

Article

Performance Analysis of Six Semi-Automated Tumour Delineation Methods on [¹⁸F] Fluorodeoxyglucose Positron Emission Tomography/Computed Tomography (FDG PET/CT) in Patients with Head and Neck Cancer

Francesco Bianconi ^{1,*}, Roberto Salis ², Mario Luca Fravolini ¹, Muhammad Usama Khan ¹, Matteo Ministrini ³, Luca Filippi ⁴, Andrea Marongiu ², Susanna Nuvoli ², Angela Spanu ² and Barbara Palumbo ³

¹ Department of Engineering, Università degli Studi di Perugia, Via Goffredo Duranti 93, 06125 Perugia, Italy; mario.fravolini@unipg.it (M.L.F.); muhammadusama.khan@studenti.unipg.it (M.U.K.)

² Unit of Nuclear Medicine, Department of Medicine, Surgery and Pharmacy, University of Sassari, 07100 Sassari, Italy; r.salis15@studenti.uniss.it (R.S.); amarongiu2@uniss.it (A.M.); snuvoli@uniss.it (S.N.)

³ Section of Nuclear Medicine and Health Physics, Department of Medicine and Surgery, Università degli Studi di Perugia, Piazza Lucio Severi 1, 06132 Perugia, Italy; matteo.ministrini@ospedale.perugia.it (M.M.); barbara.palumbo@unipg.it (B.P.)

⁴ Policlinico Tor Vergata Hospital, Viale Oxford 81, 00133 Rome, Italy; luca.filippi@ptvonline.it

* Correspondence: bianco@ieee.org; Tel.: +39-075-585-3706



Citation: Bianconi, F.; Salis, R.; Fravolini, M.L.; Khan, M.U.; Ministrini, M.; Filippi, L.; Marongiu, A.; Nuvoli, S.; Spanu, A.; Palumb, B. Performance Analysis of Six Semi-Automated Tumour Delineation Methods on [¹⁸F] Fluorodeoxyglucose Positron Emission Tomography/Computed Tomography (FDG PET/CT) in Patients with Head and Neck Cancer. *Sensors* **2023**, *23*, 7952. <https://doi.org/10.3390/s23187952>

Academic Editor: Manuchehr Soleimani

Received: 15 August 2023

Revised: 1 September 2023

Accepted: 15 September 2023

Published: 18 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: *Background.* Head and neck cancer (HNC) is the seventh most common neoplastic disorder at the global level. Contouring HNC lesions on [¹⁸F] Fluorodeoxyglucose positron emission tomography/computed tomography (FDG PET/CT) scans plays a fundamental role for diagnosis, risk assessment, radiotherapy planning and post-treatment evaluation. However, manual contouring is a lengthy and tedious procedure which requires significant effort from the clinician. *Methods.* We evaluated the performance of six hand-crafted, training-free methods (four threshold-based, two algorithm-based) for the semi-automated delineation of HNC lesions on FDG PET/CT. This study was carried out on a single-centre population of $n = 103$ subjects, and the standard of reference was manual segmentation generated by nuclear medicine specialists. Figures of merit were the Sørensen–Dice coefficient (DSC) and relative volume difference (RVD). *Results.* Median DSC ranged between 0.595 and 0.792, median RVD between -22.0% and 87.4% . Click and draw and Nestle’s methods achieved the best segmentation accuracy (median DSC, respectively, 0.792 ± 0.178 and 0.762 ± 0.107 ; median RVD, respectively, $-21.6\% \pm 1270.8\%$ and $-32.7\% \pm 40.0\%$) and outperformed the other methods by a significant margin. Nestle’s method also resulted in a lower dispersion of the data, hence showing stronger inter-patient stability. The accuracy of the two best methods was in agreement with the most recent state-of-the-art results. *Conclusions.* Semi-automated PET delineation methods show potential to assist clinicians in the segmentation of HNC lesions on FDG PET/CT images, although manual refinement may sometimes be needed to obtain clinically acceptable ROIs.

Keywords: head and neck cancer; positron emission tomography; segmentation; region of interest; radiomics

1. Introduction

Head and neck cancer (HNC) is the seventh most common oncological disorder worldwide with over 660,000 new cases and 325,000 related deaths per year [1]. Current estimates place the four-year survival rate after diagnosis at 72% for men and 76% for women [2].

The primary prevention of HNC involves reducing the exposure to risk factors such as tobacco smoking, alcohol consumption, human papillomavirus (HPV) and Epstein–Barr virus (EBV) infections. Secondary prevention could also give excellent results due to the

good prognosis when the disease is diagnosed at an early stage [3,4]. In this context, positron emission tomography/computed tomography with [¹⁸F] Fluorodeoxyglucose (FDG PET/CT in the remainder) is currently one of the most reliable methods for the management of patients with HNC, its role being crucial for the diagnosis of the primary lesion, staging (detection of loco-regional lymph-node metastasis and distant metastasis) and follow-up after therapy [5,6]. However, a qualitative analysis of FDG PET/CT images is not always sufficient to obtain correct diagnosis and/or treatment monitoring. Quantitative data are necessary to better define the staging and re-staging of the disease.

Advances in technology and the advent of radiomics in nuclear medicine have the potential to obtain the accurate quantification of data extracted from images, thus providing non-invasive, personalised modelling, treatment planning and response assessment. The aim of radiomics is to convert medical images into mineable data that can be used for computer-assisted clinical decision making [7]. Radiomics relies on the ability to detect clinically relevant, sub-visual features that would go unnoticed to the human eye [8,9]. It also leverages on artificial intelligence methods and large datasets of pre-classified cases to make predictions about, for instance, disease phenotypes, or survival and response to therapy.

There are two main branches of radiomics: *conventional* (or *hand-crafted*) and based on *deep learning* [10]. The main difference is that image features are defined a priori in conventional radiomics, whereas, in deep learning, they are learnt from data. Both approaches have their pros and cons, and it is currently under debate which is the most suitable for translation into the clinical practice [11]. Although deep learning radiomics may produce more accurate prediction models, the gain comes at a cost—i.e., large datasets and computational resources needed for training, and the limited interpretability of the models (the ‘black box’ problem). On the other hand, conventional radiomics is computationally less expensive and relies on relatively easy-to-interpret image features which can be directly linked to the underlying characteristics of the tissue [12]. The existence of various, easy-to-use radiomics tools based on conventional radiomics [13] also facilitates testing and application in clinical environments. The above pros and cons translate seamlessly into lesion delineation (more on this later). Although methods based on deep learning may achieve better performance in some cases, they require large datasets for training and highly skilled personnel for model set-up and tuning. On the other hand, conventional methods are essentially training-free, easy to interpret and user-friendly.

Conventional radiomics (the focus of this work) comprises six steps [14,15]: acquisition, pre-processing, segmentation, feature extraction, post-processing and data analysis. Among them, the segmentation and feature extraction step are crucial. The first the identification and delineation of the region(s) of interest (ROI), the second involves the computation of sets of pre-defined (hand-crafted) image parameters such as first-order statistics, shape and texture features from the ROI.

