Title

Validation of MRI-based auto-contouring software tool for gross tumour delineation in head and neck cancer radiotherapy planning

Authors

Trushali Doshi, PhD Student, Department of Electronic & Electrical Engineering, University of Strathclyde, Glasgow

Corresponding author

Contact Information:

Address: Royal College Building, University of Strathclyde,

204 George Street,

Glasgow, G1 1XW

Scotland, United Kingdom (UK)

Email: trushali.doshi@strath.ac.uk; trushalid@gmail.com

Tel: (+44) 0755 700 2284; (+44) 0141 548 2205

Dr Christina Wilson, Speciality Registrar, Beatson West of Scotland Cancer Centre, Glasgow, Scotland, UK, G12 0YN

Email: Christina.Wilson@ggc.scot.nhs.uk

Dr Claire Paterson, Consultant Clinical Oncologist, Beatson West of Scotland Cancer Centre, Glasgow, Scotland, UK, G12 0YN

Email: Claire.Paterson2@ggc.scot.nhs.uk

Dr Carolynn Lamb, Consultant Clinical Oncologist, Beatson West of Scotland Cancer Centre, Glasgow, Scotland, UK, G12 0YN

Email: Carolynn.Lamb@ggc.scot.nhs.uk

Dr Allan James, Consultant Clinical Oncologist, Beatson West of Scotland Cancer Centre, Glasgow, Scotland, UK, G12 0YN

Email: Allan.James@ggc.scot.nhs.uk

Mr Kenneth MacKenzie, Consultant ENT Surgeon, Glasgow Royal Infirmary, Glasgow, Scotland, UK, G4 0ET

Email: Kenneth.MacKenzie@ggc.scot.nhs.uk

Prof John Soraghan, Professor, Department of Electronic & Electrical Engineering, University of Strathclyde, Glasgow, Scotland, UK

Email: j.soraghan@strath.ac.uk, G1 1XW

Dr Lykourgos Petropoulakis, Senior Knowledge Exchange Fellow, Department of Electronic & Electrical Engineering, University of Strathclyde, Glasgow, Scotland, UK, G1 1XW

Email: l.petropoulakis@strath.ac.uk

Dr Gaetano Di Caterina, Research Associate, Department of Electronic & Electrical Engineering, University of Strathclyde, Glasgow, Scotland, UK, G1 1XW

Email: gaetano.di-caterina@strath.ac.uk

Dr Derek Grose, Consultant Clinical Oncologist, Beatson West of Scotland Cancer Centre, Glasgow, Scotland, UK, G12 0YN

Email: Derek.Grose@ggc.scot.nhs.uk

Number of Pages: **32**

Number of Figures: **4**

Number of Tables: **2**

Abstract

Purpose

To perform statistical validation of a newly developed magnetic resonance imaging (MRI) auto-contouring software tool for gross tumour volume (GTV) delineation in head and neck tumours to assist in radiotherapy planning.

Material and Methods

Axial MRI baseline scans were obtained for 10 oropharyngeal and laryngeal cancer patients. GTV was present on 102 axial slices and auto-contoured using the modified fuzzy c-means clustering integrated with level set method (FCLSM).

Peer reviewed (C-gold) manual contours were used as the reference standard to validate auto-contoured GTVs (C-auto) and mean manual contours (C-manual) from 2 expert clinicians (C1 and C2). Multiple geometrical metrics, including Dice Similarity Coefficient (DSC) were used for quantitative validation. A DSC $\geq$0.7 was deemed acceptable. Inter- and intra- variabilities amongst the manual contours were also validated.

The 2-dimension (2D) contours were then reconstructed in 3D for GTV volume calculation, comparison and 3D visualisation.

Results

The mean DSC between C-gold and C-auto was 0.79. The mean DSC between C-gold and C-manual was 0.79 and that between C1 and C2 was 0.80.

The average time for GTV auto-contouring per patient was 8 minutes (range 6-13mins; mean 45seconds per axial slice) compared to 15 minutes (range 6-23mins; mean 88 seconds per axial slice) for C1.

The average volume concordance between C-gold and C-auto volumes was 86.51% compared to 74.16% between C-gold and C-manual. The average volume concordance between C1 and C2 volumes was 86.82%.

Conclusions

This newly-designed MRI-based auto-contouring software tool shows initial acceptable results in GTV delineation of oropharyngeal and laryngeal tumours using FCLSM.

This auto-contouring software tool may help reduce inter- and intra- variability and can assist clinical oncologists with time-consuming, complex radiotherapy planning.

Keywords

Automatic delineation

Magnetic Resonance Imaging

Head and Neck GTV

Pharyngeal and Laryngeal tumours

Radiotherapy treatment planning

Volume calculation and 3D visualisation

Introduction

The widespread implementation of intensity modulated radiotherapy (IMRT) has allowed improved conformity of high dose radiation to the gross tumour (GTV) and planning target volume. By minimising the dose to organs at risk, this aims to improve late toxicity [1-2] and has generated renewed interest in dose-escalation strategies. However, one concern with modern highly-conformal techniques is the potential for geographical misses and the accurate definition of target volumes is critical. Clinician voluming time and the complexity of this process have also increased significantly as a result [3].

