Head and Neck target delineation using a novel PET automatic segmentation algorithm

B. Berthon[1a], M. Evans[2], C. Marshall[1], N. Palaniappan[2], N. Cole[2], V. Jayaprakasam[2], T. Rackley[2], E. Spezi[2,3]

1. Wales Research & Diagnostics PET Imaging Centre, Cardiff, UK
2. Velindre Cancer Centre, Cardiff, UK
3. School of Engineering, Cardiff University, Cardiff, UK

Keywords: Positron Emission Tomography, Image Segmentation, Intensity Modulated Radiation Therapy, Automatic PET segmentation

Corresponding author details:

Beatrice Berthon
Institut Langevin
17 rue Moreau, 75012 Paris
France
Tel: +33 1 80 96 30 88
e-mail: beatrice.walker@espci.fr

Number of pages :19
Number of tables :2
Number of figures :2

Running head: Novel PET segmentation for H&N IMRT

[a] Dr. Berthon is now at Institut Langevin, ESPCI Paris, PSL Research University, CNRS UMR 7587, INSERM U979

Abstract

Purpose: To evaluate the feasibility and impact of using a novel advanced PET auto-segmentation method in Head and Neck (H&N) radiotherapy treatment (RT) planning.

Methods: ATLAAS, Automatic decision Tree-based Learning Algorithm for Advanced Segmentation, previously developed and validated on pre-clinical data, was applied to ^{18}F-FDG-PET/CT scans of 20 H&N patients undergoing Intensity Modulated Radiation Therapy. Primary Gross Tumour Volumes (GTVs) manually delineated on CT/MRI scans ($GTVp_{CT/MRI}$), together with ATLAAS-generated contours ($GTVp_{ATLAAS}$) were used to derive the RT planning GTV ($GTVp_{final}$). ATLAAS outlines were compared to CT/MRI and final GTVs qualitatively and quantitatively using a conformity metric.

Results: The ATLAAS contours were found to be reliable and useful. The volume of $GTVp_{ATLAAS}$ was smaller than $GTVp_{CT/MRI}$ in 70% of the cases, with an average conformity index of 0.70. The information provided by ATLAAS was used to grow the $GTVp_{CT/MRI}$ in 10 cases (up to 10.6 mL) and to shrink the $GTVp_{CT/MRI}$ in 7 cases (up to 12.3 mL). ATLAAS provided complementary information to CT/MRI and $GTVp_{ATLAAS}$ contributed to up to 33% of the final GTV volume across the patient cohort.

Conclusions: ATLAAS can deliver operator independent PET segmentation to augment clinical outlining using CT and MRI and could have utility in future clinical studies.

1. INTRODUCTION

Positron Emission Tomography (PET) imaging using 18F-Fluorodeoxyglucose (FDG) plays an increasingly valuable role in Radiotherapy Treatment (RT) planning for a number of cancers [1]. Loco-regional recurrences have been shown to correlate with PET-avid volumes [2], with studies demonstrating the feasibility and usefulness of PET/CT-guided Intensity Modulated Radiation Therapy (IMRT) [3]. PET/CT-based outlining can lead to more accurate and reproducible delineation of the Gross Tumour Volume (GTV), compared to outlining done using CT alone [4]. The PET-based GTV is usually smaller than the CT based volume [5], [6]. Nishioka *et al.* showed with 21 oropharyngeal and nasopharyngeal cancer patients that adjacent normal tissue, particularly parotids, could be spared in 71% of patients when using PET in the delineation [7], which could potentially lead to reduced long term morbidity, xerostomia and improved quality of life.

Although FDG-PET has been adopted in oncology as a key tool in diagnostic imaging, its use in RT planning has, until now, been limited due to a lack of consensus on

GTV delineation method. The low resolution of PET coupled with the proximity to the tumour of other metabolically active structures make the delineation challenging. In particular in the Head and Neck (H&N), organs such as the pharyngeal muscles, spinal cord and salivary glands, which should be spared to minimise morbidity and improve quality of life, can generate additional background FDG uptake.

Manual PET-based GTV delineation, currently used in most centres, is time consuming and highly operator-dependent and several studies have shown significant variations in the GTV delineated by different operators using PET [5], [8]. This has led to the development and recommended use of various PET automatic segmentation (PET-AS) methods for H&N [9]. However, only a small number of prospective clinical studies have reported on the use of PET-AS in RT planning [6]. Comparing different studies is difficult because of the different PET-AS methods used. Basic thresholding methods lack accuracy and reliability [10], [11], but more advanced PET-AS methods, such as gradient-based, clustering or region-growing approaches are rarely used, and their impact on RT planning is still unclear. There is a need for studies investigating the feasibility and clinical benefits of using advanced PET-AS in RT planning.

This prospective study investigated the use of an optimised PET-AS tool, developed and validated in house using phantom and clinical PET data [12], [13], for GTV delineation in the RT planning of 20 oropharyngeal cancer patients. We evaluated the feasibility and impact of including this method into the RT planning process.