In recent years, conventional radiomics applied to FDG PET/CT has demonstrated its potential to assist clinical decision making in patients with head and neck cancer [16]. Recent studies have supported the use of FDG PET/CT radiomics for predicting overall survival [17–20], disease-free survival [17–19,21], metastasis-free survival [19], loco-regional recurrence [22], local control [23] and response to therapy [24]. The outcome of radiomics studies, however, can be affected by a number of factors including study design (e.g., perspective vs. retrospective); image acquisition and reconstruction settings; spatial re-sampling; lesion delineation; signal quantisation and others. Such sources of uncertainty may easily lead to models that fail to generalise to new research trials [16,25,26]. Lesion delineation (also referred to as *segmentation* or *contouring*), in particular, is a critical step in the radiomics process, as the correct identification of the ROI is crucial to the development of robust prediction models.

Contouring suspicious head and neck lesions poses specific challenges due to the intrinsic complexity of the anatomical region. Manual delineation performed by experienced physicians is usually regarded as the standard of reference. Unfortunately, this is a tedious

and time-consuming procedure, and, as such, can represent a significant bottleneck to the whole procedure. To overcome this problem, a number of automated and semi-automated approaches have been proposed in the literature, some of which are already available in radiomics tools.

Although comparing different methods for lesion segmentation has been the subject of a number previous studies—particularly in lung cancer [27,28]—the delineation of HNC lesions on FDG PET/CT has received much less attention in the literature. Prior related works include a paper by Zaidi et al. [29], in which image segmentation techniques were compared against the reference specimens of pharyngolaryngeal squamous cell carcinoma collected ‘en bloc’ and frozen—although the analysis was essentially two-dimensional as the specimens were cut into 1.7–2 mm thick slices. More recently, Trada et al. [30] assessed the impact of tree semi-automated tumour delineation methods on the stability of semi-quantitative parameters computed from FDG PET/CT, but no direct assessment of ROI overlap was carried out. Other works have advocated for the use of convolutional network for segmentation [31], but these are outside the scope of this study as we focus on ready-to-use, training-free methods.

In this work, we benchmarked six semi-automated, training-free methods for segmenting HNC lesions on FDG PET/CT. The analysis was carried out on a retrospective, single-centre patient population using LIFEx, a freeware tool for radiomic feature calculation. The Sørensen–Dice coefficient and relative volume difference were, respectively, the primary and secondary figure of merit for the segmentation goodness. The standard of reference was manual delineation operated by nuclear medicine physicians. To the best of our knowledge, this is the first study of this kind in head and neck cancer; a disease that requires particular attention due to its poor prognosis and the heterogeneous characteristics of the anatomical district.

2. Materials and Methods

2.1. Patient Population

For this study, we retrospectively evaluated $n = 103$ head and neck lesions (volume = 22.7 ± 27.8 [0.5–157.7 cm³]) from as many subjects (male = 79, female = 24; age = 63.2 ± 9.2 (42–88 years)) who received baseline FDG PET/CT scans for clinical examination at the Unit of Nuclear Medicine of Department of Medicine, Surgery and Pharmacy; Università degli Studi di Sassari, Sassari, Italy, between December 2014 and March 2023. The inclusion criteria were:

- Age > 18;
- No previous surgical, chemotherapy and/or radiotherapy treatment for the suspicious lesion before the FDG PET/CT scan was obtained;
- Presence of a histology record for the region of interest after surgical resection.

All lesions were subsequently confirmed as head and neck squamous cell carcinomas (HNSCCs) by histological evaluation. Only primary lesions were considered for the analysis, and therefore, lymph-nodes and secondary lesions were not included in this study. The anatomical site distribution was: Oral cavity/oropharynx ($n = 71$), laryngopharynx ($n = 22$), nasopharynx ($n = 7$), and multiple sites ($n = 3$).

2.2. Image Acquisition

The scans were performed according to the AIMN guidelines [32] on a Discovery 710 PET/CT machine (GE HealthCare Technologies Inc., Chicago, IL, United States). Patients were required to fast for at least 6 h prior to the procedure and checked for glucose level lower than 150 mg/dL before the examination; then, image acquisition started 60 min after radiotracer injection. The PET images of matrix size 256 px × 256 px, in-plane pixel spacing 2.73 mm × 2.73 mm and slice thickness 3.27 mm were reconstructed by ordered subset expectation maximisation with point spread function recovery and time-of-flight (VFPXS). The dimension of the field of view of the PET camera was 700 × 153 mm. Low-dose CT scans for attenuation correction and anatomical correlation were acquired in

helicoidal mode under 120–140 kVp tube voltage and reconstructed into images of size of 512 px × 512 px, in-plane pixel spacing of 1.37 mm × 1.37 mm, slice thickness of 3.75 mm and spacing between slices of 3.27 mm. The total number of slices for patient ranged from 83 to 551 for PET and from 83 to 558 for CT. No pre-processing step like filtering, resampling or spatial rescaling was applied to the resulting images.

2.3. Lesion Delineation

Six semi-automated, training-free lesion delineation methods of the FDG PET signal were considered in this study (Sections 2.3.1–2.3.6) and compared with manually generated ground-truth segmentation (Section 2.3.7). Specifically, we evaluated one-fixed absolute threshold method, two fixed relative threshold methods, one background threshold method and two algorithm-based methods (see Table 1 for a recap; the overall experimental set-up is also summarised in Figure 1). All the semi-automated and manual delineation procedures were carried out using LIFEx v7.3.0 (LITO-Curie, SHFJ-CEA, CNRS, Univ. Paris Sud, University Paris Saclay, Orsay, France [33]). Figures 2 and 3 show the ground-truth and the segmentation results for two sample cases.

Table 1. Summary table of the methods considered in this study. Classification follows Im et al. [34].

Method	Class	Procedure	Refs.
2.5 SUV	Fixed absolute threshold	Draw and refine	[34,35]
40% SUVmax	Fixed relative threshold	Draw and refine	[17,34]
50% SUVmax	Fixed relative threshold	Draw and refine	[22,34]
Nestle	Background threshold	Draw and refine	[36]
Click and draw	Algorithm based	Click and draw	[37]
Click 40	Algorithm based	Click and draw	[37]

