Research Article

An Adaptive Thresholding Method for BTV Estimation Incorporating PET Reconstruction Parameters: A Multicenter Study of the Robustness and the Reliability

M. Brambilla,[1] R. Matheoud,[1] C. Basile,[2] C. Bracco,[3] I. Castiglioni,[4] C. Cavedon,[5] M. Cremonesi,[6] S. Morzenti,[7] F. Fioroni,[8] M. Giri,[5] F. Botta,[6] F. Gallivanone,[4] E. Grassi,[8] M. Pacilio,[2] E. De Ponti,[7] M. Stasi,[3] S. Pasetto,[9] S. Valzano,[1] and D. Zanni[9]

[1]*Department of Medical Physics, University Hospital Maggiore della Carità, 28100 Novara, Italy*
[2]*Department of Medical Physics, Hospital S. Camillo Forlanini, 00152 Roma, Italy*
[3]*Department of Medical Physics, Institute for Cancer Research and Treatment (IRCC), 10060 Candiolo, Italy*
[4]*Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), 20093 Milan, Italy*
[5]*Department of Medical Physics, University Hospital, 37126 Verona, Italy*
[6]*Department of Nuclear Medicine, European Institute of Oncology, 20141 Milan, Italy*
[7]*Department of Medical Physics, Hospital San Gerardo, 20900 Monza, Italy*
[8]*Department of Medical Physics, Arcispedale S.Maria Nuova, IRCCS, 42123 Reggio Emilia, Italy*
[9]*Department of Medical Physics, Hospital Niguarda, 20162 Milan, Italy*

Correspondence should be addressed to I. Castiglioni; castiglioni.isabella@hsr.it

Objective. The aim of this work was to assess robustness and reliability of an adaptive thresholding algorithm for the biological target volume estimation incorporating reconstruction parameters. *Method*. In a multicenter study, a phantom with spheres of different diameters (6.5–57.4 mm) was filled with ^{18}F-FDG at different target-to-background ratios (TBR: 2.5–70) and scanned for different acquisition periods (2–5 min). Image reconstruction algorithms were used varying number of iterations and postreconstruction transaxial smoothing. Optimal thresholds (TS) for volume estimation were determined as percentage of the maximum intensity in the cross section area of the spheres. Multiple regression techniques were used to identify relevant predictors of TS. *Results*. The goodness of the model fit was high (R^2: 0.74–0.92). TBR was the most significant predictor of TS. For all scanners, except the Gemini scanners, FWHM was an independent predictor of TS. Significant differences were observed between scanners of different models, but not between different scanners of the same model. The shrinkage on cross validation was small and indicative of excellent reliability of model estimation. *Conclusions*. Incorporation of postreconstruction filtering FWHM in an adaptive thresholding algorithm for the BTV estimation allows obtaining a robust and reliable method to be applied to a variety of different scanners, without scanner-specific individual calibration.

1. Introduction

In the last years the coregistration of ^{18}F-fluorodeoxyglucose positron emission tomography (^{18}F-FDG PET) images with computed tomography (CT) images has gained an increasing interest in the staging and treatment planning for radiotherapy of several tumor sites. However, a standardized way of converting PET signals into target volumes is not yet available [1].

New semiautomatic and automatic segmentation methods have been developed implying gradient, region growing, clustering, statistical methods and other approaches [2–9]. While referring to these promising methods, it should be pointed out that some of these new methods suffer from the need of extensive preprocessing of the images (e.g., edge-detection). Moreover, the majority of these new algorithms are not widely available, so their use is currently restricted to

Sphere ID (mm)	Sphere A (mm^2)	Sphere volume (mL)
6.5	33.2	0.14
8.1	51.5	0.28
10	79	0.52
13	133	1.15
17	227	2.57
22	383	5.57
28	615	11.49
37	1075	26.52
57.4	2570	99.02

FIGURE 1: Inserts in the IEC phantom used for the multicenter measurements comprising 9 fillable spheres of different diameters.

the developers and, as a consequence, they are not independently validated.

Apart from visual inspection of PET scans, which suffers from interobserver variability [10], thresholding methods are widely used as PET segmentation approach in clinical practice for biological target volume (BTV) delineation for radiotherapy planning. Adaptive thresholding methods based on contrast-oriented contouring algorithms have been developed independently by many groups and validated in patient data in head and neck, in lung cancer, and in lymph nodes [11–13]. These methods are based on phantom measurements to derive a relationship between the "true" volume and the threshold to be applied to PET images.

These threshold-volume curves for one PET/CT scanner have been previously obtained varying target-to-background ratio (TBR), target dimensions, and postreconstruction smoothing [14]. It has been previously demonstrated that the emission scan duration (ESD) and background activity concentration, related to the level of image noise, are not predictors of the thresholding level of PET images [15]. Moreover, adaptive-threshold segmentation algorithms are not influenced by the different conditions of attenuation and scatter which may be encountered in different anatomical districts [16] and by the degree of convergence of iterative reconstruction algorithms [14].

Although adaptive thresholding methods are applicable to every PET scanner, it is generally assumed that the values of the parameters obtained during model building are system dependent so that a specific calibration for each PET system is required. On the other hand, a less hardware-dependent solution to the problem of PET segmentation could provide a robust algorithm, easily usable with images acquired by different scanner models without needing any previous optimization of the individual image quality.

Our hypothesis is that this goal can be accomplished by incorporating in our algorithm the reconstruction parameters that impact on threshold determination. To validate this hypothesis we firstly developed an original method to adapt the thresholds (TS) used to estimate the BTV in PET images. The proposed method incorporates the PET reconstruction parameters that influence the threshold determination. Secondly, we investigated in a multicenter

trial the robustness of this method with respect to various scanner models, reconstruction settings, and acquisition conditions: a multivariable approach was adopted to study the dependence of the TS that define the boundaries of ^{18}F-FDG uptake on object characteristics (contrast, size), acquisition parameters (scan duration), and reconstruction modalities (reconstruction algorithm, number of iterations, and amount of postreconstruction smoothing) in eleven state-of-the-art PET/CT scanners installed in eight different institutions. Finally, we assessed the reliability of the regression models through the use of split-sample analysis.

2. Materials and Methods

2.1. Phantoms. Measurements were performed on the NEMA IEC Body Phantom Set (Data Spectrum Corporation, Hillsborough, NC). This phantom contains 6 coplanar spheres, with internal diameters (ID) of 10, 13, 17, 22, 28, and 37 mm. A supplemental set of 2 microhollow spheres of 6.5 and 8.1 mm ID and 1 sphere of 57.4 mm ID were positioned at the bottom of the phantom. The experimental setup is depicted in Figure 1, together with sphere IDs (mm), maximum cross section areas (A) (mm^2), and volumes (mL). The same positioning of the phantom was ensured through laser localizer and a scout CT acquisition.