2. METHODS

2.A. THE ATLAAS OPTIMISED SEGMENTATION MODEL

PET-AS was performed using the Automatic decision Tree-based Learning Algorithm for Advanced Segmentation (ATLAAS)[b] method developed at our centre. The

[b] Patent pending No PCT/GB2015/052981

ATLAAS model is designed to select the most accurate PET-AS method for a given PET image. This is achieved using a decision tree supervised machine learning method, optimised with a training dataset for which the segmentation outcome is known, to achieve optimal performance for cases in which the outcome is not known. ATLAAS is described elsewhere [14], and its accuracy was shown for 6 classes of advanced PET-AS methods used to segment a large range of data including simulated H&N tumours, and phantom H&N images of complex and realistic tumours obtained with a sub-resolution printed sandwich phantom [15]. ATLAAS was optimised for H&N data using 65 sub-resolution printed sandwich phantom images. The optimised version included the two algorithms Adaptive Thresholding method (AT) and Gaussian mixture models Clustering Method using 5 clusters (GCM5), described in previous work [12]. The best method was predicted on the basis of TBR_{peak} defined as the ratio between the tumour peak intensity value, (mean value in a 1 cm^3 sphere centred on the maximum intensity voxel) and the background intensity (mean intensity in a 1 cm thick extension of a thresholded volume at 50% of the peak intensity value). An example of the typical steps involved in the segmentation with ATLAAS is given in **Error! Reference source not found.**. The ATLAAS model was implemented for this work in the Computational Environment for Radiotherapy Research (CERR)[16]. The segmentation accuracy was evaluated by quantifying the overlap between the segmented and true contour using the Dice Similarity Coefficient (DSC) described in other work [17].

2.B. ACQUISITION OF CLINICAL DATA

The POSITIVE (Optimization of Positron Emission Tomography based Target Volume Delineation in Head and Neck Radiotherapy) study was set up to test ATLAAS for the first time in patients undergoing H&N radiotherapy (REC No. 12/WA/0083) and was carried out at Velindre Cancer Centre (UK). Twenty stage III/IVa-b oropharyngeal cancer patients were recruited after informed consent to the study. The patients were treated with neoadjuvant (induction) chemotherapy followed by radical chemoradiotherapy (66 Gy in 30 fractions over 6 weeks) using IMRT. A planning FDG PET/CT scan was carried out on a GE Discovery 690 PET/CT scanner before chemotherapy to avoid changes in tumour volumes prior to outlining. The scans were acquired 90 minutes after FDG administration in the treatment position with an RT immobilisation shell. The PET was acquired using 6-8 bed positions of 3 min each. The patient was injected with contrast for a subsequent CT used in the planning process. The images were reconstructed to 512 x 512 voxels for CT and 256 x 256 voxels for PET, using the algorithm Vue Point FX (24 subsets, 2 iterations, 6.4 mm cut-off) including CT-based attenuation-, scatter- and Time-Of-Flight corrections.

Six weeks on average separated the PET/CT planning scan and the start of RT. The fit of the immobilisation shells was adjusted if needed after induction chemotherapy and the patient was re-outlined and re-planned if necessary using the original CT/MRI/PET scan. Reporting was done by PET specialist radiologists after acquisition of the planning scans.

2.C. WORKFLOW AND ANALYSIS

MRI scans acquired before recruitment were available for all patients and were fused to the planning PET-CT scan using the Mutual Information registration algorithm in the ProSoma software (MedCom GmbG, Darmstadt, Germany).

Planning scans for the first 10 patients recruited were used to validate the workflow and verify that ATLAAS provided relevant contours for use. In this subgroup the primary GTVs were manually outlined by three consultant radiation oncologists, in discussion with a specialist PET radiologist, on the registered PET/CT, using the software VelocityAI (Varian Medical Systems, Palo Alto, USA). The resulting $GTVp_{PET/CT}$ contours were compared with ATLAAS contours in terms of their volume and geometrical overlap, using the DSC index.

Once the ATLAAS output was verified, another 10 oropharyngeal cancer patients were recruited for the study. Manual delineation of the primary GTV was performed by the consultant radiation oncologists on the fused MRI and CT images in ProSoma ($GTVp_{CT/MRI}$). ATLAAS ($GTVp_{ATLAAS}$) contours were then imported into ProSoma where the final GTV ($GTVp_{final}$) was drawn by the treating clinician using all the available contour data.

The use of the additional information brought by ATLAAS contours was evaluated by comparing the different contours ($GTVp_{CT/MRI}$, $GTVp_{ATLAAS}$, and $GTVp_{final}$) for each patient, in terms of volume and geometrical overlap using the DSC. In addition, the clinicians were asked to report any changes made to the $GTVp_{final}$ due to the ATLAAS contour. Lymph nodes, which are well defined on CT/MRI images, were not outlined using ATLAAS, and are therefore not reported on in this paper.

3. RESULTS

The patient cohort included 17 men and 3 women with a median age of 63 years. Ten patients had tonsillar tumours, 8 base of tongue tumours and 2 soft palate tumours. Two patients needed re-planning after induction chemotherapy.