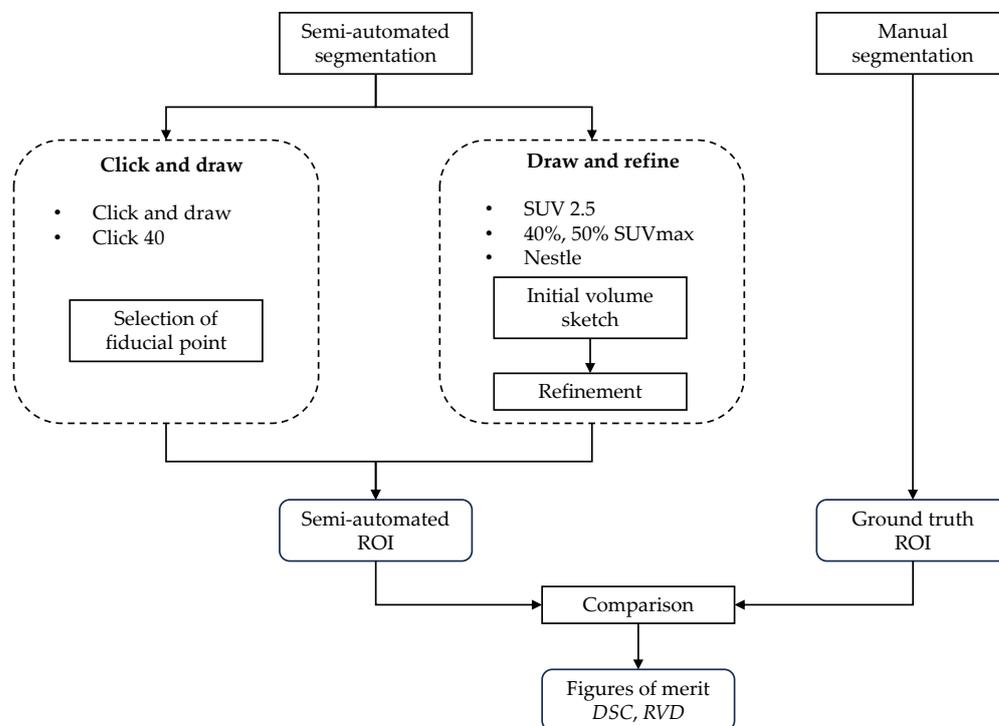


Figure 1. Experimental set-up: Summary flow-chart.

2.3.1. Absolute Threshold at SUV 2.5 (SUV 2.5)

Absolute thresholding is one of the simplest and most widespread method for FDG PET/CT image segmentation. It consists of setting a threshold value to separate the lesion from the background, and in this case, we set the threshold at 2.5 of the standardised uptake value (SUV), which is the typical choice in the literature [29]. As a result, all voxels with an SUV above the threshold are assigned to the ROI.

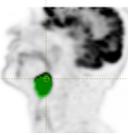
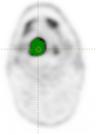
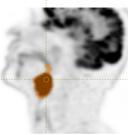
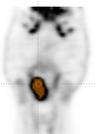
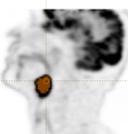
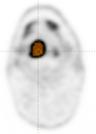
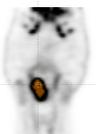
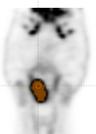
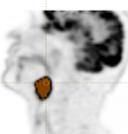
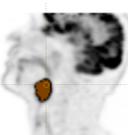
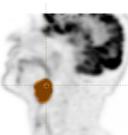
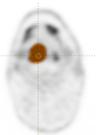
Method	View			Score	
	Coronal	Sagittal	Axial	DSC	RVD
Manual				—	—
SUV 2.5				0.606	126.6%
40% SUVmax				0.555	−61.4%
50% SUVmax				0.446	−71.3%
Nestle				0.669	−48.4%
Click and draw				0.706	−43.1%
Click 40				0.857	21.3%

Figure 2. Squamous cell carcinoma of the oropharynx in a 56-year-old male (0d64). Rows refer to the segmentation method used with manual segmentation on the top row. The columns Coronal, Sagittal and Axial report snapshots of the obtained segmentation on the three planes. Note that green overlays denote manual (ground truth) segmentation and brown overlays the semi-automatically generated ROIs. Finally, the last two columns on the right report the figures of merit (DSC, RVD) that each semi-automated segmentation method achieved in this case. Observe the spillover effect that affects SUV 2.5 and which leads to a markedly overestimated volume.

Operatively, we can define this as a ‘draw and refine’ procedure (Table 1, Figure 1), which involves two steps. In the first, the user sketches an initial three-dimensional volume that completely encircles the putative ROI; then all the voxels with uptake below the threshold are filtered out of the initial volume. We used a three-dimensional spherical painting tool for sketching the initial volume, and the physician was free to adjust the radius and position of the sphere(s) during this operation.

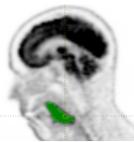
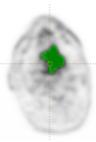
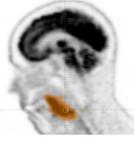
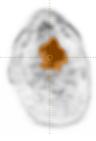
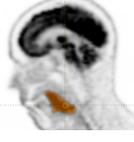
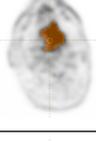
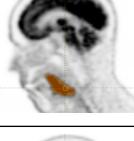
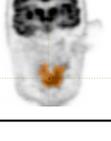
Method	View			Score	
	Coronal	Sagittal	Axial	DSC	RVD
Manual				—	—
SUV 2.5				0.555	160.3%
40% SUVmax				0.820	41.6%
50% SUVmax				0.911	−7.5%
Nestle				0.886	16.4%
Click and draw				0.875	21.5%
Click 40				0.573	148.9%

Figure 3. Squamous cell carcinoma of the oropharynx in a 66-year-old male (76b9). Rows refer to the segmentation method used with manual segmentation on the top row. The columns Coronal, Sagittal and Axial report snapshots of the obtained segmentation on the three planes. Note that green overlays denote manual (ground truth) segmentation and brown overlays denote the semi-automatically generated ROIs. Finally, the last two columns on the right report the figures of merit (DSC, RVD) that each semi-automated segmentation method achieved in this case. Like in Figure 2, we note that the spillover effect is again evident on SUV 2.5, but Click 40 also produces a patently overestimated volume in this case.

2.3.2. Relative Threshold at 40% of SUVmax (40% SUVmax)

Relative thresholding is also a very common approach for lesion delineation on FDG PET/CT. Operatively, this is analogous to SUV 2.5 (draw and refine); however, in this case, the threshold is defined as a percentage (40%) of the maximum SUV in the initial volume.

2.3.3. Relative Threshold at 50% of SUVmax (50% SUVmax)

Same as 40% SUVmax (Section 2.3.2), but with the threshold set at 50% of SUVmax. Note that we also performed preliminary tests with other threshold values (30% and 70%), but discarded these options as the results were not satisfactory.

2.3.4. Nestle's Method (Nestle)

This technique is also based on relative thresholding, and is therefore conceptually similar to 40% SUVmax and 50% SUVmax (Sections 2.3.2 and 2.3.3). The calculation of the threshold, however, is more involved in this case as it takes into account not only the radiotracer uptake of the initial region, but also that of the background. The threshold is computed as follows: [36]:

$$t_N = \beta \times I_{70} + I_{bgd} \quad (1)$$

where I_{70} indicates the mean uptake value in a volume containing all the voxels with an uptake of >70% of the maximum value (70% I_{max} isocontour volume), and I_{bgd} denoting the average background uptake. The latter is defined as the mean uptake in a shell of 4 mm thickness generated by offsetting the 70% SUVmax iso-contour boundary towards the exterior by 12 mm. Voxels in the shell with an uptake greater than SUV 2.5 units are excluded from the calculation. Finally, β is a tuning parameter which depends on the acquisition device, the acquisition settings and the reconstruction settings. The value of β should in principle be estimated through phantom experiments. Since no phantom data were available for this study, we preliminarily tested $\beta = 0.15$ and $\beta = 0.30$ (as, respectively, proposed in [36,38]) and chose the latter as it provided better qualitative results. Further details about this method are also available in Ref. [37].