2.2. PET/CT Scanners. Eleven PET/CT scanners were used for the robustness study: n. 2 Discovery ST (S1, S2) [17], n. 1 Discovery STE (S3) [18], n. 2 Discovery 600 (S4, S5) [19], and n. 1 Discovery 690 (S6) [20] (GE Healthcare, Milwaukee, WI), n. 1 Biograph HI-REZ (S7) [21] and n. 1 Biograph TRUEV (S8) [22] (SIEMENS Medical Solutions, Knoxville, TN), n. 1 Gemini XL (S9) and n. 2 Gemini TF (S10, S11) [23] (Philips Medical Systems, Cleveland, OH). The technical characteristics and physical performances of the PET/CT scanners were derived from factory data and/or previous publications and are reported in Table 1.

2.3. Phantom Acquisition. The background of the IEC phantom was filled with 3 kBq/mL activity concentration of ^{18}F-FDG. A standard protocol was designed to generate the following acquisitions for each scanner model.

TABLE 1: PET/CT scanners: main technical characteristics and physical performances.

	Discovery ST (general electric) S1-2	Discovery STE (general electric) S3	Discovery 600 (general electric) S4-5	Discovery 690 (general electric) S6	Biograph16 Hi-REZ (Siemens) S7	Biograph 6 True V (Siemens) S8	Gemini XL (Philips) S9	Gemini TF (Philips) S10
Detector ring diameter (cm)	88.6	88.6	80.1	80.1	83.0	83.0	88.5	90.3
Detector material	BGO	BGO	BGO	LYSO	LSO	LSO	GSO	LYSO
Acquisition mode	2D/3D	2D/3D	3D	3D	3D	3D	3D	3D
Number of individual crystals	10.080	13.440	12.288	13824	24.336	32448	17.864	28.336
Number of crystal/ring	420	560	512	576	624	624	616	NA
Number of image planes	47	47	47	47	81	109	90	90
Crystal size (mm^3)	$6.3 \times 6.3 \times 30$	$4.7 \times 6.3 \times 30$	$4.7 \times 6.3 \times 30$	$4.2 \times 6.3 \times 25$	$4 \times 4 \times 20$	$4 \times 4 \times 20$	$4 \times 6 \times 30$	$4 \times 4 \times 22$
Patient port diameter (cm)	70	70	70	70	70	70	70	71.7
Axial field of view (cm)	15.7	15.7	15.7	15.7	16.2	21.8	18.0	18.0
Transaxial filed view (cm)	70	70	70	70	58.5	60.4	57.6	57.6
Axial sampling interval (mm)	3.27	3.27	3.27	3.27	2.0	2.0	2.0	2.0
Coincidence window width (ns)	11.7	9.3	9.0	4.9	4.5	4.5	7.5	6.0
Lower energy threshold (keV)	375	425	425	425	425	425	410	440
Physical performances								
Transverse resolution								
FWHM (mm) at 1 cm	6.29	5.1	4.9	4.70	4.61	4.1	5.2	4.8
FWHM (mm) at 10 cm	6.82	5.7	5.6	5.06	5.34	4.8	5.8	5.0
Axial resolution								
FWHM (mm) at 1 cm	5.68	5.2	5.6	4.74	5.10	4.7	5.8	4.8
FWHM (mm) at 10 cm	6.05	5.9	6.4	5.55	5.93	5.7	6.6	5.2
System sensitivity (cps/KBq)	8.99	8.8	9.6	7.5	4.87	8.0	8.0	6.6
Scatter fraction (%)	45	34	36.6	37	34.1	32.7	35	27

(a) Nine different TBRs (2.5 : 1, 4 : 1, 8 : 1, 16 : 1, 25 : 1, 35 : 1, 47 : 1, 55 : 1, and 70 : 1), determined by the dose calibrator and dilution, were imaged in different acquisition sessions. The measured TBRs were determined in the reconstructed image as the maximum pixel intensity in a region of interest (ROI) encircling the cross sectional area of the target, divided by the average pixel intensity of ROIs surrounding the sphere. These TBRs ranged from 70 down to 2.5 and were within the full range observed in patients.

(b) Four different ESD (2, 3, 4, and 5 min) were acquired to provide independent replicates of the experiments.

2.4. PET Image Reconstruction

2.4.1. Discovery ST, Biograph HI-REZ, and Biograph TRUEV. These systems use a 2D Fourier-rebinning (FORE) ordered subset expectation maximization (OSEM) algorithm with all corrections (scatter, random, dead time, attenuation, and normalization) incorporated into the iterative reconstruction scheme. In these systems the user can independently specify the number of iterations and subsets and the amount of the transaxial postreconstruction Gaussian smoothing, through the filter full-width-at-half-maximum (FWHM) expressed in mm.

2.4.2. Discovery STE and Discovery 600. The D-600 system uses a fully 3D-OSEM algorithm with all corrections incorporated into the iterative reconstruction scheme. The reconstruction settings are the same as above with the only difference that the axial filter is a mean filter with available kernels of 1 : 2 : 1, 1 : 4 : 1, and 1 : 6 : 1.

2.4.3. Discovery 690. The D-690 system uses a fully 3D-OSEM algorithm with all corrections incorporated into the iterative reconstruction scheme. Furthermore, new reconstruction algorithms are available on the D-690, which add to the standard configuration the time of flight information (TOF) and/or a 3D model of the D-690 PET point spread function (PSF). The activation of TOF and/or PSF does not require the setting of any new parameter compared to those used with the 3D-OSEM algorithm (number of subsets, number of iterations, reconstructed field of view (FOV), image matrix, and axial and transaxial postfilters). In this study, both TOF and PSF information were included in the reconstruction scheme.

2.4.4. Gemini XL. This system uses a fully 3D line-of-response (LOR) based iterative reconstruction algorithm named row-action maximum likelihood algorithm (RAMLA) [24]. The number of iterations is fixed (2 iterations and 33 subsets) and the reconstruction protocols contain one modifiable parameter that can be set to adjust the quality of the images as normal, smooth, or sharp.

2.4.5. Gemini TF. This system uses the TOF maximum likelihood expectation-maximization reconstruction algorithm (TF-MLEM) [24]. The reconstruction protocols contain three modifiable parameters that can be set to adjust the quality of the images: the first is the number of iterations (3 iterations and 20 subsets or 3 iterations and 33 subsets); the second is a so called relaxation parameter that can be set between 1, 0.7 and 0.5 and controls the magnitude of change that each iteration makes to the image. A third parameter, the kernel width of the TOF, can be set by the user at two levels (Gemini TF manual) [25].