In the preliminary group of 10 patients, ATLAAS successfully delineated the PET-avid tumour for all patients. The segmentation of the tumour ROI was fully automatic and took no more than 2 minutes on a dual core 3.1 GHz processor. $GTVp_{ATLAAS}$ were smaller than the manually delineated $GTVp_{PET/CT}$ for 7 out of 10 patients. The mean DSC between $GTVp_{PET/CT}$ and $GTVp_{ATLAAS}$ was 0.82, when 0.7 is considered to be an indicator of good overlap [18]. On the basis of these results, it was decided that only ATLAAS and CT/MRI contours would be used for the subsequent 10 patients recruited.

A comparison in terms of volume and conformity between the GTVp delineated using ATLAAS and both CT/MRI-based and final contours delineated by the investigators is presented in **Error! Reference source not found.**. ATLAAS volumes were smaller than the corresponding CT/MRI volumes in 7 out of 10 cases, and were within 10% of CT/MRI volumes in 4 out of 10 cases. The spatial conformity of $GTVp_{ATLAAS}$ and $GTVp_{CT/MRI}$ was 0.70 DSC on average. $GTVp_{ATLAAS}$ and $GTVp_{final}$ were close, with the larger of the two no bigger than 30% of the smaller, in 6 out of the 10 cases. $GTVp_{final}$ volumes were larger than the $GTVp_{ATLAAS}$ in all cases. However, the ATLAAS volumes showed good conformity to the final contour, with an average DSC of 0.77.

Table 2 reports the details of the global and local changes to the final volume based on ATLAAS, and outlines the differences between ATLAAS and CT/MRI contours not taken into account in the final GTV. For instance, the data in the top row of the table shows that more than 83% of the ATLAAS volume was included in the final GTV for all patients, and 100% of the ATLAAS volume was included in the final GTV in 4 cases. The second row reports the proportion of the CT/MRI volume modified on the basis of the ATLAAS outline. This value ranged from 6.5% to 33%. This modification could include both additional extension of the volume when the ATLAAS contour was outside the $GTVp_{CT/MRI}$ or local reduction of the extension in cases where the inverse was true. This is detailed in rows 3-5 as illustrated under the table.

Figure 2 illustrates specific differences found between $GTVp_{CT/MRI}$, $GTVp_{ATLAAS}$ and $GTVp_{final}$ overlaid on the corresponding CT/PET scan, for seven clinical cases of interest.

3.A. EXTENDING THE GTV BASED ON ATLAAS

As reported in the third row of Table 2, $GTVp_{CT/MRI}$ was locally extended based on the information provided by ATLAAS (cf. Figure 2a) for all clinical cases, with up to 10 mL added to make the final volume. Visual examination and reporting by the clinicians showed that this was done when additional disease extension was detected by ATLAAS, and confirmed by clinical or CT/MRI findings. This included larger superior-inferior disease extension (for five patients and up to 1.1 cm as reported in Figure 2b), and disease extension identified across the midline (cf. Figure 2c).

3.B. REDUCING THE GTV EXTENT BASED ON ATLAAS

As shown in the fourth row of **Error! Reference source not found.**, local reduction of the extension (on one or more transverse slices) of the CT/MRI volume based on ATLAAS was observed for 7 patients, and was more than 2 mL for two patients. The extent of the contours was locally reduced when the smaller disease extension indicated by ATLAAS was in agreement with the clinical findings and the CT or the MRI information. The extension was also reduced in the superior-inferior direction for two patients (1.5 cm for patient No 16). In cases of largely conflicting information between image modalities, the CT/MRI contour extension was reduced down to a compromise following the edge of the anatomical structures, as depicted in Figure 2d.

3.C. ATLAAS INFORMATION DISCARDED

Differences between $GTVp_{CT/MRI}$ and $GTVp_{ATLAAS}$ were not considered in the final GTV when they included:

a) bone (0.1 mL for patient No 11, cf. Figure 2e),
b) air (for 5 patients, up to 6.6 mL for patient No 12, cf. Figure 2f)
c) different superior-inferior disease extension in $GTVp_{ATLAAS}$ which was not confirmed by anatomical imaging or clinical examinations (for 6 patients, up to 6.4 mL for patient No 13, cf. Figure 2g)
d) different transverse disease extension unconfirmed by anatomical imaging or clinical examinations (cf. some regions in Figure 2f)

In these cases, the differences between $GTVp_{ATLAAS}$ and $GTVp_{final}$ (expressed in mL), is given in row 5 of Table 2 and includes both over and under contouring.

4. DISCUSSION

In this study, we investigated the clinical feasibility of using the novel ATLAAS optimised segmentation model in 20 H&N cancer patients undergoing radical chemoradiotherapy. ATLAAS was applied for the first time to 20 prospectively recruited patients in a clinical trial with a strict scanning protocol, which involved expert PET radiologist and H&N radiation oncologists. It was prospectively used, in combination with manual CT/MRI data, to derive the final GTV for use in RT planning. To the best of our knowledge, advanced PET-AS methods (beyond simple thresholding) have only been included as part of RT treatment planning in two studies in H&N cancer [19], [20], which were based on the same segmentation method. In this work, we additionally evaluated the impact of using the PET-AS contour on local modifications of the planning contour.