2.3.5. Click and Draw

This is an algorithm-based method of the type 'click and draw' (Table 1, Figure 1). The only interaction required from the user is the selection of a point approximately around the centroid of the region to be delineated.

Let V_s indicate the voxel corresponding to the selected point, and denote by $I(V_s)$ the signal intensity (tissue uptake) at that voxel. The algorithm proceeds as follows (see also [37] for further details):

1. Perform a three-dimensional flood-fill [39] that extends to all voxels connected to V_s with an uptake of at least $t \times I(V_s)$, where $t = 0.7$. Let Ω denote the resulting (flood-filled) region.
 - Return an exception if $|\Omega| > 500 \text{ cm}^2$, where $|x|$ indicates the extent (volume in this case) of x .
2. Let I_N be Nestle's uptake threshold computed via Nestle's method (see Section 2.3.4) on $|\Omega|$ using $\beta = 0.3$ and I_{max} as the input parameters, where $I_{max} = \max_{V \in \Omega} [I(V)]$.
3. Perform a three-dimensional flood-fill again, this time starting at $V_M = \arg \max_{V \in \Omega} [I(V)]$, and including all voxels connected to V_M with an uptake of at least I_N .
4. Obtain the flood-filled region as the resulting ROI.

2.3.6. Click 40

This method is based on the same algorithm as Click and draw (Section 2.3.5) with the only difference being that a threshold value $t = 0.4$ instead of $t = 0.7$ is used for step 1.

2.3.7. Ground Truth

Reference (ground-truth) segmentation of the PET signal was generated via manual, slice-by-slice contouring on transaxial slices of the fused PET/CT images. The contouring was carried out by a nuclear medicine resident (R.S., >3 years experience) under the supervision of a senior nuclear medicine specialist (B.P., >20 years experience). For the process, the operators used a ‘CTL-472 One’ graphics tablet (Wacom Co., Ltd., Saitama, Japan) and were free to adjust the size of the painting tool and the image magnification factor based on their experience and personal judgement.

2.4. Evaluation Metrics

The effectiveness of the semi-automated lesion delineation methods described in Sections 2.3.1–2.3.6 was assessed by comparing the results of each method against manual ground-truth segmentation. We considered the Sørensen–Dice coefficient (DSC) and relative volume difference (RVD) as the primary and secondary figures of merit [40–42]:

$$\text{DSC} = \frac{2|\Omega_e \cap \Omega_{gt}|}{|\Omega_e| + |\Omega_{gt}|} \quad (2)$$

$$\text{RVD} = \frac{|\Omega_e| - |\Omega_{gt}|}{|\Omega_{gt}|} \quad (3)$$

In Equations (2) and (3), Ω_e and Ω_{gt} , respectively, denote the result of semi-automated segmentation and the ground truth ROI.

Let us recall that the Sørensen–Dice coefficient measures the overlap between two sets, and is arguably the most used metric for validating image segmentation algorithms. It ranges from 0 to 1, with 1 indicating perfect overlap (i.e., perfect segmentation) and 0 indicating no overlap [43,44]. Relative volume difference quantifies the extent to which the volume of semi-automated segmentation differs from that of the ground-truth ROI [40,41]. A positive RVD indicates that the algorithm overestimates the target volume, and a negative RVD indicates that it underestimates it. Note that the relative volume difference should never be considered alone, but always in conjunction with DSC or other metrics, as RVD can be zero even when there is no overlap between the estimated and the target ROI.

2.5. Statistical Analysis

Statistical differences between the performance of the segmentation methods were assessed via pairwise non-parametric Mann–Whitney U test at a significance value $\alpha = 0.05$. Bonferroni’s correction for multiple tests was also applied.

3. Results

3.1. Segmentation Overlap Assessed by Sørensen–Dice Coefficient

The box plots/strip plots in Figure 4 show the distribution of the Sørensen–Dice coefficient (DSC) for the semi-automated methods considered in this study. Aggregate values are also reported in Table 2.

We observe that, in absolute terms, the two best methods (click and draw and Nestle), respectively, achieved an aggregate DSC of 0.792 and 0.762, figures which are in good agreement (and even slightly higher) than has been reported in the recent literature [21,31]. Figure 4 also shows a marked dispersion of the data, which indicates strong inter-case variability. This was particularly evident for click and draw and click 40, which sometimes resulted in a DSC close to zero, indicating a completely misplaced ROI. By contrast, Nestle had the lowest DSC spread (assessed by standard deviation) among the methods considered.

On a relative scale, we observe that click and draw and Nestle outperformed the other methods by a noticeable margin in terms of median DSC. Pairwise comparison by Mann–Whitney U test (Table 3) indicates that there was no statistically significant difference between these two methods. However, click and draw had a significantly higher DSC score

than all other methods except Nestle, while Nestle significantly outperformed SUV 2.5 and 50% SUVmax.

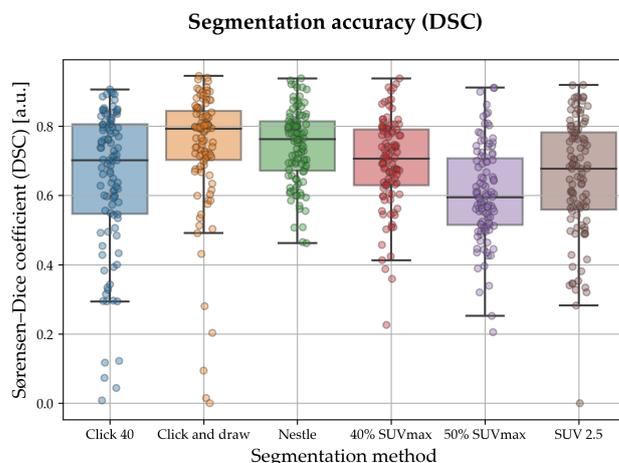


Figure 4. Box plots/strip plots showing DSC values (Sørensen–Dice coefficient) for the semi-automated segmentation methods considered in this study. Colour and x tick labels denote segmentation method; each dot in the strip plots represents one lesion/patient. Observe that click and draw achieved the highest median DSC followed by Nestle; the latter, however, had a remarkably lower dispersion of the data.

Table 2. Aggregate performance (DSC and RVD) by segmentation method. Values show median \pm standard deviation. Asterisk indicates values from the literature, for a comparison; N/A stands for not available.