In recent years, MRI linear accelerators [4-5] have been developed to integrate MRI in radiotherapy planning and allow more accurate identification of tumour and normal structures. Delineation of the target volumes, especially tumour delineation, is still done manually by clinicians and is time consuming and open to inter- and intra-observer variations [6]. The oropharynx and larynx regions, due to their geometric variability and non-convex form, are examples of areas at risk of this [7]. When treating with radical intent using highly-conformal modern radiotherapy techniques this poses challenges and scope for potential errors. Reproducible automatic contouring programmes have the potential to reduce the risk of inter- and intra-observer variability [8].

Administration of gadolinium contrast before obtaining a T1-weighted MRI scan enhances the vasculature and gross tumour, thus significantly improving soft tissue contrast, margin definition and therefore accuracy. Despite this, issues such as tumour necrosis, anatomical and geometric variability, diffused boundaries and scan artefact can make GTV contouring, manual or computer-automated, a challenging task.

In recent years, automatic delineation tools [9-11] have been developed and validated using atlas-based techniques for the segmentation of normal anatomical structures from the head and neck region. This aims to improve time, efficiency and reduce variations when compared to manual delineation. It is, however, not feasible to develop an atlas-based approach for auto-contouring of pathological disease [12]. Different non-atlas based techniques have been developed for semi-automatic or automatic delineation of head and neck GTV using MRI [13-17]. Publications relevant to oropharyngeal and laryngeal tumours are summarised below.

Two semi-automated tumour volume measurement methods [16] based on seed growing and region deformation are validated on 16 patients with tongue tumours from MRI. These techniques showed satisfactory contouring results but required manual interaction to place seed points. Auto-contouring technique for extraction of tongue tumours with the help of T1-weighted and T2-weighted MRI is proposed in [17]. In this technique, contouring is applied from coarse to fine level with selection

of five features of tumour regions. This technique was tested on only 16 axial MRI slices as compared to 102 slices in this work.

The intent of this work was to develop a novel approach for auto-contouring of oropharyngeal and laryngeal GTV from gadolinium-enhanced T1-weighted (T1+Gd) MRI axial slices with 3D reconstruction and volume calculation followed by validation against the current standard manual approach. To the best of our knowledge this is the first automatic tool developed and validated against manual approach for the delineation of oropharyngeal and laryngeal GTV using T1+Gd MRI. This tool may help reduce inter- and intra- variability and can assist clinical oncologists with time-consuming, complex radiotherapy planning.

Methods and Materials

Data acquisition

This was a cross-sectional study with retrospective imaging data. Ten patients previously treated with radical chemo-radiotherapy for Stage II/III head-and-neck squamous cell carcinoma (6 oropharynx and 4 larynx) were randomly selected and included. MRI scans for each patient had been carried out during formal staging work-up pre-treatment. These MRI scans were acquired and processed retrospectively using the automatic contouring tool.

All MRI DICOM files were processed. Scans were obtained from 3 different 1.5T MRI scanners, namely Magnetom Avanto from Siemens, Intera Neuro coils from Philips Medical Systems, and Signa HDxt from GE Medical Systems. The MRI scans were undertaken pre- and post-intravenous gadolinium (Gd) contrast and were acquired after 15-20 minutes of intravenous injection of 0.1ml/kg. The range of other imaging parameters were 3-5mm slice thickness, 3.3-6 mm spacing in between slices, 9.06-20ms echo time, 542-1066ms repetition time, 90^{o}-150^{o} flip angle, 0.43x0.43-0.94x0.94 mm in-plane resolution (pixel spacing), 256x256-512x512 acquisition matrix and 97.65-221 Hz/pixel bandwidth.

Of the 10 MRIs assessed, there was a total of 102 axial slices containing gross tumour.