ATLAAS had previously shown accuracy and robustness on phantom and simulated data for the evaluation of H&N PET scans [14]. Evaluation on images from the 10 first patients involved in this study showed that ATLAAS provided PET-avid GTVs for all patients with a high degree of similarity to PET GTVs manually delineated by experts. In addition, the segmentation was fully automatic and therefore reproducible, and lasted no more than 2 minutes per patient. The use of ATLAAS instead of manual PET/CT outlining for the 10 subsequent patients in this study, considerably reduced the clinicians' workload and removed inter-observer variability. We have shown that ATLAAS not only could segment the PET-avid areas of disease reliably in patients compared to manual PET outlining but that it could also add valuable information to guide clinical delineation of the primary GTV.

The ATLAAS contours were smaller than the CT/MRI contours in most cases, which is in agreement with findings from other studies where threshold-based delineation was used for H&N patients [21]. Furthermore, the ATLAAS derived contours provided additional information to anatomical contours manually drawn on CT and MRI. This is in line with the study by Newbold *et al*. in 19 H&N patients, where threshold-based delineation was used to derive the PET-based GTV [22]. In our study, we found that additional information from ATLAAS included (a) identification of superior-inferior disease extension , and extension across the midline not seen on CT (e.g. Figure 2c), and (b) other disease extension boundaries differing from anatomical data. The information provided by ATLAAS was used in all patients and this shows the confidence of the clinicians in the usefulness of our segmentation method for RT planning at our centre. The clinician's judgment and expertise and the additional clinical data available (endoscopy or clinical examination results) remained paramount in the process. Nevertheless ATLAAS was very useful (a) in confirming the GTV outline when this was close to the CT/MRI based contour, and (b) as a delineation guide when in disagreement with CT/MRI based contours, due for instance to different patient positioning and/or poor image registration.

We have methodically investigated the impact of ATLAAS on the final GTV for our cohort. We found that although ATLAAS led to reducing the extension in some areas of the $GTVp_{CT/MRI}$ for 7 patients, the PET information led to a globally smaller final GTV for only 1 patient. This is in line with the findings of Ciernik *et al.* for a cohort of 12 H&N patients [5], and Paulino *et al.* for 40 H&N patients [23], both using manual PET segmentation. This confirms the suggestion that clinicians may not be prepared yet to reduce the GTV volume based on PET. Indeed, although some studies have shown that PET-AS contours can accurately identify the whole tumour burden in laryngeal cancer [4], [24], it may be more useful for defining the metabolically active tumour region, especially for tumours which can be highly heterogeneous such as in the H&N [25]. This is in line with the suggestion of considering the Biological Tumor Volume as defined by Ling et al. [26], which can be used for dose escalation [27], [28] to increase the dose to the tumour while sparing the surrounding tissue. The ATLAAS model could be useful for determining, with a consistent and operator independent approach, highly metabolically active areas of the tumour requiring a radiation boost and it could therefore be extremely useful for treatment plan adaptation. Correlation with additional information and clinical input would still be required in finalising the volumes for dose escalation.

Differences between the $GTVp_{final}$ and $GTVp_{ATLAAS}$ volumes were in the range [0.6, 45] mL (cf. row 5 of Table 2). CT/MRI based outlining was preferred when: (a) no PET signal was found in abnormal mucosa (Figure 2g), (b) high PET uptake was observed in and around air cavities and/or bone (Figure 2c and 2h) due to signal spill-out or inflammation. Spill-out effects can be corrected with CT-based thresholding, whereas unconfirmed soft tissue extensions of the disease, which represent a large part of the differences observed between CT/MRI and PET contours (cf. rows 5c and 5d of Table 2), are inherent to the difference between modalities.

One of the limitations of this study is that we could not carry out a full comparison between $GTVp_{ATLAAS}$ and the PET GTV outlined manually without reference to anatomical data from the CT scan. In this case the correlation between manual and $GTVp_{ATLAAS}$ could have been greater because based on the same underlying data. In addition, this work was carried out as a single centre study. Both limitations shall be addressed in the design of a forthcoming multicentre clinical trial.

5. CONCLUSIONS

The ATLAAS optimised segmentation model based on the decision tree machine learning method is a novel, fast and operator independent tool for tumour delineation in radiotherapy treatment planning of Head and Neck cancer. ATLAAS can potentially be applied to any tumour site and tumour type and holds promise for future multi-centre clinical studies investigating the use of PET in radiotherapy outlining, prior to starting treatment and also for adaptive re-planning of residual metabolically active disease during treatment.

ACKNOWLEDGEMENTS

This work was funded by Cancer Research Wales grants No. 7061 and 2476. The funder was not involved in the study design, conception, data collection acquisition and analysis nor writing or submission of this manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] M. MacManus, U. Nestle, K. E. Rosenzweig, I. Carrio, C. Messa, O. Belohlavek, M. Danna, T. Inoue, E. Deniaud-Alexandre, S. Schipani, N. Watanabe, M. Dondi, and B. Jeremic, "Use of PET and PET/CT for radiation therapy planning: IAEA expert report 2006-2007.," *Radiother. Oncol.*, vol. 91, no. 1, pp. 85–94, Apr. 2009.