Segmentation Method	DSC	RVD
SUV 2.5	0.677 \pm 0.172	89.9% \pm 108.4
40% SUVmax	0.706 \pm 0.130	−36.9% \pm 100.8
50% SUVmax	0.595 \pm 0.138	−56.1% \pm 57.4
Nestle	0.762 \pm 0.107	−32.7% \pm 40.0
Click and draw	0.792 \pm 0.178	−21.6% \pm 1270.8
Click 40	0.701 \pm 0.210	79.7% \pm 2451.1
Bi-modal 3D U-Net * [31]	0.717 (mean)	N/A
3D nnU-Net * [21]	0.774 (mean)	N/A

Table 3. Pairwise comparison of methods (Mann–Whitney U test). Statistically significant results are marked with an asterisk followed by the direction of the inequality; $\bar{\alpha}$ indicates Bonferroni-corrected α value.

Segmentation Mode A	Segmentation Mode B	$\bar{\alpha}$	DSC	RVD
Click 40	Click and draw	0.003	<0.001 * (A < B)	<0.001 * (A > B)
Click 40	Nestle	0.003	0.004	<0.001 * (A > B)
Click 40	SUV 2.5	0.003	0.570	0.862
Click 40	40% SUVmax	0.003	0.564	<0.001 * (A > B)
Click 40	50% SUVmax	0.003	<0.001 * (A > B)	<0.001 * (A > B)
Click and draw	Nestle	0.003	0.223	0.008
Click and draw	SUV 2.5	0.003	<0.001 * (A > B)	<0.001 * (A < B)
Click and draw	40% SUVmax	0.003	<0.001 * (A > B)	<0.001 * (A > B)
Click and draw	50% SUVmax	0.003	<0.001 * (A > B)	<0.001 * (A > B)
Nestle	SUV 2.5	0.003	<0.001 * (A > B)	<0.001 * (A < B)
Nestle	40% SUVmax	0.003	0.009	0.426
Nestle	50% SUVmax	0.003	<0.001 * (A > B)	<0.001 * (A > B)
SUV 2.5	40% SUVmax	0.003	0.129	<0.001 * (A > B)
SUV 2.5	50% SUVmax	0.003	0.005	<0.001 * (A > B)
40% SUVmax	50% SUVmax	0.003	<0.001 * (A > B)	<0.001 * (A > B)

3.2. Relative Volume Difference

Relative volume difference (Figure 5, Table 2) indicating that all methods except Click 40 and SUV 2.5 had a tendency to underestimate the lesion volume. On the whole, click and draw and Nestle had the lowest RVD magnitude (respectively, -21.6% and -32.7%); SUV 2.5 the highest (89.9%). Pairwise RVD comparison (Table 3) highlighted significant differences between all pairs except between Click and draw and Nestle, Click 40 and SUV 2.5, and Nestle and 40% SUVmax. We again observe that Nestle had the lowest standard deviation (i.e., highest inter-patient stability) in terms of RVD than the other methods; click 40 and click and draw the highest variability.

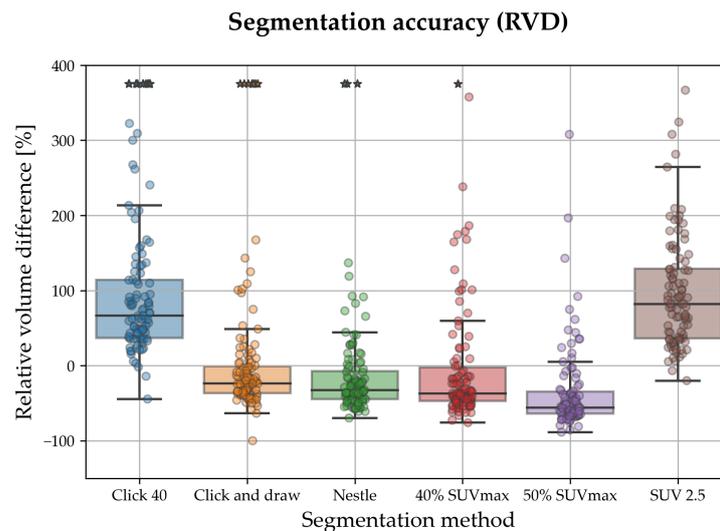


Figure 5. Box plots/strip plots showing relative volume difference (RVD) for the semi-automated segmentation methods considered in this study. Colour and x tick labels denote segmentation method; each dot in the strip plots represents one lesion/patient, stars indicate outliers (plotted out of scale). Note that all the methods except click 40 and SUV 2.5 tended to underestimate the lesion volume; on the whole, click and draw and Nestle had the lowest RVD magnitude.

3.3. Time Savings

A quantitative evaluation of the time required for manual segmentation, semi-automated ‘click and draw’ segmentation and semi-automated ‘draw and refine’ segmentation (see also Table 1 for a recap) were also carried out on a sample of five cases. The results are reported in Table 4. As can be seen, the time saving achieved by semi-automated methods was remarkable.

Table 4. Time savings: values show the total segmentation time (excluding the data loading time) for each case and by the class of segmentation method. The volume of the manually delineated ROI is also reported.

Case ID	Segmentation Time			Volume
	Manual	Click and Draw	Draw and Refine	
8229	67 s	9 s	18 s	3.0 cm ³
16f6	75 s	8 s	14 s	2.3 cm ³
a174	92 s	9 s	13 s	6.0 cm ³
58b2	88 s	11 s	21 s	12.2 cm ³
b6d2	308 s	11 s	22 s	38.3 cm ³
Avg	126.0 s	9.6 s	17.6 s	

4. Discussion

Head and neck tumour volume delineation on FDG PET/CT plays a pivotal role in clinical decisions such as risk stratification, prediction of response to therapy and radiotherapy planning. When performed manually, this is a lengthy and tedious procedure which can put significant strain on clinical personnel which are often already overworked. Consequently, several automated and semi-automated methods to generate regions of interest on PET FDG have been proposed in the literature. Although the subject has received significant attention in other oncological disorders—e.g., lung cancer [27,34,45]—few works have addressed HNC. In particular, we are not aware of any previous benchmarks comparing the performance of semi-automated methods for lesion delineation of HNC lesions on FDG PET/CT. A study by Zaidi et al. [29] assessed nine methods for segmenting pharyngolaryngeal SCC on FDG PET/CT using frozen tumour specimens as a reference; while, in [25], the authors investigated the stability of radiomics features to image segmentation and signal discretisation, but the focus was on feature robustness, not on goodness of segmentation.

In this work, we compared six training-free methods for the semi-automated lesion delineation of HNC tumour regions on FDG PET/CT. With an eye on the practical implications and for reproducible research, we carried out the study using a widespread free radiomics tool (LIFEX). We found that click and draw and Nestle methods achieved the best accuracy, albeit the first at the cost of noticeable inter-patient variability and the presence of outliers towards the low-accuracy region (DSC close to zero in some cases).

We speculate that, if a voxel exhibits an unusually large uptake value, this could lead to an overestimated threshold, and, as a result, an underestimated volume—which is what happens with 40% and 50% SUV_{max}. The same reasoning applies to the methods of the ‘click and draw’ class. Here, the problem is that if the voxel selected as the starting point has an unusually high or low uptake, this is likely to hamper the flood-fill algorithm and produce incorrect ROIs. This could explain why click and draw and click 40 fail in some cases (see Figure 4). In contrast, the Nestle method, in which threshold calculation is mediated by background uptake, emerged as the best trade-off between the overall accuracy and robustness.

