %), and from 87.71 % (phantom “a”) up to 99.43 % (phantom “d”) for the spheres with a diameter exceeding 17 mm ($DSC = 96.17 \pm 3.48$ %). Considering all spheres, phantom “b” had the best mean DSC (97.33 ± 1.87 %) and phantom “a” the worst one (93.29 ± 5.44 %). HD ranged from 1.27 mm (phantom “c”) up to 2.71 mm (phantom “b”) for the smaller spheres ($HD = 1.72 \pm 0.28$ mm), and from 0.1 mm (phantom “a”) up to 2.73 mm (phantom “b”) for the larger spheres ($HD = 0.8 \pm 0.13$ mm). The mean TPVF was 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 AK-RW method showed a DSC range from 80.92 % (phantom “c”) up to 99.98 % (phantom “b”) for the smaller spheres ($DSC = 89.55 \pm 7.16$ %), and from 93.42 % (phantom “a”) up to 99.39 % (phantom “b”) for the larger spheres ($DSC = 97.32 \pm 1.73$ %). Considering all spheres, phantom “b” had the best mean DSC (97.23 ± 3.56 %) and phantom “a” the worst one (93.45 ± 4.77 %). HD ranged from 1.35 mm (phantom “c”) up to 2.81 mm (phantom “b”) for the smaller spheres ($HD = 1.81 \pm 0.19$ mm), and from 0.1 mm (phantom “a”) up to 2.73 mm (phantom “b”) for the larger spheres ($HD = 0.9 \pm 0.11$ mm). The mean TPVF was 93.08 ± 9.66 % for phantom “a”, 98.37 ± 2.13 % for phantom “b”, 97.30 ± 1.11 % for phantom “c”, and 98.74 ± 2.32 % for phantom “d”.

The mean specificity ($100 - FPVF$) was ~ 100 % for all experiments and algorithms.

High DSC and TPVF, and low HD and FPVF values confirm the accuracy of the K-RW and AK-RW methods. The AK-RW algorithm is slightly less accurate than K-RW. This finding had been foreseen because K-RW is a semi-supervised algorithm: The user can choose the best of the probability threshold values to properly delineate the PET spheres. However, user-independent techniques achieving a good segmentation, such as AK-RW method, are crucial in a clinical environment. In addition, the AK-RW method follows the whole lesion volume, taking into account the changes in both intensity gradient and contrast of the PET lesion in different slices. This is a key feature in clinical studies.

An analysis of the time performance showed that both algorithms are fast: The segmentation time for larger spheres was around 4 s. Obviously, in K-RW delineation, the time needed for the user choice of the probability threshold was excluded.

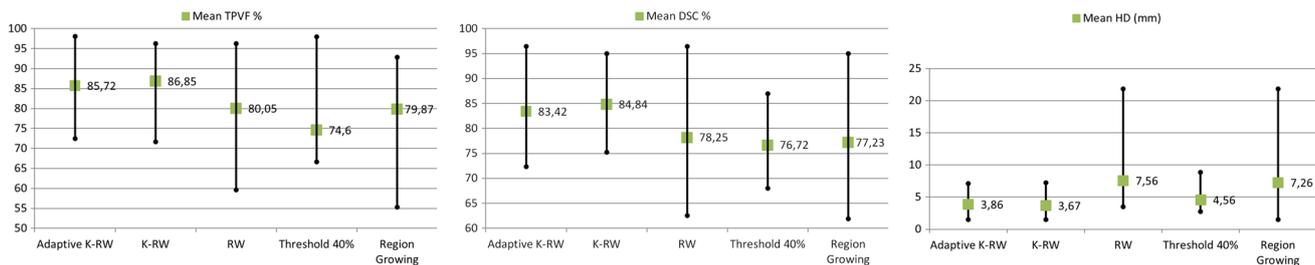


Fig. 6 Results based on 40 lesions in 18 patients, for each segmentation algorithm, are shown