[2] A. K. Due, I. R. Vogelius, M. C. Aznar, S. M. Bentzen, A. K. Berthelsen, S. S. Korreman, A. Loft, C. A. Kristensen, and L. Specht, "Recurrences after intensity modulated radiotherapy for head and neck squamous cell carcinoma more likely to originate from regions with high baseline [18F]-FDG uptake.," *Radiother. Oncol.*, vol. 111, no. 3, pp. 360–5, Jun. 2014.

[3] D. Wang, C. Schultz, and P. Jursinic, "Initial experience of FDG-PET/CT guided IMRT of Head and Neck carcinoma," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 65, pp. 143–151, 2006.

[4] J.-F. Daisne, T. Duprez, B. Weynand, M. Lonneux, M. Hamoir, H. Reychler, and V.

Grégoire, “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, Oct. 2004.

[5] I. Ciernik, E. Dizendorf, B. Baumert, B. Reiner, C. Burger, J. Davis, U. Lutolf, H. Steinert, and G. Vonschulthess, “Radiation treatment planning with an integrated positron emission and computer tomography (PET/CT): a feasibility study,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 57, no. 3, pp. 853–863, Nov. 2003.

[6] L. Deantonio, D. Beldì, G. Gambaro, G. Loi, M. Brambilla, E. Inglese, and M. Krengli, “FDG-PET/CT imaging for staging and radiotherapy treatment planning of head and neck carcinoma.,” *Radiat. Oncol.*, vol. 3, p. 29, Jan. 2008.

[7] T. Nishioka, T. Shiga, H. Shirato, E. Tsukamoto, K. Tsuchiya, T. Kato, K. Ohmori, A. Yamazaki, H. Aoyama, S. Hashimoto, T.-C. Chang, and K. Miyasaka, “Image Fusion Between 18 FDG-PETAnd MRI/CT For Radiotherapy Planning Of Oropharyngeal And Nasopharyngeal Carcinomas,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 53, no. 4, pp. 1051–1057, 2002.

[8] C. B. Caldwell, K. Mah, Y. C. Ung, C. E. Danjoux, J. M. Balogh, S. N. Ganguli, and L. E. Ehrlich, “Observer variation in contouring gross tumor volume in patients with poorly defined non-small-cell lung tumors on CT: the impact of 18FDG-hybrid PET fusion.,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 51, no. 4, pp. 923–31, Nov. 2001.

[9] V. Grégoire, R. Jeraj, J. A. Lee, and B. O’Sullivan, “Radiotherapy for head and neck tumours in 2012 and beyond: conformal, tailored, and adaptive?,” *Lancet. Oncol.*, vol. 13, no. 7, pp. e292-300, Jul. 2012.

[10] U. Nestle, S. Kremp, A. Schaefer-Schuler, C. Sebastian-Welsch, D. Hellwig, and C. Ru, “Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-small cell lung cancer.,” *J. Nucl. Med.*, vol. 46, no. 8, pp. 1342–1348, 2005.

[11] E. C. Ford, P. E. Kinahan, L. Hanlon, A. Alessio, J. Rajendran, D. L. Schwartz, and M. Phillips, “Tumor delineation using PET in head and neck cancers: Threshold contouring and lesion volumes,” *Med. Phys.*, vol. 33, no. 11, pp. 4280–4288, 2006.

[12] B. Berthon, C. Marshall, A. Edwards, M. Evans, and E. Spezi, “Influence of cold walls on PET image quantification and volume segmentation.,” *Med. Phys.*, vol. 40, no. 8, pp. 1–13, 2013.

[13] B. Berthon, C. Marshall, M. Evans, and E. Spezi, “Evaluation of advanced automatic PET segmentation methods using non-spherical thin-wall inserts,” *Med. Phys.*, vol. 41, no. 2, p. 22502, 2014.

[14] B. Berthon, C. Marshall, M. Evans, and E. Spezi, “ATLAAS: an Automatic decision Tree-based Learning Algorithm for Advanced,” *Phys. Med. Biol.*, no. In press, 2016.

[15] B. Berthon, C. Marshall, R. B. Holmes, and E. Spezi, “A novel phantom technique for evaluating the performance of PET auto-segmentation methods in delineating heterogeneous and irregular lesions,” *Eur. J. Nucl. Med. Mol. Imaging Phys.*, vol. 2, no. 13, 2015.

[16] J. O. Deasy, A. I. Blanco, and V. H. Clark, “CERR: A computational environment for radiotherapy research,” *Med. Phys.*, vol. 30, no. 5, pp. 979–985, 2003.

[17] L. Dice, “Measures of the amount of ecologic association between species.,” *Ecology*, vol. 26, pp. 297–302, 1945.

[18] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. C. Tempany, M. R. Kaus, S. J. Haker, W. M. W. Iii, and F. A. Jolesz, “Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index,” *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, 2004.