The type of reconstruction algorithm, the degree of the convergence of the iterative algorithm, and the amount of the postreconstruction smoothing applied on images were varied starting from the clinical acquisition protocols used in each institution for radiotherapy planning. Overall, in each scanner, the maximum of theoretical independent combinations of acquisition parameters available for the subsequent model fitting were 9 sphere $A \times$ 9 TBR $\times$ 4 ESD = 324. The 8.1 and 6.5 mm spheres was not always included in the analysis because they were not clearly visible in all the phantom acquisitions. The number of reconstruction modalities available for model fitting depends on the scanner capabilities: the details of the reconstruction parameters together with the voxel size of the reconstructed images and the number of data points that is actually available for model fitting in each scanner are shown in Table 2.

2.5. Image Analysis. TS were determined as a percentage of the maximum intensity in the cross section area of the spheres. Target cross sections of area A were selected in the middle of the spheres, which constitutes the largest cross section of the sphere. The values of TS were entirely based on the apparent activity concentration in the images and not on the known activities actually placed in the spheres. To find the TS value that yielded an area A best matching the true value, the cross sections were autocontoured in the attenuation corrected slices varying TS in step of 1%, until the area so determined differed by less than 10 mm^2 versus its known physical value.

The analysis was performed by means of an automatic routine, EyeLite RT v.1.1 (G-Squared, Vicenza, Italy), to avoid the influence of the operator in ROIs dimensioning and to minimize the influence of the operator in the ROIs positioning. The operator placed six 17 mm-diameter ROIs in the background area surrounding the spheres. The mean intensity of these 6 ROIs was used as a background value (BG). ROI analyses were performed only for visually detectable spheres: this accounted for the discrepancy between theoretical and experimental data points collected for each scanner.

2.6. Statistical Analysis. For each combination of EM-equivalent iteration number (i) and ESD (j), the following variables were evaluated: X_{1ij} defined as target cross section A, X_{2ij} defined as $1 - 1/\text{TBR}$, and X_{3ij} defined as the FWHM.

Multiple linear regression analysis was performed in order to define the relationship between the best TS (TS$_{ij}$) (providing the most accurate sphere cross sectional area) and X_{1ij}, X_{2ij}, and X_{3ij}. The multiple regression model used for the fit was

$$\text{TS}_{ij} = B_0 + B_1 \times X_{1ij} \left(\text{mm}^2 \right) + B_2 \times X_{2ij} + B_3 \times X_{3ij} + E,$$

(1)

TABLE 2: PET/CT scanners: reconstruction parameters.

	Discovery ST S1	Discovery ST S2	Discovery STE S3	Discovery 600 S4	Discovery 600 S5	Discovery 690 S6	Biograph 16 Hi-REZ S7	Biograph 6 true V S8	Gemini XL S9	Gemini TF S10	Gemini TF S11
Reconstruction protocol	FORE-OSEM	FORE-OSEM	3D-OSEM	3D-OSEM	3D-OSEM	3D-OSEM	FORE-OSEM	FORE-OSEM	LOR-RAMLA	TF-MLEM	TF-MLEM
Number of iterations	14, 42	16, 40	14, 42	16, 32	16, 48	54, 108	16, 24	16, 42	66	60, 99	60, 99
Transaxial smoothing FWHM (mm)	6, 9, 13	5.5, 8.2, 11	6, 8, 11	6, 8, 11	5.5, 8.2, 11	4, 6, 8	4, 6, 8	4, 6, 8	—	—	—
Axial smoothing (kernel)	—	—	1:4:1	1:4:1	1:4:1	1:6:1 1:4:1 1:2:1	—	—	—	—	—
Kernel width (cm)	—	—	—	—	—	—	—	—	—	14.1	14.1, 18.7*
Relaxation (λ)	—	—	—	—	—	—	—	—	0.5, 0.7, 1	0.5, 0.7, 1	0.5, 0.7, 1
Voxel dimensions $L \times W \times H$ (mm)	$2.7 \times 2.7 \times 3.3$	$2.7 \times 2.7 \times 3.3$	$2.7 \times 2.7 \times 3.3$	$2.7 \times 2.7 \times 3.3$	$2.7 \times 2.7 \times 3.3$	$2.7 \times 2.7 \times 3.3$	$2.6 \times 2.6 \times 2$ $5.3 \times 5.3 \times 2$	$4.1 \times 4.1 \times 5$	$4 \times 4 \times 3$	$4 \times 4 \times 3$	$4 \times 4 \times 3$
Number of reconstructions	6	6	6	6	6	18	12	6	3	6	9
Number of data points	1641	1651	1676	1785	1727	5368	3456	1676	814	1993	2466

*Only available with 99 iterations.

where B_0, B_1, B_2, and B_3 are the regression coefficients to be estimated and E is the error term. The hypothesis of linear dependence between TS and independent variables X were already demonstrated in Brambilla et al. and in Matheoud et al. [14, 15]. However, nonlinear objective functions or even indicator functions according to the different parameters range could be investigated for fitting TS.

Additional variables, reflecting the characteristics of the reconstruction protocol of each considered scanner, were inserted in the model as independent predictors. Axial smoothing was considered for the Discovery 690 and voxel dimensions were inserted for the Biograph Hi-REZ, while relaxation parameter and TOF kernel width were accounted for in the Gemini and in the Gemini TF, respectively.

Stepwise forward selection was used as a strategy for selecting the variables and F statistic was used as a criterion for selecting a model. Goodness of fit for each regression model was expressed using the adjusted coefficient of determination (R^2). Goodness of fit was reported at each stage of model building as partial R^2. The criteria for retaining a variable in a model were $F > 4$ and an increment of at least 0.01 in the R^2 in order to be cautious in including redundant variables into the models. The weight of the different independent variables in explaining TS was quantified by means of standardized regression coefficients βi.

The reliability of the regression models was assessed through split-sample analysis [26]. Using this methodology, all observations in each scanner model were randomly assigned to one of two groups, the training group or the holdout group. The regression models were derived using the training group and the sample squared multiple correlation R^2 was obtained. Then the prediction equation for the training group was used to compute predicted values for the holdout group. Finally, the univariate correlation R^{2*} (cross validation correlation) was obtained between these predicted values and the observed responses in the holdout group. The reliability of the regression models was expressed by using the shrinkage on cross validation coefficients $R^2 - R^{2*}$. As a criterion, shrinkage values of less than 0.10 were considered as indicative of a reliable model.