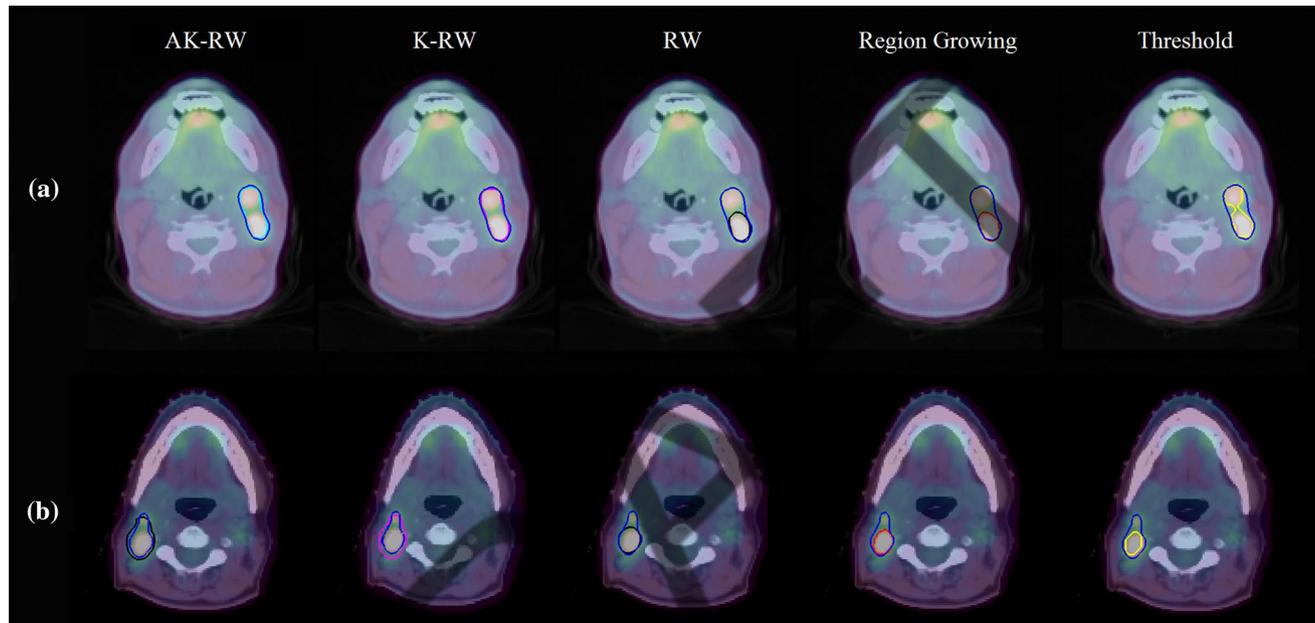


Fig. 7 Segmentation examples of two bifurcated lesions (A and B rows) are shown. AK-RW method (cyan in the first column), K-RW method (magenta in the second column), RW method (black in the

third column), region growing method (red in the fourth column), and threshold method (yellow in the fifth column) are overlaid with manual segmentation (blue in all columns)

3.2 Trials and results on patient studies

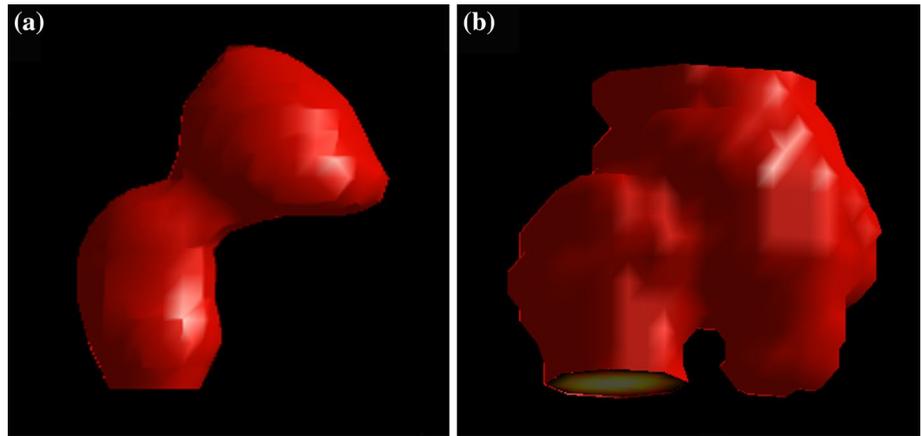
Manual delineation was obtained by averaging the segmentations performed by two nuclear medicine physicians with an inter-observer agreement of $86.51 \pm 3.65\%$.