[19] X. Geets, M. Tomsej, J. a. Lee, T. Duprez, E. Coche, G. Cosnard, M. Lonneux, and V. Grégoire, “Adaptive biological image-guided IMRT with anatomic and functional imaging in pharyngo-laryngeal tumors: Impact on target volume delineation and dose distribution using helical tomotherapy,” *Radiother. Oncol.*, vol. 85, no. 1, pp. 105–

115, Oct. 2007.

[20] M. Leclerc, E. Lartigau, T. Lacornerie, J. Daisne, A. Kramar, and V. Grégoire, “Primary tumor delineation based on 18FDG PET for locally advanced head and neck cancer treated by chemo-radiotherapy,” *Radiother. Oncol.*, vol. 116, no. 1, pp. 87–93, 2015.

[21] C. Greco, S. A. Nehmeh, H. Schoder, L. Gonen, B. Raphael, H. E. Stambuk, J. L. Humm, S. M. Larson, and N. Y. Lee, “Evaluation of different methods of 18F-FDG-PET target volume delineation in the radiotherapy of head and neck cancer,” *Am. J. Clin. Oncol.*, vol. 31, no. 5, pp. 439–445, 2008.

[22] K. L. Newbold, M. Partridge, G. Cook, B. Sharma, P. Rhys-Evans, K. J. Harrington, and C. M. Nutting, “Evaluation of the role of 18FDG-PET/CT in radiotherapy target definition in patients with head and neck cancer.,” *Acta Oncol.*, vol. 47, no. 7, pp. 1229–36, Jan. 2008.

[23] A. C. Paulino, M. Koshy, R. Howell, D. Schuster, and L. W. Davis, “Comparison of CT- and FDG-PET-defined gross tumor volume in intensity-modulated radiotherapy for head-and-neck cancer,” *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 61, no. 5, pp. 1385–1392, Apr. 2005.

[24] X. Geets, J. a Lee, A. Bol, M. Lonneux, and V. Grégoire, “A gradient-based method for segmenting FDG-PET images: methodology and validation.,” *Eur. J. Nucl. Med. Mol. Imaging*, vol. 34, no. 9, pp. 1427–38, Sep. 2007.

[25] I. Hoogsteen, H. Marres, J. Bussink, A. van der Kogel, and J. Kaanders, “Tumor microenvironment in head and neck squamous cell carcinomas: predictive value and clinical relevance of hypoxic markers. A review.,” *Head Neck*, vol. 29, pp. 591–604, 2007.

[26] C. C. Ling, J. Humm, S. Larson, H. Amols, Z. Fuks, S. Leibel, J. A. Koutcher, J. F. Kayyem, R. M. Kumar, S. E. Fraser, T. J. Meade, A. Moore, J. P. Basilion, E. A.

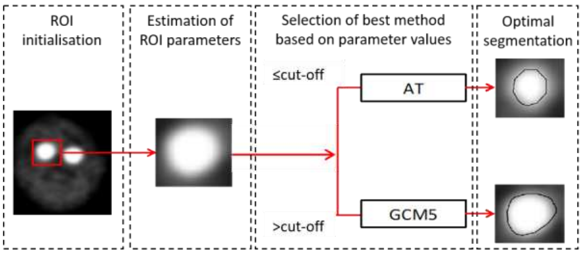
Chiocca, R. Weissleder, R. Weissleder, M. Simonova, A. Bogdanova, S. Bredow, W. S. Enochs, A. Bogdanov, J. L. Urbain, R. J. Hamilton, P. J. Sweeney, C. A. Pelizzari, E. Al., C. C. Ling, C. Burman, C. S. Chui, E. Al., C. C. Ling, Z. Fuks, T. Bortfeld, J. Burkelbach, R. Boesecke, W. Schlegel, R. Mohan, X. Wang, A. Jackson, E. Al., S. V. Spirou, C. S. Chui, C. Burman, C. S. Chui, G. Kutcher, E. Al., H. Bartelink, A. Begg, J. C. Martin, E. Al., R. B. Moon, J. H. Richards, J. A. Koutcher, A. A. Alfieri, M. L. Devitt, E. Al., P. Okunieff, C. S. Walsh, P. Vaupel, E. Al., N. Weidner, J. P. Semple, W. R. Welch, J. Folkman, W. A. Kaiser, E. Zeitler, N. Weidner, J. Folkman, F. Pozza, E. Al., S. Mussurakis, D. L. Buckley, A. Horsman, D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, M. Laval-Jeantet, R. D. Tien, G. J. Felsberg, H. Friedman, M. Brown, J. MacFall, K. Krabbe, P. Gideon, P. Wagn, U. Hansen, C. Thomsen, F. Madsen, T. Ichikawa, H. Haradome, J. Hachiya, T. Nitatori, T. Araki, K. H. Kim, N. R. Relkin, K. M. Lee, J. Hirsch, J. W. Belliveau, D. N. Kennedy, R. C. McKinstry, E. Al., S. G. Kim, J. Ashe, K. Hendrich, E. Al., I. Fried, V. I. Nenov, S. G. Ojemann, R. P. Woods, J. Kurhanewicz, D. B. Vigneron, H. Hricak, P. Narayan, P. Carroll, S. J. Nelson, Y. Kaji, J. Kurhanewicz, H. Hricak, E. Al., J. Kurhanewicz, D. B. Vigneron, H. Hricak, E. Al., F. Parivar, H. Hricak, K. Shinohara, E. Al., L. L. Wald, S. J. Nelson, M. R. Day, E. Al., O. Speck, T. Thiel, J. Hennig, G. Tedeschi, N. Lundbom, R. Raman, E. Al., K. L. Behar, J. A. den Hollander, M. E. Stromski, E. Al., G. Schwickert, S. Walenta, K. Sundfor, E. K. Rofstad, W. Mueller-Klieser, S. Walenta, A. Salameh, H. Lyng, E. Al., Z. M. Bhujwalla, J. D. Glickson, U. Mahmood, A. A. Alfieri, H. Thaler, D. Cowburn, J. A. Koutcher, J. C. Street, U. Mahmood, C. Matei, J. A. Koutcher, J. P. Freyer, P. L. Schor, K. A. Jarrett, M. Neeman, L. O. Sillerud, R. J. Gillies, J. A. Barry, B. D. Ross, M. W. Dewhirst, H. D. Sostman, K. A. Leopold, E. Al., D. M. Prescott, H. C. Charles, H. D. Sostman, E. Al., D. M. Brizel, S. P. Scully, J. M. Harrelson, E. Al., K. Scheidhauer, A. Scharl, U. Pietrzyk, E. Al., P. Rigo, P. Paulus, B. J. Kaschten, E. Al., C. I. Utech, C. S. Young, P. F. Winter, N. Avril, S. Bense, S. I. Ziegler, E. Al., C. S. Brock, S. R. Meikle, P. Price, J. D. Kiffer,