In order to compare separate multiple regressions of TS as a function of the X independent variables for two different scanners, an additional dummy variable, coding for each scanner, was inserted in the model. A regression model was then built by pooling the data coming from the two scanners and inserting this dummy variable as a predictor. The criteria for retaining this variable in the model were the ones specified above.

Statistical analysis was performed using the software Statistica 6.0 (Statsoft Inc., Tulsa OK).

3. Results and Discussion

3.1. Multiple Linear Regression. Figure 2 shows the plots of averaged TS versus cross sectional areas for a coarse grouping of TBRs for each scanner model.

The TS versus predictor variables plot was fitted only for cross sectional area $> 133\,\text{mm}^2$ that is in the range of clinically relevant volumes comprised between 1 and 100 mL. The cross sectional area of $133\,\text{mm}^2$ (that corresponds to a sphere ID of 13 mm and approximately the twofold FWHM of the scanners) was selected as a separator of the data due to the resolution characteristics of the scanners.

Following (1), the regression equations that best summarize the results obtained in a multiple regression model with TS as the predicted variable are reported for each scanner in Table 3 together with the values of the corresponding parameters B_0–B_3. In the third column of Table 3 the multiple-R^2 of model fitting are reported, while the last column shows the ranking of the independent predictors together with the standardized regression coefficients and the amount of TS variance explained by each predictor.

The emission scan duration and the degree of the convergence of the iterative algorithm were never significant predictors of TS. This provided a confirmation of previously reported findings. Also the axial smoothing in the Discovery 690, the voxel size in the Biograph Hi-REZ, the relaxation λ in the Gemini scanners, and the Kernel width of the time of flight correction in the Gemini TF were not significant predictors of TS.

The goodness of the model fit, assessed by the coefficient of determination R^2, was high, ranging from a minimum of 0.74 for the Discovery 690 to a maximum of 0.92 for both the Discovery 600 and the Biograph Hi-REZ. The most relevant variable for TS prediction was TBR with a partial R^2 accounting for 74% to 91% of TS variability. In the case of the Discovery 690, TBR only accounted for 40% of TS variability, although it remains the best individual predictor. Second came the amount of smoothing in the transaxial plane (FWHM) that showed an additional R^2 roughly explaining from 1 to 5% of TS variability. The only exceptions were the Gemini scanners, where this parameter cannot be varied by the user, and the Discovery 690, where its contribution is significantly increased to 29% of TS variability. Last came the lesion size (A) that played an independent role only in the Discovery 690 and in the Gemini scanners accounting for 5%–8% of TS variability.

The comparison of the regression lines obtained from two scanners of the same model (Discovery ST, Discovery 600, and Gemini TF) did not evidence any relevant difference. The test of the hypothesis of coincident regression lines for the two DST scanners provided an $F_{3,2145} = 0.73$ ($P = 0.53$). This F statistics is small (P is large), so we do not reject H_0 and therefore have no statistical basis for believing that the two lines are not coincident. The details of the test are reported in Table 4. Moreover, the additional R^2 of the "scanner" dummy variable as predictor of TS variance was below <0.001. Similar results were found for the D600 and GTF scanners (not shown). Accordingly, the results of the regression analysis obtained by pooling all the measurements from scanners of the same model are reported in Table 3.

3.2. Regression Model Reliability. The results of the reliability study on regression models are reported in Table 3. The shrinkage on cross validation was always below 0.07, which is quite small and indicative of an excellent reliability of
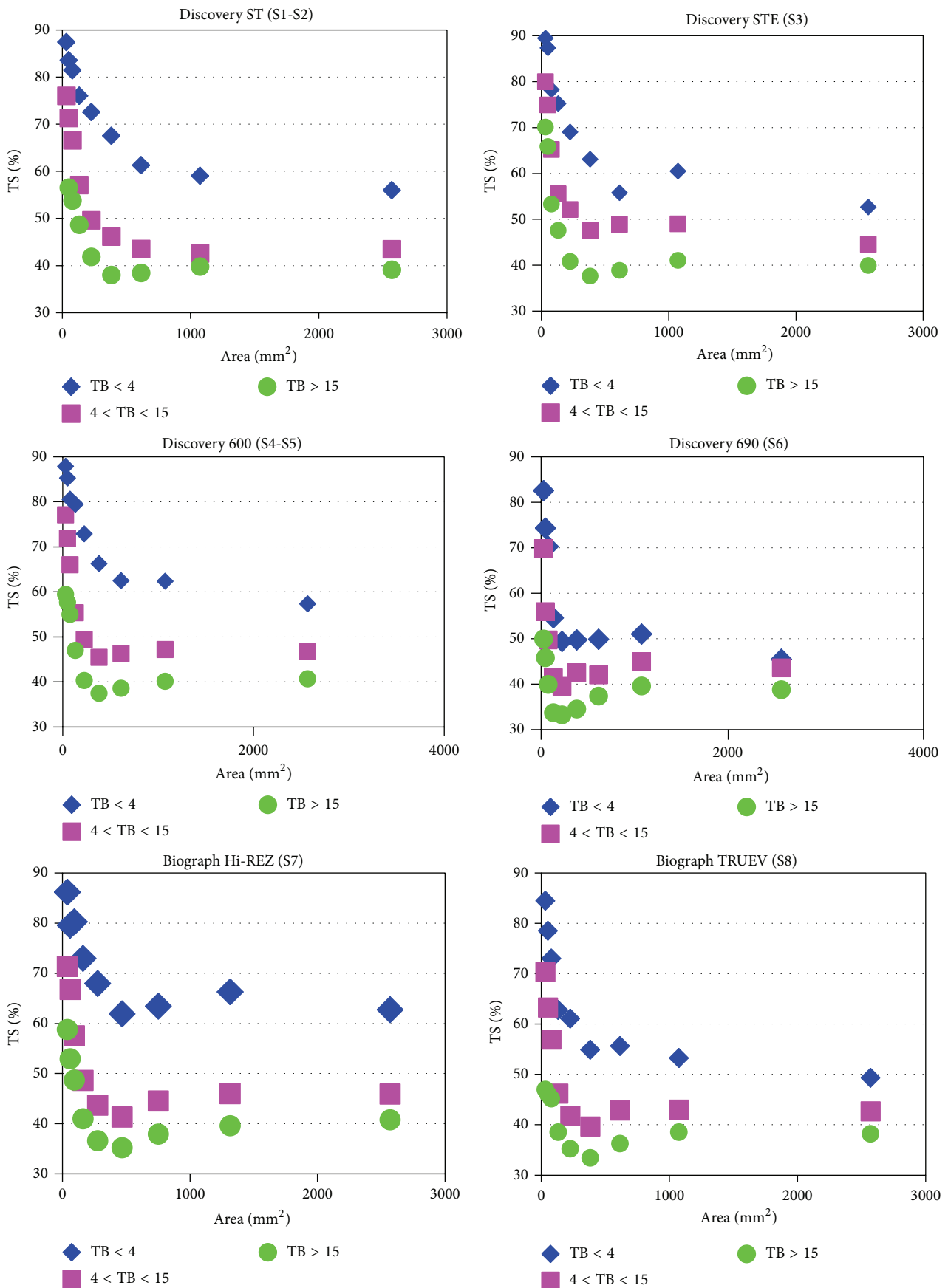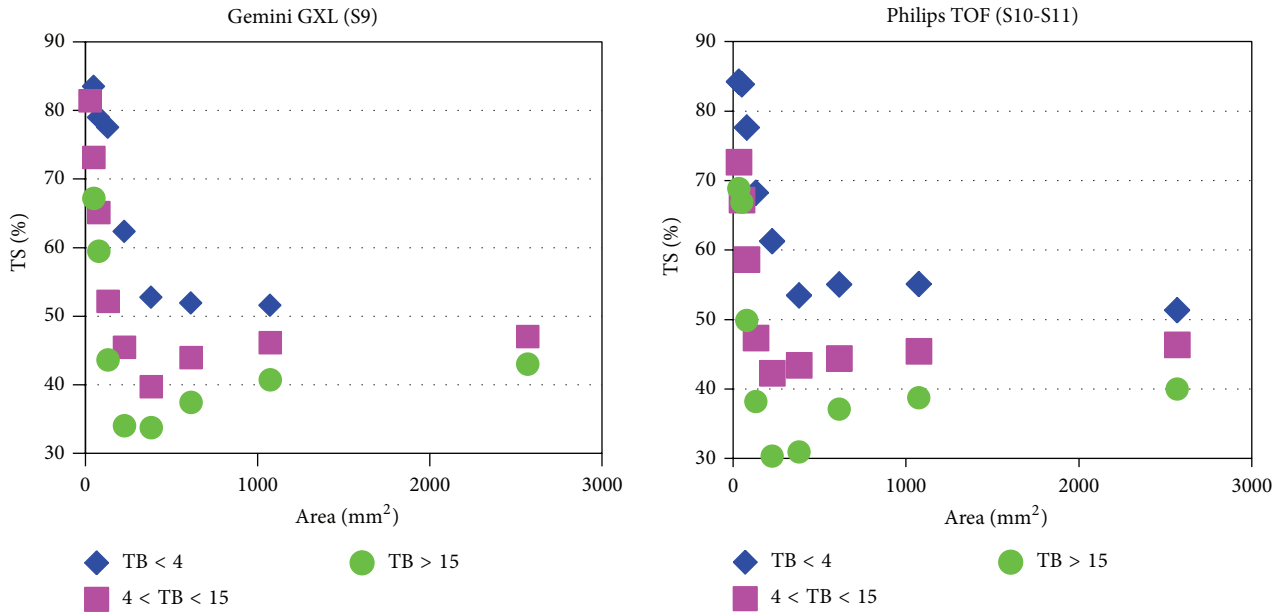
Figure 2: Continued.