Figure 6 reports the quantitative comparison between semi-automatic and manual segmentation. FPVF is very low for all algorithms since there are a lot fewer target voxels than background ones; a single PET slice consists of 65,536 voxels while the largest lesion is less than 180 voxels in a single PET slice. Results based on the DSC, HD, and TPVF values show that the K-RW and AK-RW algorithms outperform the best algorithms taken into account for comparison.

In addition, region growing and RW methods often failed to properly delineate bifurcated lesions: Fig. 7 shows the segmentation task of two lesions with a complex shape

that was 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 to delineate 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. In this way, false positives are removed, but the segmentation time increases considering the need to delineate the VOI. However, the proposed methods are able to follow the bifurcation by 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 lesion volume. The volumes of two segmented lesions are

Fig. 8 3D segmentation examples using AK-RW algorithm are shown



shown in Fig. 8: The AK-RW algorithm is efficient to properly follow the whole cancer volume.

4 Discussion

Qualitative visual interpretation of PET studies is the most commonly used method in clinical environment. The manual segmentation method depends on the experience of the nuclear physician, limiting the measurement accuracy. Due to the dependency of both operator experience and display window level settings, the process is time-consuming and affected by inter- and intra-observer variability. To reduce these issues, several automatic methods have been presented in literature, although few clinical studies are available, and there is no consensus for proper BTV determination.

In radiotherapy, CT imaging is considered the standard approach for target volume delineation in HNC. On the other hand, CT imaging does not show cancer biological features. For this reason, PET has been introduced to assist radiation oncologists in clinical routine.

In phantom studies, the CT region matches the related PET region since the radiotracer is contained in the CT visible sphere. This is not confirmed in patient studies due their complexity and variability. Head and neck lesions may have different PET margins when compared to anatomical margins [15]. The metabolic volume cannot match the cancer anatomic extension, showing different and additional information, like for CT invisible metastases and cancer extensions [25, 26]. It is not appropriate to consider a one-to-one relationship between anatomical and functional images. In addition, misregistration between the two series can occur due to a patient's motion artifacts [27]. The assumption of identical boundary in PET and CT images is questionable with special reference to the HNC region [25–27]. For these reasons, we have independently extracted the BTV from anatomical imaging, although many studies use

co-registered CT images information to identify features and distinguish a lesion from the background and, consequently, for PET image segmentation [21, 28, 29]. Finally, a whole automatic detection method cannot be implemented to identify oncological lesions in whole-body PET scans, since healthy organs like the brain, heart, bladder, and kidneys normally have a high FDG uptake. As a result, user interaction is mandatory.

In this study, we optimized the performance of an existing semi-automatic segmentation algorithm based on Grady's RW formulation [18]. The key strategies include a k-means clustering algorithm to obtain refined target seed locations within pre-segmented lesions and a strategy to adaptively select the optimum threshold value to be applied on the RW probabilistic output, in order to obtain the final cancer segmentation.

Initially, the RW algorithm with an embedded k-means algorithm to identify hot region centroids has been proposed to obtain optimized segmentation results with complex lesions (see Fig. 1). Unlike the fixed 50 % threshold value of the original RW method, the user can manually change the probability value to discriminate between target and background voxels in order to select the value optimizing the segmentation results. We call this method "K-RW".