S. U. Berlangieri, A. M. Scott, E. Al., M. T. Munley, L. B. Marks, C. Scarfone, E. Al., U. Haberkorn, L. G. Strauss, A. Dimitrakopoulou, E. Al., H. Minn, M. Lapela, P. J. Klemi, E. Al., A. J. Fischman, A. F. Thornton, M. P. Frosch, B. Swearinger, R. G. Gonzalez, N. M. Alpert, J. G. Tjuvalev, H. A. Macapinlac, F. Daghighian, E. Al., A. F. Shields, D. A. Mankoff, M. M. Graham, E. Al., A. F. Shields, D. A. Mankoff, J. M. Link, E. Al., T. Hara, N. Kosaka, H. Kishi, R. H. Thomlinson, L. H. Gray, J. Overgaard, M. R. Horsman, M. Hockel, K. Schlenger, M. Mitze, U. Schaffer, P. Vaupel, D. M. Brizel, S. P. Scully, J. M. Harrelson, E. Al., M. Hockel, K. Schlenger, B. Aral, M. Mitze, U. Schaffer, P. Vaupel, J. S. Rasey, W. J. Koh, M. L. Evans, E. Al., J. D. Chapman, E. L. Engelhardt, C. C. Stobbe, R. F. Schneider, G. E. Hanks, F. Blankenberg, P. Katsikis, J. Tait, E. Al., P. Vriens, F. Blankenberg, J. Stoot, E. Al., J. G. Tjuvajev, R. Finn, K. Watanabe, E. Al., S. M. Larson, J. Tjuvajev, R. Blasberg, B. E. Rogers, M. E. Rosenfeld, M. B. Khazaeli, E. Al., M. I. Nunez, T. J. McMillan, M. T. Valenzuela, J. M. R. de Almodovar, V. Pedraza, G. C. Li, H. Ouyang, X. Li, E. Al., C. C. Ling, B. Endlich, W. G. McKenna, M. C. Weiss, B. Endlich, E. Al., R. G. Bristow, S. Benchimol, R. P. Hill, J. M. Brown, B. G. Wouters, M. Austin-Seymour, G. T. Chen, J. Rosenman, J. Michalski, K. Lindsley, M. Goitein, J. A. Raleigh, M. W. Dewhirst, and D. E. Thrall, "Towards multidimensional radiotherapy (MD-CRT): biological imaging and biological conformality.," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 47, no. 3, pp. 551–60, 2000.

[27] W. van Elmpt, D. De Ruysscher, A. van der Salm, A. Lakeman, J. van der Stoep, D. Emans, E. Damen, M. Öllers, J.-J. Sonke, and J. Belderbos, "The PET-boost randomised phase II dose-escalation trial in non-small cell lung cancer," *Radiother. Oncol.*, vol. 104, no. 1, pp. 67–71, 2012.

[28] A. J. G. Even, J. van der Stoep, C. M. L. Zegers, B. Reymen, E. G. C. Troost, P. Lambin, and W. van Elmpt, "PET-based dose painting in non-small cell lung cancer: Comparing uniform dose escalation with boosting hypoxic and metabolically active

sub-volumes," *Radiother. Oncol.*, vol. 116, no. 2, 2015.