FIGURE 2: Plots of averaged TS versus cross sectional areas for a coarse grouping of TBRs for each scanner model. Measured data points have been omitted here to prevent obscuring differences in trends.

estimation. An important aspect related to assessing the reliability of a model involves considering difference score of the form $TS_{observed} - TS_{predicted}$, where only holdout cases are used and when the training sample equation is used to compute the predicted values. The "unstandardized residuals" can be subjected to various residual analyses. The most helpful entails univariate descriptive statistics such as the box and whiskers plots depicted in Figure 3. In our case a few large residuals are present, but they are neither sufficiently implausible nor influential to require further investigations.

In our investigation, we derived the calibration curve for eleven PET scanners (eight models, three manufacturers, and eight sites) to apply the adaptive-threshold algorithm for PET-based contouring. The eight scanner types investigated in this study differ in scintillation crystal, scanner electronics, and reconstruction methodologies. Methods of retrospective image resolution recovery such as PSF-reconstruction or TOF measurements were also characterized in the present study.

At present, there is considerable variability in the way standard PET/CT scans are performed in different centers [27, 28]. Thus, there was no chance for a multicenter standardization of all scanners and all imaging protocols in use. Instead, we chose to directly incorporate in our adaptive thresholding algorithm the reconstruction parameters that can be selected by the user and that are relevant for TS determination. This should increase the robustness of the proposed method by avoiding the need to perform individual calibrations in each center of the algorithm. Noteworthy, the comparison of the regression lines obtained from two scanners of the same model did not evidence any relevant difference, at least for the three scanner models tested. This brings another relevant consequence; that is, with the incorporation of the reconstruction parameters in the regression

models the calibration curve in a specific scanner model need not to be obtained at each site. Instead, it can be derived once and applied irrespectively of the specific scanner being utilized provided that is of the same model.

3.3. Comparison with Previous Published Papers. To the best of our knowledge only three studies have been published so far on the integration of PET/CT scans from different hospitals into radiotherapy treatment planning. In the first study, Öllers et al. [29] used a TBR algorithm to evaluate head-and-neck tumors. To this purpose only small spheres of volumes ranging from 2 to 16 mL (i.e., sphere ID less than 3 cm) were used. TBRs, as determined by the dose calibrators, ranged from 2 to 12. The authors performed phantom measurements on three scanners of the same manufacturer (Biograph Accel, Siemens) equipped with Pico 3D (2 scanner) or standard (1 scanner) detector electronics. Identical acquisition and reconstruction protocols were used. To study the effect of different reconstruction parameters on the results PET raw data were reconstructed varying the number of iterations (IT from 2 to 64) with a fixed smoothing of FWHM = 5 mm. They found that the standardized uptake value (SUV) threshold of the scanner equipped with standard electronic differed significantly from those of the other two scanners and that at least 16 iterations are required in order to produce reliable SUV thresholds. Our own results support these findings. On the one hand, the regression lines did not differ significantly between scanners of the same type equipped with similar electronics, while the calibration curves for scanners of different type clearly differ (Figure 2). On the other hand, the number of iteration is not a significant predictor of TS, provided that this number is kept above a certain level which is both recommended by the manufactures and necessary to have good image quality.

Table 3: Scanner-model specific calibration curves, model R^2, shrinkage on cross validation and independent predictor of TS ranked in order of significance.