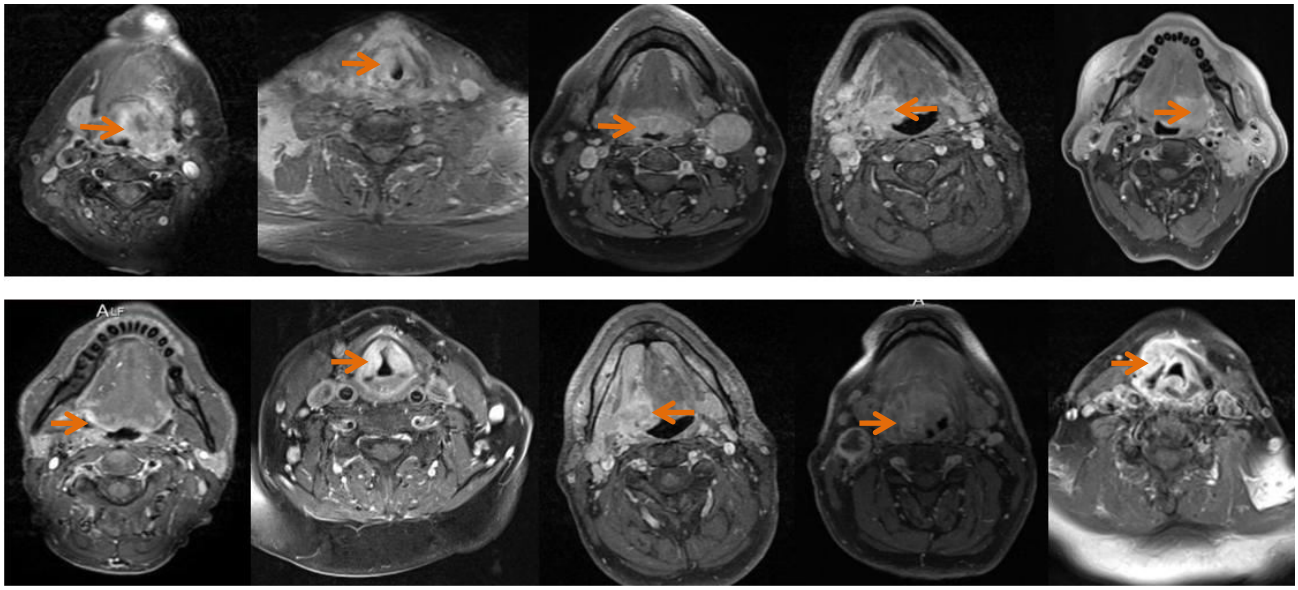


Fig. 1: Typical MRI dataset with GTV (shown with orange arrow) used in this work to validate auto-contouring tool (1 T1+Gd MRI axial slice from each Patient)

The MRI scans not only demonstrated variations in the size and shape of the tumour regions, but also variability in contrast uptake, ill-defined margins and artefact presence. Typical axial slices used to test the auto-contouring tool are shown in Fig. 1.

Auto-contouring software tool

The segmentation framework using this novel and fully-automatic contrast enhanced T1-weighted (T1+Gd) MRI auto-contouring tool for oropharynx GTV in one patient was presented in [18-19]. T1+Gd MRI is preferred imaging modality to define tumour spread for oropharyngeal and laryngeal tumours [20], [21] compared to unenhanced (normal) T1, proton-density and T2 MRI, due to its significantly higher contrast-to-noise ratios for the primary tumour and lymph nodes and it significantly improves soft tissue contrast and tumour margin definition.

This paper improves the segmentation framework presented in [18-19], in terms of pre-processing, post-processing techniques and new algorithm for the detection of pharynx region, to assess the tool further in oropharynx and larynx GTV, including 3D reconstruction and volume calculation.

Compared to the initial framework [18-19] additional pre-processing techniques such as histogram equalisation and log-exponential transform are applied to the MRI slice to increase the contrast and reduce noise. The intensity inhomogeneity present in MRI slice is reduced using local entropy minimisation technique with modifications

for adaptive knot spacing. Further, the pharynx region is detected using fuzzy rule-based system. Two fuzzy rules based on intensity and location of pharynx region are created for pharynx region detection. Further, as in [19], the information of detected pharynx region is utilised in fuzzy c-means (FCM) clustering for initial segmentation of the pre-processed MRI slice into five clusters. The automatically selected cluster containing suspected tumour undergoes new automatic post-processing including morphological and non-linear filtering to remove aberrant regions. Finally, the localised region-based level set method (LSM) [18] obtains a smooth 2D GTV contour for these regions.

This auto-contouring tool processes each axial slice individually (2D segmentation) to avoid inter- and intra- patient intensity and spatial normalisation. With regards to initial 3D segmentation, this can be a challenging problem in the presence of pathology [22]. Furthermore, MRI data in this study is highly anisotropic with large slice spacing which may provide inaccurate results if using initial 3D segmentation.

The GTV from the contiguous 2D contours were then reconstructed in 3D using interpolation and rectangular mesh generation technique. The irregularities in 2D contours are handled by automatic smoothing of the 3D reconstructed volume. The volume of the contoured region is calculated as contour area multiplied by slice thickness and spacing in between slices.

Endpoints and analysis

Two consultant clinical oncologists (C1 & C2) subspecialising in head-and-neck malignancy manually outlined the GTV according to the published EORTC/GORTEC/RTOG endorsed consensus guidelines and PARSPORT/ ARTDECO protocols [23-25]. The C1 and C2 repeated the process at one month interval in order to allow inter- and intra-variability assessment.

To obtain a gold standard manual contour for comparison, the MRIs were peer reviewed in a tertiary oncology centre at the head-and-neck radiotherapy weekly team meeting for expert general consensus of the GTV for all 10 patients.

The peer reviewed (gold standard) manual dataset was referred as C-gold.

The first manual dataset contoured by C1 and C2 was referred as C1-A and C2-A respectively. The second dataset was referred as C1-B and C2-B respectively. The mean (C-manual) was calculated from these data-sets by employing STAPLE algorithm [26] and was used for comparison against C-gold.

Inter-variability was assessed between C1 and C2. Intra-variability was assessed between C1-A and C1-B and between C2-A and C2-B.

All GTV contours were reconstructed into 3D. Volume concordance was calculated between C-gold and the automatic (C-auto) and C-manual contoured volumes.