Subsequently, an extension of the K-RW method has been developed to adaptively determine the probability threshold for discriminating between cancer and background voxels. We call this method "Adaptive K-RW" (AK-RW). The two methods are able to deal with PET image segmentation, speeding-up considerably when compared with the time needed for manual segmentation.

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

small cancer volume [30]. The PVE for smaller objects is one of the most important factors impacting the qualitative and the quantitative accuracy in PET imaging [31]. The images are blurred due to the limited spatial resolution of PET scanners and small lesions appear larger. For this reason, the method described in [20] has been used in the proposed approach.

The AK-RW method is slightly less accurate than the supervised K-RW method, but this is to be expected because of the automatic selection of probability threshold value. However, AK-RW achieves good segmentation results with the benefit of requiring a lower user interaction effort and lower levels of the user's specialist knowledge than the first method. In addition, AK-RW does not depend on the choice of the probability threshold value to discriminate between target and background regions. The development of user-independent techniques capable of performing a good segmentation step is crucial in a clinical environment.

Clinical studies show that the proposed methods provide better results in minimizing the difference between manual and automated segmentation than the other state-of-the-art methods. K-RW and AK-RW methods are able to deal with complex volume delineation, unlike the other ones that have shown an acceptable delineation under specific conditions, such as homogeneous uptake concentration. Nevertheless, lesions in PET studies can have complex and bifurcated shapes and inhomogeneous uptake concentration. In these cases, literature methods fail in BTV delineation. The proposed study takes into account these issues to prevent potential disease progression, in accordance with the accuracy required in the radiotherapy environment.

5 Conclusions

An enhanced RW algorithm embedding both a k-means clustering algorithm and an adaptive probability threshold (AK-RW) has been described in this paper. The new AK-RW maintains all the properties of the original RW algorithm, but it is capable of selecting refined target seed locations for initializing the RW algorithm. AK-RW is also able to deal with intensity gradient and contrast changes of complex, bifurcated and inhomogeneous lesions over the whole target volume. The proposed method is very powerful in terms of PET image segmentation accuracy and time performance. It may be used as a Medical Decision Support System to enhance the current daily methodology performed by healthcare operators in radiotherapy treatments.

Acknowledgments This work was partially supported by CIPE1 (n. DM45602).