	Equation	R^2	$R^2(1) - R^2 * (2)$	TS predictors
Discovery ST	TS = 90.68 − 62.44 · (1 − 1/TBR) + 1.05 · FWHM (mm)	0.91	−0.008	1 − 1/TBR (β_2 = −0.91, partial R^2 = 0.85) FWHM (β_3 = 0.23, additional R^2 = 0.06)
Discovery STE	TS = 88.68 − 59.39 · (1 − 1/TBR) + 1.03 · FWHM (mm)	0.87	0.004	1 − 1/TBR (β_2 = −0.90, partial R^2 = 0.82) FWHM (β_3 = 0.23, additional R^2 = 0.05)
Discovery 600	TS = 93.52 − 64.38 · (1 − 1/TBR) + 0.99 · FWHM (mm)	0.92	0.000	1 − 1/TBR (β_2 = −0.93, partial R^2 = 0.88) FWHM (β_3 = 0.20, additional R^2 = 0.04)
Discovery 690	TS = 63.04 − 0.015 ·A (mm^2) − 40.51 · (1 − 1/TBR) + 1.92 · FWHM (mm)	0.74	0.007	1 − 1/TBR (β_2 = −0.61, partial R^2 = 0.40) FWHM (β_3 = 0.54, additional R^2 = 0.29) A (β_1 = 0.22, additional R^2 = 0.05)
Biograph Hi-REZ	TS = 88.19 − 56.18 · (1 − 1/TBR) + 0.67 · FWHM (mm)	0.92	−0.012	1 − 1/TBR (β_2 = −0.95, partial R^2 = 0.91) FWHM (β_3 = 0.10, additional R^2 = 0.01)
Biograph TRUEV	TS = 90.43 − 61.86 · (1 − 1/TBR) + 0.95 · FWHM (mm)	0.89	−0.013	1 − 1/TBR (β_2 = −0.91, partial R^2 = 0.85) FWHM (β_3 = 0.19, additional R^2 = 0.04)
Gemini XL	TS = 92.04 + 0.0025 ·A (mm^2) − 59.15 · (1 − 1/TBR)	0.82	0.067	1 − 1/TBR (β_2 = −0.89 partial R^2 = 0.74) A (β_1 = 0.27, additional R^2 = 0.08)
Gemini TF	TS = 88.57 + 0.0027 ·A (mm^2) − 57.44 · (1 − 1/TBR)	0.84	−0.009	1 − 1/TBR (β_2 = −0.88 partial R^2 = 0.76) A (β_1 = 0.28, additional R^2 = 0.08)

TABLE 4: Analysis of variance table. Test of H_0 = coincident regression lines for the two DST scanners.

$A > 133 \, \text{mm}^2$	Sum of squares (SS)	Degrees of freedom	Mean square (MS)	F	P
Reduced model					
Regression	281844	2	140922	10445	$< 10^{-6}$
Residuals	28980	2148	13.5		
Full model					
Regression	281874	3	93958	6968	$< 10^{-6}$
Residuals	28951	2147	13.5		

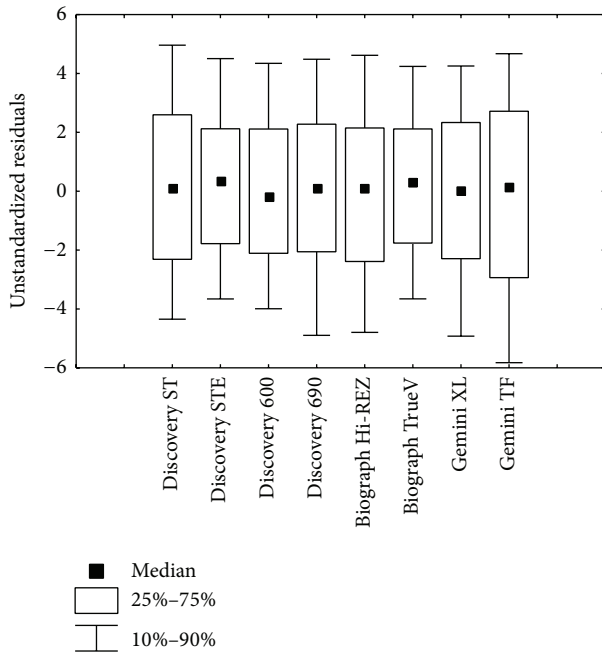$F = (28980,3 - 28950,6)/3/13.48 = 0.73; P = 0.53.$

FIGURE 3: Box and whiskers plot of unstandardized residuals $(\text{TS}_{\text{observed}} - \text{TS}_{\text{predicted}})$ for the different scanners where only holdout cases are used and when the training sample equation is used to compute the predicted values.

In a second study, Hatt et al. [30] evaluated the robustness and repeatability of a TBR algorithm in comparison to fuzzy C-means clustering and fuzzy locally adaptive Bayesian algorithm. The authors performed phantom measurements on four different PET/CT scanners (Philips Gemini and Gemini TF, Siemens Biograph, and GE Discovery LS) using a standard acquisition protocol with two TBR (4 and 8) and Three ESD (1, 2, and 5 min). PET raw data were reconstructed using routine clinical image reconstruction and two voxel size volume for all scanners. They reported a higher robustness of the fuzzy locally adaptive Bayesian algorithm while the repeatability provided by all segmentation methods was very high with a negligible variability of <5% in comparison to that associated with manual delineation. However, as recognized by the same authors, in order to assess the robustness of the TBR approach they applied an adaptive thresholding

using the parameters optimized on other scanners to the image datasets acquired with the Siemens Biograph, which is sort of misleading since the TBR approach is system dependent. Instead, by adopting scanner-model specific calibration curves, similar mean classification error (~10%) and variability (~5%) would have been obtained for the TBR algorithm and for the fuzzy locally adaptive Bayesian approach. By first principles, the inclusion of the postreconstruction smoothing (not considered in the study of Hatt) should increase both the accuracy and robustness of adaptive thresholding algorithms possibly leading to results even superior to those achieved by advanced image segmentation methods. Noteworthy, the coefficient of regression for the TBR variable reported by Hatt for the Gemini TF is very similar to the one obtained for the same variable in the present work ($B_{\text{TBR}} = 61.4$ versus 59.3), also considering that the two regression models are not identical. This provides, although indirectly, a further confirmation of the robustness of the scanner-model specific approach in deriving TS calibration curves.