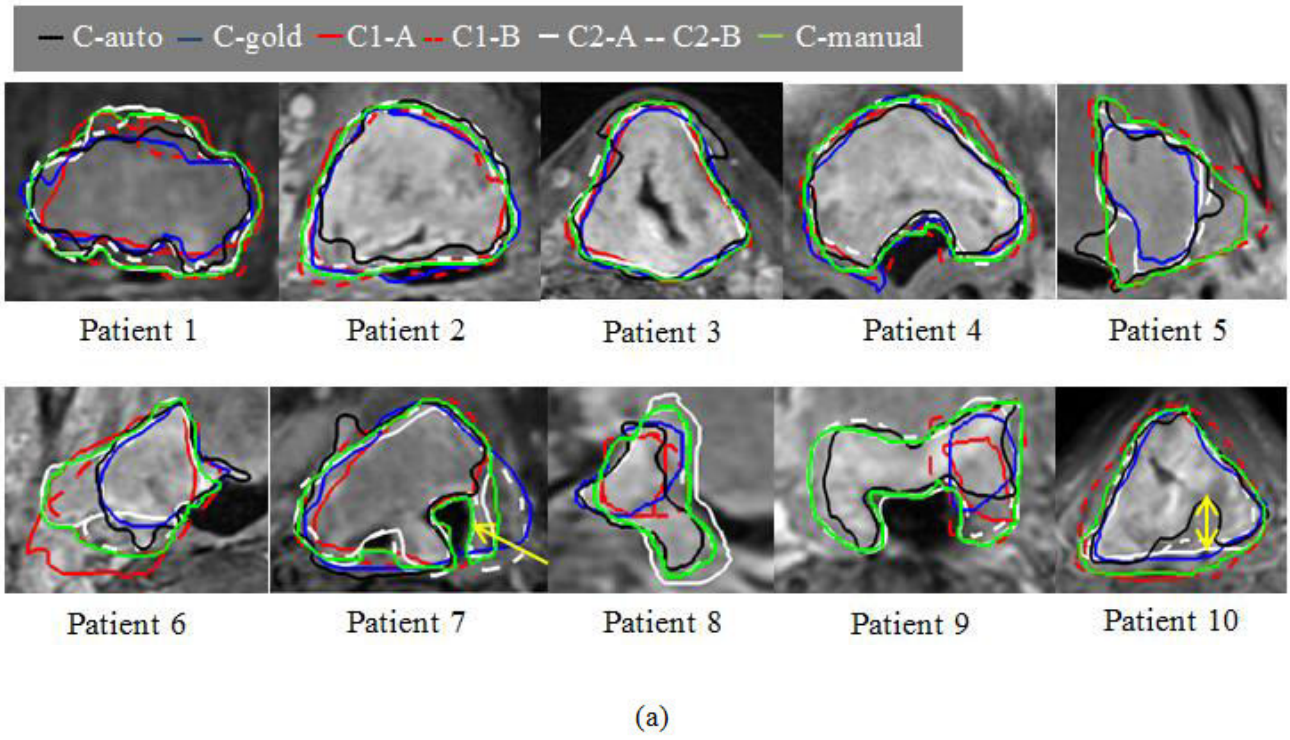
Variations were quantified using Dice Similarity Coefficient (DSC) [27] and Hausdorff Distance (HD) [28]. Based on other relevant studies [12, 29-30], a value of DSC $\geq$0.7 was utilised as an acceptable threshold.

Statistical significance between the C-gold compared to the C-auto and C-manual were assessed using Pearson's Correlation Coefficient (CC). A value of >0.6 indicates a strong positive correlation and >0.8 very strong [29-30].

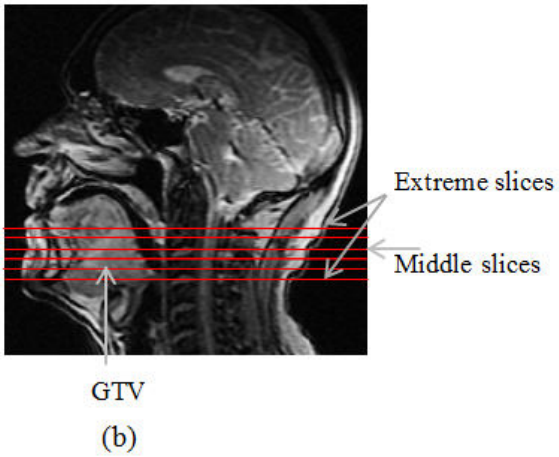
Results

Evaluation between C-gold, C- auto and C-manual contours

The auto-contouring tool was implemented in Matlab 2014a (Mathworks, Natick, MA), on Windows 7 system. The Matlab software was executed on a Dell U2412M machine with 8GB RAM. The average GTV auto-contouring time per patient was 8 minutes (range 6-13 minutes, mean 45 seconds per slice) compared to 15 minutes (range 6-23 minutes; mean 88 seconds per axial slice) for C1 contours. It was observed that it takes less time to obtain C1 outline for larynx GTV (mean 80 seconds per slice) than for the oropharyngeal GTV (mean 96 seconds per slice).