References

1. Lauve A, Morris M, Schmidt-Ullrich R et al (2004) Simultaneous integrated boost intensity-modulated radiotherapy for locally advanced head-and-neck squamous cell carcinomas: II—clinical results. *Int J Radiat Oncol Biol Phys*. doi:[10.1016/j.ijrobp.2004.03.010](https://doi.org/10.1016/j.ijrobp.2004.03.010)
2. Kim Y, Tomé WA (2007) On the radiobiological impact of metal artifacts in head-and-neck IMRT in terms of tumor control probability (TCP) and normal tissue complication probability (NTCP). *Med Biol Eng Comput* 45:1045–1051. doi:[10.1007/s11517-007-0196-8](https://doi.org/10.1007/s11517-007-0196-8)
3. Wahl RL, Jacene H, Kasamon Y, Lodge MA (2009) From RECIST to PERCIST: evolving Considerations for PET response criteria in solid tumors. *J Nucl Med* 50(Suppl 1):122S–150S. doi:[10.2967/jnumed.108.057307](https://doi.org/10.2967/jnumed.108.057307)
4. Stefano A, Russo G, Ippolito M et al (2016) Evaluation of erlotinib treatment response in non-small cell lung cancer using metabolic and anatomic criteria. *Q J Nucl Med Mol Imaging* 60(3):264–273
5. Newbold KL, Partridge M, Cook G et al (2008) Evaluation of the role of ^{18}F FDG-PET/CT in radiotherapy target definition in patients with head and neck cancer. *Acta Oncol (Madr)* 47:1229–1236
6. Ciernik IF, Dizendorf E, Baumert BG et al (2003) Radiation treatment planning with an integrated positron emission and computer tomography (PET/CT): a feasibility study. *Int J Radiat Oncol* 57:853–863. doi:[10.1016/S0360-3016\(03\)00346-8](https://doi.org/10.1016/S0360-3016(03)00346-8)
7. Belhassen S, Zaidi H (2010) A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET. *Med Phys* 37:1309–1324. doi:[10.1118/1.3301610](https://doi.org/10.1118/1.3301610)
8. Li H, Thorstad WL, Biehl KJ et al (2008) A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours. *Med Phys* 35:3711–3721. doi:[10.1118/1.2956713](https://doi.org/10.1118/1.2956713)
9. Geets X, Lee JA, Bol A et al (2007) A gradient-based method for segmenting FDG-PET images: methodology and validation. *Eur J Nucl Med Mol Imaging* 34:1427–1438. doi:[10.1007/s00259-006-0363-4](https://doi.org/10.1007/s00259-006-0363-4)
10. Wanet M, Lee JA, Weynand B 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. *Radiother Oncol* 98:117–125. doi:[10.1016/j.radonc.2010.10.006](https://doi.org/10.1016/j.radonc.2010.10.006)
11. Namiás R, D'Amato JP, Del Fresno M et al (2016) Multi-object segmentation framework using deformable models for medical imaging analysis. *Med Biol Eng Comput*. 54(8):1181–1192. doi:[10.1007/s11517-015-1387-3](https://doi.org/10.1007/s11517-015-1387-3)
12. Hatt M, Cheze Le Rest C, Albarghach N et al (2011) PET functional volume delineation: a robustness and repeatability study. *Eur J Nucl Med Mol Imaging* 38:663–672. doi:[10.1007/s00259-010-1688-6](https://doi.org/10.1007/s00259-010-1688-6)
13. Schinagl DAX, Vogel WV, Hoffmann AL 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. *Int J Radiat Oncol Biol Phys* 69:1282–1289. doi:[10.1016/j.ijrobp.2007.07.2333](https://doi.org/10.1016/j.ijrobp.2007.07.2333)
14. Zaidi H, El Naqa I (2010) PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. *Eur J Nucl Med Mol Imaging* 37:2165–2187. doi:[10.1007/s00259-010-1423-3](https://doi.org/10.1007/s00259-010-1423-3)
15. Stefano A, Vitabile S, Russo G et al (2013) A Graph-Based Method for PET Image Segmentation in Radiotherapy Planning: a Pilot Study. *Lect Notes Comput Sci* 8157:711–720. doi:[10.1007/978-3-642-41184-7_72](https://doi.org/10.1007/978-3-642-41184-7_72)