In a recent multi-centre study [21], the authors used convolutional neural networks (CNNs) to achieve an aggregate DSC of 0.774, a figure just in between that obtained here with click and draw and Nestle; whereas Andrearczyk et al. reported an average DSC of 0.717 using a bi-modal U-Net model [31]. A comparison with the recent literature therefore indicates that the accuracy attained by our two best methods was in agreement with state-of-the-art approaches. We conclude observing that a fixed SUV 2.5 threshold, a commonly used approach [34,35], should be considered with care in HNC as it consistently overestimated the tumour volume by a large margin. This is likely the consequence of the ‘spillover’ effect as already discussed in [34].

The extent to which the accuracy of tumour segmentation can impact clinical practice, whether in terms of time-saving benefits or patient outcomes, is a matter for reflection. Surely one tangible advantage of computer-assisted lesion delineation is time saving, which should help streamline clinical workflows and reduce the risk of overwork. To better demonstrate the potential time benefits, we assessed the time required for manual segmentation, semi-automated ‘click and draw’ segmentation and semi-automated ‘draw and refine’ segmentation. The analysis was carried out on a sample of five cases and the results are presented in Table 4. As can be seen, the amount of time saved by semi-automated segmentation was remarkable.

In a recent work, Buteau et al. [46] investigated the influence of automated segmentation to quantify the total tumour burden (TTB) in 68Ga-PSMA-11 PET/CT scans of patients undergoing radioligand therapy (RLT) with 177Lu-PSMA. The authors concluded that automated segmentation significantly reduced the amount of time required for TTB quantification while not producing significant differences in the SUV_{mean}. With a focus on patient outcome, a study by Li et al. [47] (MRI in colorectal cancer) demonstrated the capability of an automated segmentation tool (DeepTOP) to accurately segment tumours and predict a pathologically complete response to chemo/radiotherapy. However, large-scale

clinical investigations are needed to better define the clinical relevance of semi-automated methods, especially when assisted by AI-based algorithms.

In conclusion, the main findings of the present investigation can be summarised as follows:

- Semi-automated, training-free approaches can greatly help the clinician delineate HNC lesions on FDG PET/CT;
- Among the methods considered in this study, Nestle was the best trade-off between accuracy and robustness;
- The time saving granted by semi-automated methods can be considerable;
- Fixed SUV 2.5 threshold segmentation (commonly used in the practice) should be considered with care in HNC as it consistently overestimated the tumour volume in our study.

5. Limitations and Future Work

There are some limitations to this work, including the relatively small sample size, the single-centre population and the retrospective nature of the study. The results should be therefore validated in larger, ideally perspective and multi-centre cohorts of patients. The scan images were used ‘as is’, therefore the potential effects of noise, filtering and/or system artefacts were considered outside the scope of the present work. Furthermore, by design, we used one single ROI as the ground-truth segmentation, hence inter-observer variability was not assessed in this study. All of the above would represent an interesting subject for future investigations.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/s23187952/s1>, Table S1: By-patient results.

Author Contributions: Conceptualisation, F.B., R.S., S.N., A.S. and B.P.; Data curation, R.S., M.U.K., M.M., A.M., S.N. and A.S.; Formal analysis, F.B., L.F. and B.P.; Funding acquisition, F.B.; Investigation, F.B., R.S., A.M. and B.P.; Methodology, F.B., R.S., M.L.F., M.M. and S.N.; Software, F.B. and M.U.K.; Supervision, M.L.F., M.M., L.F., S.N., A.S. and B.P.; Validation, F.B., M.L.F., L.F., A.M., A.S. and B.P.; Visualisation, M.U.K.; Writing—original draft, F.B.; Writing—review and editing, R.S. and B.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the Department of Engineering, Università degli Studi di Perugia, Italy, under the fundamental research grant scheme D.D. 132/2022 within project STARTLE—Shape and Texture Analysis of Planar and Volumetric models: Methods and Applications. Muhammad Usama Khan’s Ph.D. scholarship ICT technologies for health support was funded by the Italian Ministry of University and Research (MUR) under the National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza—PNRR), action D.M. 351/2022.

Institutional Review Board Statement: Patients underwent PET/CT scan for clinical purposes (HNC cancer staging) in compliance with the international guidelines. All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Formal ethical approval was not required due to the retrospective nature of the study and the analysis of anonymous clinical data.

Informed Consent Statement: Written informed consent was obtained from all subjects. In the informed consent, patients signed an item accepting that their data could be used for scientific purposes in anonymised form.

Data Availability Statement: Anonymised by-patient results are provided as supplementary material (anonymised_by_patient_accuracy.csv). Original FDG PET/CT scans are not available due to privacy restrictions.

Acknowledgments: The corresponding author wishes to thank Constantino Carlos Reyes-Aldasoro (City, University of London, United Kingdom) for his valuable remarks and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional neural network(s)
EBV	Epstein–Barr virus
FDG PET/CT	Positron emission tomography/computed tomography with [¹⁸ F] fluorodeoxyglucose
HNSCC	Head and neck squamous cell carcinoma
HNC	Head and neck cancer
HPV	Human papillomavirus
MRI	Magnetic resonance imaging
PSMA	Prostate-specific membrane antigen
ROI	Region(s) of interest
SUV	Standardised uptake value
TTB	Total tumour burden