(a)



(b)

Fig. 2: (a) Examples of oropharynx and larynx GTV outline in single axial MRI slice for each of the 10 patients (b) Sagittal slice showing position of axial slices.

- **Top row of Fig. 2(a): examples of very strong correlation (>0.8)**
- **Bottom row Fig. 2(a): examples of greater variability between contours (correlation 0.6-0.8).**

Ten examples of GTV outlines from single axial MRI slices are shown in Fig. 2(a). All five slices of top row (Patient 1 – Patient 5) of Fig. 2(a) demonstrate examples of

slices with excellent correlation (>0.8) between automatic and manual contours. These tumour regions were more likely to exhibit distinct boundaries and homogeneity. These types of features were generally observed in middle slices (Fig. 2(b)).

All five slices of bottom row (Patient 6 – Patient 10) of Fig. 2(a) demonstrate lower correlation (0.6-0.8) between automatic and manual contours and exhibits greater inter- and intra-variability. These variabilities are mostly observed in extreme slices (Fig. 2(b)) where the tumour is smaller; less defined (Fig. 2(a) Patient 8-9) and often fused into nearby normal structures (see Fig. 2(a) Patient 6). One other factor contributing to these variabilities includes inclusion/exclusion of the whole pharynx region in the GTV outline (see Fig. 2(a) Patient 7, yellow arrow).

The quantitative measures between C-gold and C-auto and between C-gold and C-manual contours are presented in Table 1. This demonstrates a DSC ≥0.75 for all patients except for Patient 1 and 4. The mean DSC between C-gold and C-auto contours is 0.79 (>0.85 for 49 slices out of 102), indicating satisfactory spatial agreement between automatic and standard manual contours. It is also comparable with the DSC of C-gold versus C-manual contours with value of 0.79. It was observed that the auto-contouring tool performed better in oropharyngeal cases (mean DSC: 0.79) compared to manual contouring (mean DSC: 0.77).

Table1: Quantitative evaluation between gold standard, automatic and manual contours (Oro: Oroharyngeal cases). Underline values in Table 1 illustrate a positive concordance, i.e. greater C-auto volume.

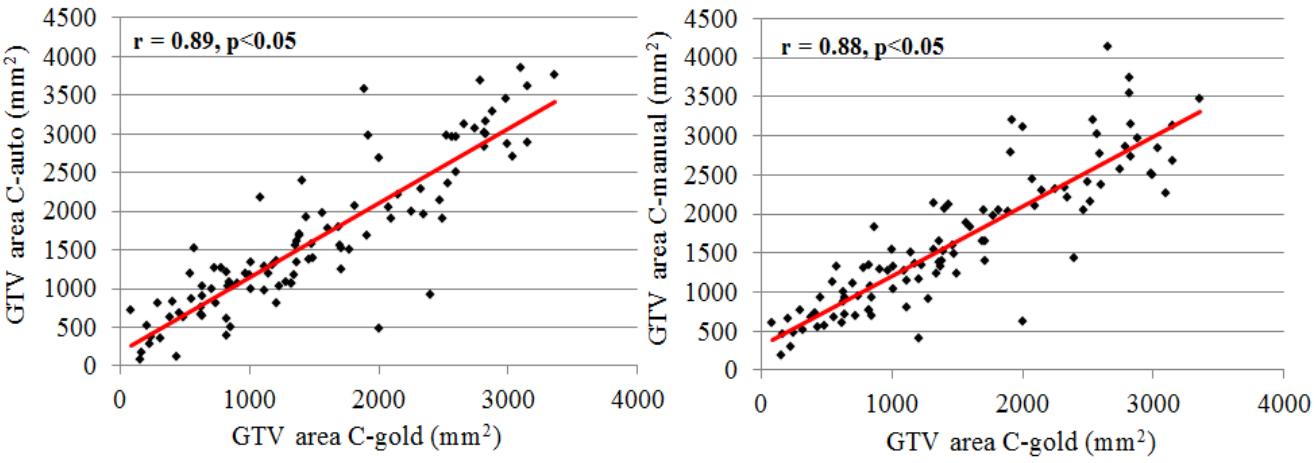
Patient No.	No. MRI axial slices inclusive of GTV	DSC (C-gold vs. C-auto)	DSC (C-gold vs. C-manual)	HD (C-gold vs. C-auto) (mm)	HD (C-gold vs. C-manual) (mm)	Volume Concordance (%) (C-gold vs. C-auto)	Volume Concordance (%) (C-gold vs. C-manual)
1	08 (Oro)	0.68	0.66	8.05	6.13	55.85	24.51
2	12 (Oro)	0.83	0.81	6.77	7.50	87.01	90.5
3	07 (Oro)	0.84	0.82	5.18	5.72	96.63	81.71
4	12 (Oro)	0.74	0.69	6.83	7.45	74.58	50.39
5	07 (Oro)	0.79	0.81	3.42	3.56	79.67	72.15
6	09 (Oro)	0.83	0.81	6.95	7.91	95.38	63.64
7	12 (Larynx)	0.75	0.8	9.56	7.91	90.65	80.31
8	07 (Larynx)	0.86	0.88	5.02	3.80	95.25	93.6
9	17 (Larynx)	0.77	0.82	6.78	6.03	94.49	95.53
10	11 (Larynx)	0.77	0.76	8.30	7.56	95.63	89.25
Absolute Mean:		**0.79**	**0.79**	**6.69**	**6.36**	**86.51**	**74.16**
CoV (Coefficient of Variation) (%)		**6.66**	**8.06**	**25.27**	**24.20**	**14.39**	**28.82**