16. Larson SM, Erdi Y, Akhurst T 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. *Clin Positron Imaging* 2:159–171. doi:[10.1016/S1095-0397\(99\)00016-3](https://doi.org/10.1016/S1095-0397(99)00016-3)
17. Boykov Y, Veksler O, Zabih R (2001) Fast approximate energy minimization via graph cuts. *Pattern Anal Mach Intell IEEE Trans* 23:1222–1239. doi:[10.1109/34.969114](https://doi.org/10.1109/34.969114)
18. Grady L (2006) Random Walks for Image Segmentation. *IEEE Trans Pattern Anal Mach Intell* 28:1768–1783
19. Bagci U, Yao J, Caban J et al (2011) A Graph-Theoretic Approach for Segmentation of PET Images. *Conf Proc IEEE Eng Med Biol Soc* 2011:8479–8482. doi:[10.1109/IEMBS.2011.6092092](https://doi.org/10.1109/IEMBS.2011.6092092)
20. Onoma DP, Ruan S, Thureau S et al (2014) Segmentation of heterogeneous or small FDG PET positive tissue based on a 3D-locally adaptive random walk algorithm. *Comput Med Imaging Graph* 38:753–763. doi:[10.1016/j.compmedimag.2014.09.007](https://doi.org/10.1016/j.compmedimag.2014.09.007)
21. Bagci U, Udupa JK, Mendhiratta N 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. *Med Image Anal* 17:929–945. doi:[10.1016/j.media.2013.05.004](https://doi.org/10.1016/j.media.2013.05.004)
22. Udupa JK, Leblanc VR, Zhuge Y et al (2006) A framework for evaluating image segmentation algorithms. *Comput Med Imaging Graph* 30:75–87. doi:[10.1016/j.compmedimag.2005.12.001](https://doi.org/10.1016/j.compmedimag.2005.12.001)
23. Day E, Betler J, Parida D et al (2009) A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients. *Med Phys* 36:4349–4358. doi:[10.1118/1.3213099](https://doi.org/10.1118/1.3213099)
24. Rundo L, Militello C, Vitabile S, Casarino C, Russo G, Midiri M, Gilardi MC (2016) Combining Split-and-Merge and Multi-Seed Region Growing Algorithms for Uterine Fibroid Segmentation in MRgFUS Treatments. *Med Biol Eng Comput* 54(7):1071–1084. doi:[10.1007/s11517-015-1404-6](https://doi.org/10.1007/s11517-015-1404-6)
25. Troost EGC, Schinagel DAX, Bussink J et al (2010) Clinical evidence on PET–CT for radiation therapy planning in head and neck tumours. *Radiother Oncol* 96:328–334. doi:[10.1016/j.radonc.2010.07.017](https://doi.org/10.1016/j.radonc.2010.07.017)
26. Paulino AC, Koshy M, Howell R et al (2005) Comparison of CT- and FDG-PET-defined gross tumor volume in intensity-modulated radiotherapy for head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 61:1385–1392. doi:[10.1016/j.ijrobp.2004.08.037](https://doi.org/10.1016/j.ijrobp.2004.08.037)
27. Wang J, del Valle M, Goryawala M et al (2010) Computer-assisted quantification of lung tumors in respiratory gated PET/CT images: phantom study. *Med Biol Eng Comput* 48:49–58. doi:[10.1007/s11517-009-0549-6](https://doi.org/10.1007/s11517-009-0549-6)
28. Han D, Bayouth J, Song Q et al (2011) Globally optimal tumor segmentation in PET-CT images: a graph-based co-segmentation method. *Inf Process Med Imaging*. 22:245–256
29. Song Q, Bai J, Han D et al (2013) Optimal Co-segmentation of tumor in PET-CT images with context information. *IEEE Trans Med Imaging* 32:1685–1697. doi:[10.1109/TMI.2013.2263388](https://doi.org/10.1109/TMI.2013.2263388)
30. Stefano A, Gallivanone F, Messa C et al (2014) 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(4):413–423
31. Soret M, Bacharach SL, Buvat II (2007) Partial-volume effect in PET tumor imaging. *J Nucl Med* 48:932–945. doi:[10.2967/jnumed.106.035774](https://doi.org/10.2967/jnumed.106.035774)



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.



Salvatore Vitabile is an associate professor with the DIBI-MED, University of Palermo, Italy. He is the Editor-in-Chief of the International Journal of Adaptive and Innovative Systems, Inderscience Publishers.



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.



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.



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



Roberto Pirrone received his Master Degree in electronic engineering (1991). In 1995, he received the Ph.D. in computer and electronic engineering. From 2005, he became Associate Professor. His research interest is medical image analysis.



Daniele Sardina received the Laurea degree in physics in 2008 and the medical physics qualification in 2013. His research interests include the nuclear medicine imaging and the radiotherapy.



Edoardo Ardizzone is a Full Professor of computer systems at the University of Palermo, where he teaches image processing. His current research interests include image processing and analysis, medical imaging, and image restoration.



Orazio Gambino is graduated in computer science engineering at the University of Palermo. On 2008, he became Assistant Professor. His research interests include artefacts removal on digital and medical images, medical imaging segmentation.



Maria Carla Gilardi received laurea degree in Physics in 1980. She is Full Professor of Bioengineering at the University of Milano-Bicocca and Director of IBFM-CNR. Research interests in the field of medical imaging.