In the last study Schaefer et al. [31] evaluated the calibration of an adaptive SUV thresholding algorithm in eleven centers equipped with 5 Siemens Biograph, 5 Philips Gemini, and one Siemens ECAT ART scanners. They reported only minor differences in calibration parameters for scanners of the same type provided that identical imaging protocols were used, whereas significant differences were found comparing scanners of different types. Moreover, they reported no statistically significant differences among SUV thresholds calculated for each site by use of the "site-specific" calibration neither among scanners of the same type at different sites nor among scanners of different types at different sites. Our own results support these findings only partially. In our study both acquisition and reconstruction parameters were varied and relevant parameters were incorporated into the "site-specific" algorithms so that there is no need to force individual centers to adopt a fixed protocol of image acquisition and image reconstruction. Bearing in mind this relevant difference, also in our study the calibration curves were not significantly different between scanners of the same type, whereas significant differences were found comparing scanners of different types. On the contrary, both the measured (Figure 2) and the calculated TS (Table 3) were significantly different among scanners of different types. For instance, the measured TS averaged over the entire spectrum of acquisition and reconstruction parameters for larger targets (sphere $A > 133 \, \text{mm}^2$) for the Discovery 690 (S6) and the Discovery 600 (S4-5) were $39.0 \pm 5.8\%$ versus 45.4 ± 10.8, respectively ($P < 0.0001$). This difference largely reflects the hot-contrast recovery capabilities of the different scanners which, in the case of the Discovery 690, are emphasized by the introduction of PSF techniques in the reconstruction process.

3.4. Study Advantages and Limitations.

The proposed method for the definition of BTV has several advantages, even if the results of this study must be interpreted in the context of some limitations.

Our method is feasible in a clinical context in those lesions presenting a uniform radiotracer uptake, as for

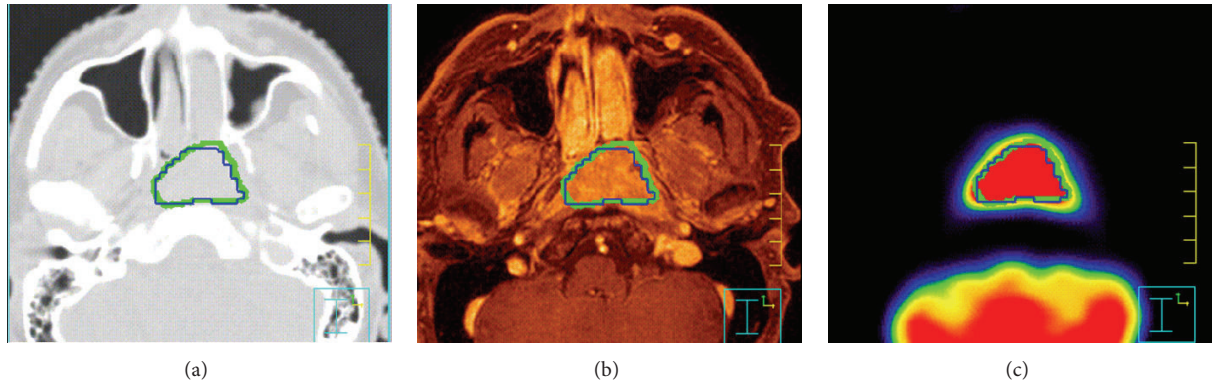(a)   (b)   (c)

FIGURE 4: Application of the proposed method to a head and neck tumor: CT (a), MR (b), and PET (c) images (acquired on the biograph Hi-REZ PET scanner). The green ROI corresponds to the GTV delineation manually performed by the radiation oncologist, while the blue one is the result of the application of the 39% TS derived from the thresholding algorithm (TBR = 12, FWHM = 4 mm). The differences between manual GTV and GTV obtained from the thresholding algorithm is 4.3%.

different oncological lesions. In this case, the proposed method can be effective for extracting functional biomarkers and for using PET imaging in image-guided radiotherapy treatments. As a representative example, Figure 4 shows the application of the proposed method to a head and neck oncological patient, a candidate for image-guided radiotherapy. In these patients, BTV can be used to optimize radiotherapy treatment taking advantages from the information of functional imaging.

The effects of lesion movement in lung tumors have been recently incorporated in an adaptive thresholding algorithm using multiple regression techniques similar to those in the present study [32]. Though the effects of lesion movement were not included in this study, we believe that the conclusions regarding the effect of smoothing and TBR on thresholds still apply in the case of moving targets.

Threshold techniques do not take into account variations in tumor heterogeneity. This has motivated the investigation of advanced segmentation techniques not based on thresholding. While referring to these important methods for segmentation of nonuniform tracer concentration it should be pointed out that until they are further developed and validated, adaptive threshold segmentation methods are and will be used in most clinics and therefore need to be accurately characterized.

Furthermore, it has to be pointed out that in our work BTVs were fitted only for cross sections larger than 133 mm^2. This choice is justified by the fact that several studies found severe errors in the volume estimation for tumor volume < 2 mL corresponding to cross sections < 192 mm^2 (in terms of sphere-equivalent cross section) [33–35].

4. Conclusion

This study demonstrated that the calibration curves for the proposed adaptive thresholding method were not significantly different between scanners of the same type at different sites. The incorporation of the postreconstruction Gaussian smoothing in the algorithms avoids the need of system-dependent optimization procedures. This, together with the

demonstrated high level of reliability of this approach, may provide robust and reliable tools to aid physicians as an initial guess in segmenting biological volumes on FDG-PET images.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] U. Nestle, W. Weber, M. Hentschel, and A.-L. Grosu, "Biological imaging in radiation therapy: role of positron emission tomography," *Physics in Medicine and Biology*, vol. 54, no. 1, pp. R1–R25, 2009.

[2] X. Geets, J. A. Lee, A. Bol, M. Lonneux, and V. Grégoire, "A gradient-based method for segmenting FDG-PET images: methodology and validation," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 34, no. 9, pp. 1427–1438, 2007.

[3] 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," *Medical Physics*, vol. 35, pp. 3711–3721, 2008.

[4] A. J. Green, R. J. Francis, S. Baig, and R. H. J. Begent, "Semiautomatic volume of interest drawing for 18F-FDG image analysis—method and preliminary results," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 35, no. 2, pp. 393–406, 2008.

[5] 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," *Medical Physics*, vol. 36, no. 10, pp. 4349–4358, 2009.

[6] D. W. G. Montgomery, A. Amira, and H. Zaidi, "Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model," *Medical Physics*, vol. 34, no. 2, pp. 722–736, 2007.

[7] M. Halt, C. C. Le Rest, A. Turzo, C. Roux, and D. Visvikis, "A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET," *IEEE Transactions on Medical Imaging*, vol. 28, no. 6, pp. 881–893, 2009.

[8] I. El Naqa, D. Yang, A. Apte et al., "Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning," *Medical Physics*, vol. 34, no. 12, pp. 4738–4749, 2007.

[9] S. Belhassen and H. Zaidi, "A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET," *Medical Physics*, vol. 37, no. 3, pp. 1309–1324, 2010.