The mean HD between C-gold and C-auto is 6.69 mm (range 3.42mm to 9.56mm) and mean HD between C-gold and C-manual is 6.36mm (range 3.56mm to 7.91mm). The auto-contouring tool follows all the details in the concavity regions, unlike humans who usually produce a more general outline, which can result in HD variation between C-gold and C-auto (Coefficient of Variation: 25.27 %). For instance in Fig.

2(a) Patient 10, the DSC value is 0.87. The HD (maximum distance between two point sets) however is 7.98mm (yellow arrow).

The DSC and HD values show that the auto-contouring tool produced comparable results to those seen in expert-led manual contouring. The quantitative comparison between C1 and C-auto and between C2 and C-auto is given in supplementary data (Table S1) which demonstrate mean DSC of 0.77 and 0.81 respectively and mean HD of 7.29mm and 6.52mm.

The linear relationship between C-gold and C-auto and between C-gold and C-manual GTV contours is highlighted in Fig. 3. The CC between C-gold Vs. C-auto and C-gold Vs. C-manual was 0.89 and 0.88 respectively with $p<0.05$ (two-sided). The CC between C1 and C-auto is 0.81 and between C2 and C-auto is 0.90 (supplementary data Fig. S1).



(a) (b)

Fig. 3: Correlation analysisfor 102 T1+Gd MRI axial slices

(a) between C-gold Vs. C-auto (b) between C-gold Vs. C-manual

Table 2: Mean quantitative values for GTV contours from 102 T1+Gd MRI axial slices between C1 and C2 (inter-variability) and for C1-A Vs. C1-B and C2-A Vs. C2-B (intra-variability)

Inter- and Intra- variability calculations	DSC	HD (mm)	Volume Concordance (%)
C1 vs. C2	0.80	6.53	86.82
C1-A vs. C1-B	0.85	4.75	90.45
C2-A vs. C2-B	0.86	4.68	89.32

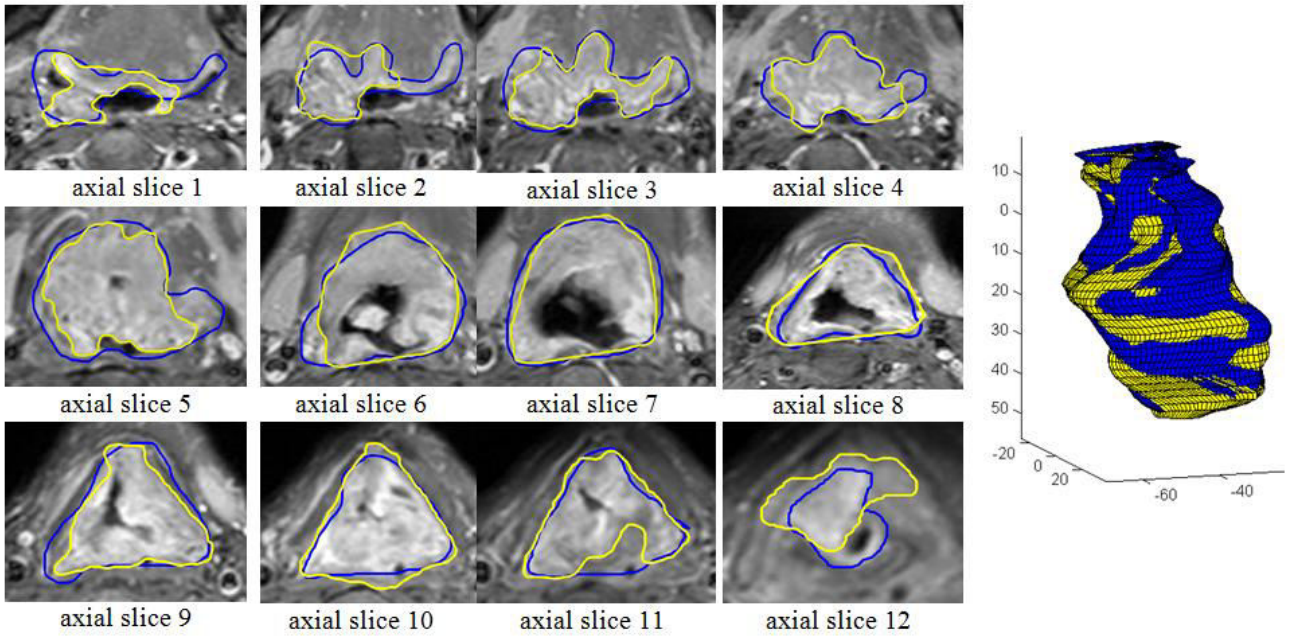
Inter- and Intra-variability calculations

Summarized results for inter- and intra-variability calculations are shown in Table 2. From qualitative observations (Fig. 2(a)) and from Table 2 (for intra-variability) there appears to be no significant difference in intra-varability calculations for this set of examples. For inter-variablity the main variation is observed in oropharyngeal cases.