<u>Figure 1.Example of steps in the decision tree method implemented in the ATLAAS segmentation model.</u>

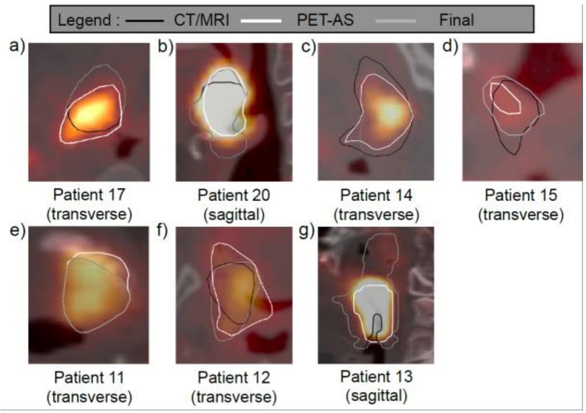


Figure 2. $GTV_{CT/MRI}$, $GTVp_{ATLAAS}$, and $GTVp_{final}$ compared for 7 clinical cases.

Table 1. GTVp volumes and DSC index for manual and ATLAAS contours.

	Patient No										
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**Mean**
Final volume (mL)	33.1	45.9	21.5	38.2	27.5	54.7	33.1	19.0	33.8	17.4	-
CT/MRI volume (mL)	27.1	40.6	19.8	32.8	26.9	60.8	28.5	16.6	31.3	15.6	-
ATLAAS volume (mL)	27.3	41.9	7.8	26.2	15.6	52.5	29.0	16.1	23.7	8.6	-
DSC(ATLAAS vs CT/MRI)	0.77	0.76	0.44	0.75	0.67	0.82	0.74	0.74	0.76	0.51	0.70
DSC(ATLAAS vs final)	0.90	0.77	0.53	0.81	0.68	0.97	0.84	0.92	0.83	0.58	0.77

Table 2. Quantification of the changes to the final volume (growth and shrinkage) based on the ATLAAS outlines, and differences between ATLAAS and CT/MRI not taken into account in the final GTV. Calculations corresponding to the different rows are schematically described under the table.

		Patient No									
		11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	**% ATLAAS included in $GTVp_{final}$**	99.6	83.2	100	100	94.0	99.1	91.8	100	100	88.0
2	**Modifications to $GTVp_{CT/MRI}$ (% $GTVp_{CT/MRI}$)**	25.1	32.0	7.9	17.0	6.5	33.1	16.3	28.4	7.7	9.8
3	**$GTVp_{CT/MRI}$ grown based on ATLAAS (mL)**	8.3	10.6	1.7	5.4	1.7	5.8	4.9	3.9	2.6	1.5
3a	**of which superior-inferior extent**	0.9	-	-	0.7	0.3	-	-	0.6	-	1.1
4	**$GTVp_{CT/MRI}$ shrunk based on ATLAAS (mL)**	-	4.1	-	1.1	0.1	12.3	0.5	1.5	-	0.2
4a	**of which superior-inferior extent**	-	0.9	-	-	-	1.5	-	-	-	-
5	**Difference between $GTVp_{final}$ and ATLAAS (mL), of which:**	6.9	20.0	12.7	12.0	45.1	0.6	11.5	3.4	10.0	11.1
5a	**Bone regions (%)**	2	3	0	0	0	0	0	0	0	0
5b	**Air cavities or vicinity (%)**	0	37	6	0	9	0	10	12	0	7
5c	**Superior-inferior extent (%)**	14	0	30	0	13	0	19	1	4	30
5d	**Transverse soft tissue extent (%)**	84	60	64	100	78	100	71	87	96	63

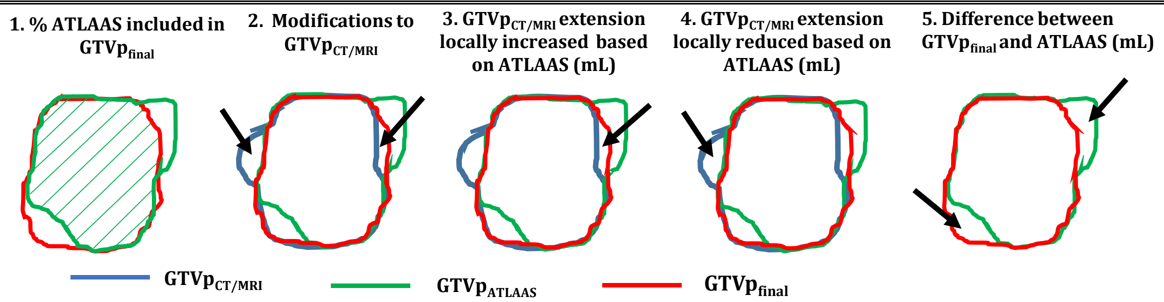


Figure legends:

Figure 1. Example of steps in the decision tree method implemented in the ATLAAS segmentation model.

Figure 2. $GTV_{CT/MRI}$, $GTVp_{ATLAAS}$, and $GTVp_{final}$ compared for 7 clinical cases.