References

- Gormley, M.; Creaney, G.; Schache, A.; Ingarfield, K.; Conway, D.I. Reviewing the epidemiology of head and neck cancer: definitions, trends and risk factors. *Br. Dent. J.* **2022**, *233*, 780–786. [[CrossRef](#)] [[PubMed](#)]
- Merlano, M.C.; Garrone, O. Vie aerodigestive superiori (Testa e collo). In *I Numeri del Cancro in Italia*; Intermedia Editore: Brescia, Italy, 2022; Chapter 3.1, pp. 49–50.
- Hashim, D.; Genden, E.; Posner, M.; Hashibe, M.; Boffetta, P. Head and neck cancer prevention: From primary prevention to impact of clinicians on reducing burden. *Ann. Oncol.* **2019**, *30*, 744–756. [[CrossRef](#)] [[PubMed](#)]
- Migliaro, M.; Massuh, D.; Infante, M.F.; Brahm, A.M.; San Martín, M.T.; Ortuño, D. Role of Epstein–Barr virus and human papilloma virus in the development of oropharyngeal cancer: A literature review. *Int. J. Dent.* **2022**, *2022*, 3191569. [[CrossRef](#)]
- Eyassu, E.; Young, M. *Chapter Nuclear Medicine PET/CT Head and Neck Cancer Assessment, Protocols, and Interpretation*; StatPearls Publishing: Treasure Island, FL, USA, 2022.
- Ferrari, C.; Santo, G.; Mammucci, P.; Rubini, D.; Sciacqua, A.; Sardaro, A.; Pisani, A.R.; Rubini, G. [¹⁸F]FDG PET/CT in head and neck squamous cell carcinoma: A head-to-head between visual point-scales and the added value of multi-modality imaging. *BMC Med. Imaging* **2023**, *23*, 34. [[CrossRef](#)] [[PubMed](#)]
- Gillies, R.; Kinahan, P.; Hricak, H. Radiomics: Images are more than pictures, they are data. *Radiology* **2016**, *278*, 563–577. [[CrossRef](#)] [[PubMed](#)]
- Thawani, R.; McLane, M.; Beig, N.; Ghose, S.; Prasanna, P.; Velcheti, V.; Madabhushi, A. Radiomics and radiogenomics in lung cancer: A review for the clinician. *Lung Cancer* **2018**, *115*, 34–41. [[CrossRef](#)] [[PubMed](#)]
- Vial, A.; Stirling, D.; Field, M.; Ros, M.; Ritz, C.; Carolan, M.; Holloway, L.; Miller, A. The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: A review. *Transl. Cancer Res.* **2018**, *7*, 803–816. [[CrossRef](#)]
- Hosny, A.; Aerts, H.J.; Mak, R.H. Handcrafted versus deep learning radiomics for prediction of cancer therapy response. *Lancet Digit. Health* **2019**, *1*, e106–e107. [[CrossRef](#)]
- Demircioğlu, A. Are deep models in radiomics performing better than generic models? A systematic review. *Eur. Radiol. Exp.* **2023**, *7*, 11. [[CrossRef](#)] [[PubMed](#)]
- Abbasian Ardakani, A.; Bureau, N.J.; Ciaccio, E.J.; Acharya, U.R. Interpretation of radiomics features—A pictorial review. *Comput. Methods Programs Biomed.* **2022**, *215*, 106609. [[CrossRef](#)] [[PubMed](#)]
- Fornacon-Wood, I.; Mistry, H.; Ackermann, C.J.; Blackhall, F.; McPartlin, A.; Faivre-Finn, C.; Price, G.J.; O'Connor, J.P.B. Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. *Eur. Radiol.* **2020**, *30*, 6241–6250. [[CrossRef](#)] [[PubMed](#)]
- Bianconi, F.; Palumbo, I.; Spanu, A.; Nuvoli, S.; Fravolini, M.L.; Palumbo, B. PET/CT radiomics in lung cancer: An overview. *Appl. Sci.* **2020**, *5*, 1718. [[CrossRef](#)]
- Bianconi, F.; Palumbo, I.; Fravolini, M.L.; Rondini, M.; Ministrini, M.; Pascoletti, G.; Nuvoli, S.; Spanu, A.; Scialpi, M.; Aristei, C.; et al. Form factors as potential imaging biomarkers to differentiate benign vs. malignant lung lesions on CT scans. *Sensors* **2022**, *22*, 5044. [[CrossRef](#)] [[PubMed](#)]
- Tortora, M.; Gemini, L.; Scaravilli, A.; Ugga, L.; Ponsiglione, A.; Stanzione, A.; D'Arco, F.; D'Anna, G.; Cuocolo, R. Radiomics Applications in Head and Neck Tumor Imaging: A Narrative Review. *Cancers* **2023**, *15*, 1174. [[CrossRef](#)] [[PubMed](#)]
- Liu, Z.; Cao, Y.; Diao, W.; Cheng, Y.; Jia, Z.; Peng, X. Radiomics-based prediction of survival in patients with head and neck squamous cell carcinoma based on pre-and post-treatment ¹⁸F-PET/CT. *Aging* **2020**, *12*, 14593–14619. [[CrossRef](#)] [[PubMed](#)]
- Xie, C.; Du, R.; Ho, J.W.; Pang, H.H.; Chiu, K.W.; Lee, E.Y.; Vardhanabhuti, V. Effect of machine learning re-sampling techniques for imbalanced datasets in ¹⁸F-FDG PET-based radiomics model on prognostication performance in cohorts of head and neck cancer patients. *Eur. J. Nucl. Med. Mol. Imaging* **2020**, *47*, 2826–2835. [[CrossRef](#)]
- Lv, W.; Feng, H.; Du, D.; Ma, J.; Lu, L. Complementary value of intra-and peri-tumoral PET/CT radiomics for outcome prediction in head and neck cancer. *IEEE Access* **2021**, *9*, 81818–81826. [[CrossRef](#)]