Volume Concordance evaluation

The volume for each GTV is demonstrated in Supplementary data Fig. S2. Volumes ranged from 19.74 cm^3 to 136.93 cm^3.

The axial slices for one patient comparing C-gold and C-auto GTV outline is shown in Fig 4(a). The 3D reconstructed volume of GTV from axial slices of Fig. 4(a) is shown in Fig. 4(b). A yellow colour represents a volume greater in C-auto than C-gold. Another example of 3D reconstructed volume of GTV from axial slices is shown in supplementary data Fig. S3.



(a) (b)

Fig. 4:

(a) Sequential MRI axial slices with automatic (C-auto) (yellow) and gold standard (C-gold) (blue) GTV outline for one patient

(b) 3D reconstructed volume from GTV outlines shown in Fig. 4(a). (automatic (C-auto) - yellow; gold standard (C-gold) - blue)

Volume concordance between C-gold and C-auto is demonstrated in Table 1. Underlined figures represent a volume where C-auto is greater than C-gold. The automatic tool overestimated laryngeal GTV in Patients 7-10. The mean volume concordance between C-gold and C-auto is 86.51% compared to 74.16% volume concordance between C-gold and C-manual. The correlation coefficient between C-gold and C-auto is 0.98.

This shows that the automatic volumes are comparable to the peer-reviewed gold standard volumes.

Discussion

This study has shown that this novel, fully-automated software tool can provide acceptable results in MRI-based auto-contouring of primary oropharyngeal and laryngeal tumours.

This tool is able to delineate GTV without any manual interaction even in the presence of ill-defined margins, geometric variability and MRI artefacts. It is able to successfully avoid uninvolved regions even when geometrically close and of similar intensity to the GTV. It can also perform robustly against variations caused by different MRI protocols with different manufacturer and scanner models.

To our knowledge, this is the first study that validates the use of an MRI-based auto-contouring tool for GTV delineation and 3D visualisation in oropharyngeal and laryngeal tumours.

One of the main advantages with this tool is the reproducibility of the contouring results, thus, potentially reducing the risk of inter- and intra-operator variability, unlike manual contouring. For this auto-contouring tool, if the parameters values are unchanged, the tool obtains similar results for repeated number of times, indicating the reproducibility of the system as no manual interaction is involved.

The proposed tool does not require any other imaging modality other than a single (T1+Gd contrast) MRI for contouring.

The proposed approach demonstrates high levels of correlation for oropharyngeal and laryngeal tumours when compared to an expert peer-reviewed standard reference with mean DSC of 0.79, correlation coefficient of 0.89, and volume concordance of 86.51% and volume correlation coefficient of 0.98.

There is some discrepancy between the values of DSC and HD when comparing automatic (C-auto) and gold standard (C-gold) outlines. This discrepancy can be attributed to the manual bias of including areas of high risk in the GTV outline (Fig 4(a) axial slice 1-2) or processing each axial slice in a 2D approach by the auto-contouring tool (Fig 4(a) axial slice 6). Other factor contributing to this discrepancy is that, the auto-contouring tool follows a contrast edge more closely than a human observer (Fig. 2(a) Patient 10) thus, increasing variation. In future this discrepancy can be resolved by either making the segmentation protocol for the human observers more specific for the purpose of this analysis, or some surface smoothing could be applied to the auto-contouring results. Either way, a like-for-like comparison would likely improve conformity measurements. However, despite differences in the axial slices between the automatic and gold standard results, there was a very strong agreement in volume comparison with volume concordance of 86.51%.

Another advantage to the software is that, no manual interaction is required for gross tumour contour construction and thus, it saves manual labour. The results obtained from this fully-automatic tool are comparable to the results demonstrated

from the existing semi-automatic systems. In our study a mean DSC of 0.79, HD of 6.69mm and volume concordance of 86.51% were obtained.

A semi-automatic approach in [13] for nasopharyngeal lesion segmentation from 7 patients demonstrated percentage match of 78.5% and correspondence ratio of 66.5%. This algorithm [13] is based on information obtained from both T1 and T2-weighted images as opposed to contrast-enhanced T1 images used in this study.

Another semi-automatic approach [16] validated on 16 patients with tongue tumours demonstrated mean percentage match of 87.07% and mean volume concordance of 92.74%. This technique required manual-placing of seed points in the tumour region or drawing of close loop outside the tumour from expert and have no results to prove any validation on laryngeal tumours.

This tool enables 3D reconstruction of GTV which can provide clinicians with additional information not easily available from 2D MRI slices. Automatic tumour volume measurement could contribute to further GTV staging and prognostication. It has been shown in previous studies [16, 31] that GTV volume is a significant factor in determining the outcome following primary radiotherapy in head-and-neck cancer.