[10] A. C. Riegel, A. M. Berson, S. Destian et al., "Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion," *International Journal of Radiation Oncology Biology Physics*, vol. 65, no. 3, pp. 726–732, 2006.

[11] J.-F. Daisne, T. Duprez, B. Weynand et al., "Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen," *Radiology*, vol. 233, no. 1, pp. 93–100, 2004.

[12] A. Schaefer, S. Kremp, D. Hellwig, C. Rübe, C. M. Kirsch, and U. Nestle, "A contrast-oriented algorithm for FDG-PET-based delineation of tumour volumes for the radiotherapy of lung cancer: derivation from phantom measurements and validation in patient data," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 35, no. 11, pp. 1989–1999, 2008.

[13] U. Nestle, A. Schaefer-Schuler, S. Kremp et al., "Target volume definition for ^{18}F-FDG PET-positive lymph nodes in radiotherapy of patients with non-small cell lung cancer," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 34, no. 4, pp. 453–462, 2007.

[14] R. Matheoud, P. D. Monica, G. Loi et al., "Influence of reconstruction settings on the performance of adaptive thresholding algorithms for FDG-PET image segmentation in radiotherapy planning," *Journal of Applied Clinical Medical Physics*, vol. 12, no. 2, pp. 3363–3381, 2011.

[15] M. Brambilla, R. Matheoud, C. Secco, G. Loi, M. Krengli, and E. Inglese, "Threshold segmentation for PET target volume delineation in radiation treatment planning: the role of target-to-background ratio and target size," *Medical Physics*, vol. 35, no. 4, pp. 1207–1213, 2008.

[16] R. Matheoud, P. D. Monica, C. Secco et al., "Influence of different contributions of scatter and attenuation on the threshold values in contrast-based algorithms for volume segmentation," *Physica Medica*, vol. 27, no. 1, pp. 44–51, 2011.

[17] V. Bettinardi, M. Danna, A. Savi et al., "Performance evaluation of the new whole-body PET/CT scanner: discovery ST," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 31, no. 6, pp. 867–881, 2004.

[18] M. Teräs, T. Tolvanen, J. J. Johansson, J. J. Williams, and J. Knuuti, "Performance of the new generation of whole-body PET/CT scanners: discovery STE and Discovery VCT," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 34, no. 10, pp. 1683–1692, 2007.

[19] E. de Ponti, S. Morzenti, L. Guerra et al., "Performance measurements for the PET/CT Discovery-600 using NEMA NU 2-2007 standards," *Medical Physics*, vol. 38, no. 2, pp. 968–974, 2011.

[20] V. Bettinardi, L. Presotto, E. Rapisarda, M. Picchio, L. Gianolli, and M. C. Gilardi, "Physical performance of the new hybrid PETCT Discovery-690," *Medical Physics*, vol. 38, no. 10, pp. 5394–5411, 2011.

[21] M. Brambilla, C. Secco, M. Dominietto, R. Matheoud, G. Sacchetti, and E. Inglese, "Performance characteristics obtained for a new 3-dimensional lutetium oxyorthosilicate-based whole-body PET/CT scanner with the National Electrical Manufacturers Association NU 2-2001 standard," *Journal of Nuclear Medicine*, vol. 46, no. 12, pp. 2083–2091, 2005.

[22] B. W. Jakoby, Y. Bercier, M. Conti, M. E. Casey, B. Bendriem, and D. W. Townsend, "Physical and clinical performance of the mCT time-of-flight PET/CT scanner," *Physics in Medicine and Biology*, vol. 56, no. 8, pp. 2375–2389, 2011.

[23] S. Surti, A. Kuhn, M. E. Werner, A. E. Perkins, J. Kolthammer, and J. S. Karp, "Performance of Philips Gemini TF PET/CT scanner with special consideration for its time-of-flight imaging capabilities," *Journal of Nuclear Medicine*, vol. 48, no. 3, pp. 471–480, 2007.

[24] J. Browne and A. R. De Pierru, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography," *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 687–699, 1996.

[25] T. F. Gemini, *Key to Success, Philips Medical Systems*, Koninklijke Philips Electronics N.V., Cleveland, Ohio, USA, 2007.

[26] D. G. Kleinbaum, L. L. Kupper, and K. E. Muler, *Applied Regression Analysis and Other Multivariable Methods*, PWS-Kent, Boston, Mass, USA, 1988.

[27] M. M. Graham, R. D. Badawi, and R. L. Wahl, "Variations in PET/CT methodology for oncologic imaging at U.S. academic medical centers: an imaging response assessment team survey," *The Journal of Nuclear Medicine*, vol. 52, no. 2, pp. 311–317, 2011.

[28] T. Beyer, J. Czernin, and L. S. Freudenberg, "Variations in clinical PET/CT operations: results of an international survey of active PET/CT users," *Journal of Nuclear Medicine*, vol. 52, no. 2, pp. 303–310, 2011.

[29] M. Öllers, G. Bosmans, A. van Baardwijk et al., "The integration of PET-CT scans from different hospitals into radiotherapy treatment planning," *Radiotherapy and Oncology*, vol. 87, no. 1, pp. 142–146, 2008.

[30] M. Hatt, C. Cheze le Rest, N. Albarghach, O. Pradier, and D. Visvikis, "PET functional volume delineation: a robustness and repeatability study," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 38, no. 4, pp. 663–672, 2011.

[31] A. Schaefer, U. Nestle, S. Kremp et al., "Multi-centre calibration of an adaptive thresholding method for PET-based delineation of tumour volumes in radiotherapy planning of lung cancer," *NuklearMedizin*, vol. 51, no. 3, pp. 101–110, 2012.

[32] A. C. Riegel, M. K. Bucci, O. R. Mawlawi et al., "Target definition of moving lung tumors in positron emission tomography: Correlation of optimal activity concentration thresholds with object size, motion extent, and source-to-background ratio," *Medical Physics*, vol. 37, no. 4, pp. 1742–1752, 2010.

[33] J. F. Daisne, M. Sibomana, A. Bol, T. Doumont, M. Lonneux, and V. Grégoire, "Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: Influence of reconstruction algorithms," *Radiotherapy and Oncology*, vol. 69, no. 3, pp. 247–250, 2003.

[34] P. Tylski, S. Stute, N. Grotus et al., "Comparative assessment of methods for estimating tumor volume and standardized uptake value in ^{18}F-FDG PET," *The Journal of Nuclear Medicine*, vol. 51, no. 2, pp. 268–276, 2010.

[35] F. Gallivanone, A. Stefano, E. Grosso et al., "PVE correction in PET-CT whole-body oncological studies from PVE-affected images images," *IEEE Transactions on Nuclear Science*, vol. 58, no. 3, pp. 736–747, 2011.