20. Wang, B.; Liu, J.; Zhang, X.; Wang, Z.; Cao, Z.; Lu, L.; Lv, W.; Wang, A.; Li, S.; Wu, X.; et al. Prognostic value of ^{18}F -FDG PET/CT-based radiomics combining dosiomics and dose volume histogram for head and neck cancer. *EJNMMI Res.* **2023**, *13*, 14. [[CrossRef](#)]
21. Salahuddin, Z.; Chen, Y.; Zhong, X.; Woodruff, H.C.; Rad, N.M.; Mali, S.A.; Lambin, P. From Head and Neck Tumour and Lymph Node Segmentation to Survival Prediction on PET/CT: An End-to-End Framework Featuring Uncertainty, Fairness, and Multi-Region Multi-Modal Radiomics. *Cancers* **2023**, *15*, 1932. [[CrossRef](#)]
22. Martens, R.M.; Koopman, T.; Noij, D.P.; Pfaehler, E.; Übelhör, C.; Sharma, S.; Vergeer, M.R.; Leemans, C.R.; Hoekstra, O.S.; Yaqub, M.; et al. Predictive value of quantitative ^{18}F -FDG-PET radiomics analysis in patients with head and neck squamous cell carcinoma. *EJNMMI Res.* **2020**, *10*, 102. [[CrossRef](#)]
23. Bogowicz, M.; Riesterer, O.; Stark, L.S.; Studer, G.; Unkelbach, J.; Guckenberger, M.; Tanadini-Lang, S. Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma. *Acta Oncologica* **2017**, *56*, 1531–1536. [[CrossRef](#)] [[PubMed](#)]
24. Zhang, Q.; Wang, K.; Zhou, Z.; Qin, G.; Wang, L.; Li, P.; Sher, D.; Jiang, S.; Wang, J. Predicting local persistence/recurrence after radiation therapy for head and neck cancer from PET/CT using a multi-objective, multi-classifier radiomics model. *Front. Oncol.* **2022**, *12*, 955712. [[CrossRef](#)] [[PubMed](#)]
25. Lu, L.; Lv, W.; Jiang, J.; Ma, J.; Feng, Q.; Rahmim, A.; Chen, W. Robustness of Radiomic Features in ^{11}C Choline and ^{18}F FDG PET/CT Imaging of Nasopharyngeal Carcinoma: Impact of Segmentation and Discretization. *Mol. Imaging Biol.* **2016**, *18*, 935–945. [[CrossRef](#)] [[PubMed](#)]
26. Varghese, A.J.; Gouthamchand, V.; Sasidharan, B.K.; Wee, L.; Sidhique, S.K.; Rao, J.P.; Dekker, A.; Hoebbers, F.; Devakumar, D.; Irodi, A.; et al. Multi-centre radiomics for prediction of recurrence following radical radiotherapy for head and neck cancers: Consequences of feature selection, machine learning classifiers and batch-effect harmonization. *Phys. Imaging Radiat. Oncol.* **2023**, *26*, 100450. [[CrossRef](#)] [[PubMed](#)]
27. Mercieca, S.; Belderbos, J.; van Loon, J.; Gilhuijs, K.; Julyan, P.; van Herk, M. Comparison of SUVmax and SUVpeak based segmentation to determine primary lung tumour volume on FDG PET-CT correlated with pathology data. *Radiother. Oncol.* **2018**, *129*, 227–233. [[CrossRef](#)]
28. Tamal, M. Intensity threshold based solid tumour segmentation method for Positron Emission Tomography (PET) images: A review. *Heliyon* **2020**, *6*, e05267. [[CrossRef](#)] [[PubMed](#)]
29. Zaidi, H.; Abdoli, M.; Fuentes, C.L.; El Naqa, I.M. Comparative methods for PET image segmentation in pharyngolaryngeal squamous cell carcinoma. *Eur. J. Nucl. Med. Mol. Imaging* **2012**, *39*, 881–891. [[CrossRef](#)]
30. Trada, Y.; Lin, P.; Lee, M.T.; Jameson, M.G.; Chlap, P.; Keall, P.; Moses, D.; Fowler, A. Impact of tumour region of interest delineation method for mid-treatment FDG-PET response prediction in head and neck squamous cell carcinoma undergoing radiotherapy. *Quant. Imaging Med. Surg.* **2023**, *13*, 2822–2836. [[CrossRef](#)] [[PubMed](#)]
31. Andrearczyk, V.; Oreiller, V.; Jreige, M.; Castelli, J.; Prior, J.O.; Depeursinge, A. Segmentation and classification of head and neck nodal metastases and primary tumors in PET/CT. In Proceedings of the 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, Glasgow, UK, 11–15 July 2022; Volume 2022, pp. 4731–4735.
32. Picchio, M.; Crippa, F. *Raccomandazioni Procedurali per l'Imaging Oncologico Con ^{18}F -FDG PET/TC*; Technical Report; Italian Society for Nuclear Medicine and Molecular Imaging (AIMN): Milan, Italy, 2017; Revision 4/2017.
33. Nioche, C.; Orhlac, F.; Boughdad, S.; Reuze, S.; Goya-Outi, J.; Robert, C.; Pellot-Barakat, C.; Soussan, M.; Frouin, F.E.; Buvat, I. LIFEx: A freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res.* **2018**, *78*, 4786–4789. [[CrossRef](#)] [[PubMed](#)]
34. Im, H.J.; Bradshaw, T.; Solaiyappan, M.; Cho, S.Y. Current methods to define metabolic tumor volume in positron emission tomography: Which one is better? *Nucl. Med. Mol. Imaging* **2018**, *52*, 5–15. [[CrossRef](#)] [[PubMed](#)]
35. Rijo-Cedeño, J.; Mucientes, J.; Villarreal, I.M.; Royuela, A.; García Vicente, P.; García-Berrocá, J.R. Utility and limitations of metabolic parameters in head and neck cancer: Finding a practical segmentation method. *Eur. Arch. Oto-Rhino-Laryngol.* **2022**, *279*, 4577–4586. [[CrossRef](#)]
36. Nestle, U.; Kremp, S.; Schaefer-Schuler, A.; Sebastian-Welsch, C.; Hellwig, D.; Rube, C.; Kirsch, C. Comparison of different methods for delineation of ^{18}F -FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-small cell lung cancer. *J. Nucl. Med.* **2005**, *46*, 1342–1348.
37. Nioche, C.; Orhlac, F.; Buvat, I. Local Image Feature Extraction–LIFEx. User Guide. 2023. Available online: <https://www.lifexsoft.org/images/phocagallery/documentation/LIFEx/UserGuide/LIFExUserGuide.pdf> (accessed on 31 May 2023)
38. Maisonobe, J.; Garcia, C.A.; Necib, H.; Vanderlinden, B.; Hendlisz, A.; Flamen, P.; Buvat, I. Comparison of PET metabolic indices for the early assessment of tumour response in metastatic colorectal cancer patients treated by polychemotherapy. *Eur. J. Nucl. Med. Mol. Imaging* **2013**, *40*, 166–174. [[CrossRef](#)]
39. Smith, A.R. Tint fill. *ACM SIGGRAPH Comput. Graph.* **1979**, *13*, 276–283. [[CrossRef](#)]
40. Heimann, T.; Van Ginneken, B.; Styner, M.A.; Arzhaeva, Y.; Aurich, V.; Bauer, C.; Beck, A.; Becker, C.; Beichel, R.; Bekes, G.; et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans. Med. Imaging* **2009**, *28*, 1251–1265. [[CrossRef](#)] [[PubMed](#)]
41. Yeghiazaryan, V.; Voiculescu, I. Family of boundary overlap metrics for the evaluation of medical image segmentation. *J. Med. Imaging* **2018**, *5*, 015006. [[CrossRef](#)] [[PubMed](#)]

42. Bianconi, F.; Fravolini, M.; Pizzoli, S.; Palumbo, I.; Minestrini, M.; Rondini, M.; Nuvoli, S.; Spanu, A.; Palumbo, B. Comparative evaluation of conventional and deep learning methods for semi-automated segmentation of pulmonary nodules on CT. *Quant. Imaging Med. Surg.* **2021**, *11*, 3286–3305. [CrossRef] [PubMed]
43. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*, 29. [CrossRef] [PubMed]
44. Moore, C.; Bell, D. Dice Similarity Coefficient. Radiopaedia.org. Available online: <https://radiopaedia.org/articles/dice-similarity-coefficient> (accessed on 27 August 2023).
45. Wang, X.Y.; Zhao, Y.F.; Liu, Y.; Yang, Y.K.; Zhu, Z.; Wu, N. Comparison of different automated lesion delineation methods for metabolic tumor volume of 18 F-FDG PET/CT in patients with stage I lung adenocarcinoma. *Medicine* **2017**, *96*, e9365. [CrossRef]
46. Buteau, J.P.; Niman, R.; Wilson, P.; Susman, T.; Nelson, A.; Alipour, R.; Hofman, M. Time-savings analysis of total tumor burden quantification on 68Ga-PSMA-11 PET/CT with deep learning auto-segmentation of organs for automatic physiological uptake removal in men with metastatic castration-resistant prostate cancer (mCRPC). *J. Nucl. Med.* **2022**, *63*, 2205.
47. Li, L.; Xu, B.; Zhuang, Z.; Li, J.; Hu, Y.; Yang, H.; Wang, X.; Lin, J.; Zhou, R.; Chen, W.; et al. Accurate tumor segmentation and treatment outcome prediction with DeepTOP. *Radiother. Oncol.* **2023**, *183*, 109550. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.