Modern conformal radiotherapy planning is time-consuming and complex. This tool may reduce volume delineation time. It delineated GTV in an average of 8 minutes/patient (range 6-13 minutes) compared to 15 minutes (range 6-23 minutes) for manual contouring from expert. Further, this auto-contouring tool is an unsupervised

segmentation, and could therefore be scheduled into the clinical workflow before a human user (observer) sits to contour a case. Also, the auto-contouring tool is highly parallelisable as each 2D contour is processed individually which may further reduce volume delineation time.

A possible limitation of the current tool may be the evaluation of GTV using MRI information alone as GTV is normally defined by information from clinical examination (including endoscopic findings), EUA (examination under anaesthesia) as well as imaging. However, the aim of this auto-contouring tool is to aid clinical oncologists in accurate GTV delineation for oropharyngeal and laryngeal GTV for radiotherapy treatment planning. Thus, this tool can be used in conjunction with other diagnostic imaging information to help improve the radiotherapy planning process.

Another limitation of the current tool may be increased variability when contouring tumour at the extreme slices or small tumour geographically close and of similar intensity to healthy tissue. This is also the same area where there is greatest clinical and radiological-observer variability.

This tool has shown acceptable and promising results, with potential for further modifications. Currently, the tool processes each axial slice separately and reconstructs in 3D. Future work will focus on ensuring the robustness of the tool and integrating before and after slice information to enable delineating in 3D. Future work

will continue focusing on contouring more challenging cases, such as multiple primary tumour regions and diffusely infiltrating GTVs.

Although our primary results are encouraging and satisfactory, before this auto-contouring tool can be established in clinical setting, further validation with more subjects and rigorous testing would be required. Clinical judgement, using all clinical and diagnostic information, will however always remain vital.

In conclusion, this newly-designed MRI-based auto-contouring software tool shows initial acceptable statistically-validated results in GTV delineation of oropharyngeal and laryngeal tumours using FCLSM. This tool may help reduce inter- and intra- variability and can assist clinical oncologists with time-consuming, complex radiotherapy planning.

Acknowledgements

The authors would like to acknowledge the Beatson Cancer Charity and University of Strathclyde for their financial support with this study.

Conflict of Interest Statement

None

References

[1] PowEH, Kwong DL, McMillan AS et al. Xerostomia and quality of life after intensity-modulated radiotherapy vs. conventional radiotherapy for early-stage nasopharyngeal carcinoma: Initial report on a randomized controlled clinical trial, Int J Radiat Oncol Biol Phys 2006; 66(4):981–991

[2] Nutting CM, Morden JP, Harrington KJ et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial, Lancet Oncol 2011; 12(2): 127–136.

[3] Watkinson J and Gilbert R. Stell and Maran's Textbook of Head and Neck Surgery and Oncology, Fifth Edition, London, UK: Hodder Arnold, 2012

[4] Paulson ES, Erickson B, Schultz C, Allen Li X. Comprehensive MRI simulation methodology using a dedicated MRI scanner in radiation oncology for external beam radiation treatment planning. Med Phys. 2015 Jan; 42(1):28-39

[5] Lagendijk JJ, Raaymakers BW, Raaijmakers AJ et al. MRI/linac integration. Radiother Oncol. 2008 Jan; 86(1):25-29.

[6] Hamilton CS and Ebert MA. Volumetric Uncertainty in Radiotherapy. Clinical Oncology, 2005;17(6):456–464

[7] Anderson CM, Sun W, Buatti JM, et al. Interobserver and intermodality variability in GTV delineation on simulation CT, FDG-PET, and MR Images of Head and Neck Cancer. Jacobs J Radiat Oncol. 2014 Sep; 1(1):006.

[8] Valentini V, Boldrini L, Damiani A, Muren LP, Recommendations on how to establish evidence from auto-segmentation software in radiotherapy. Radiother Oncol. 2014 Sep; 112(3):317-320

[9] Yang X, Wu N, Cheng G, et al. Automated Segmentation of the Parotid Gland Based on Atlas Registration and Machine Learning: A Longitudinal MRI Study in Head-and-Neck Radiation Therapy. Int J Radiat Oncol Biol Phys 2014 Dec; 90 (5):1225–1233

[10] Cheng G, Yang X, Wu N, et al., Multi-atlas-based Segmentation of the Parotid Glands of MR Images in Patients Following Head-and-neck Cancer Radiotherapy. Proc SPIE Int Soc Opt Eng. 2013 Feb; 8670.

[11] Bacher MG, PekarVand KausMR. Model-Based Segmentation of Anatomical Structures in MR Images of the Head and Neck Area. Bildverarbeitungfür die Medizin 2005: 113-117

[12] Liu X, Niethammer M, Kwitt R, McCormick M, Aylward S. Low-rank to the rescue - atlas-based analyses in the presence of pathologies. Med Image Comput Comput Assist Interv. 2014; 17(Pt 3):97-104.

[13] Lee FK, Yeung DK, King AD, Leung SF, Ahuja A. Segmentation of nasopharyngeal carcinoma (NPC) lesions in MR images. Int J Radiat Oncol Biol Phys. 2005 Feb; 61(2):608-620.

[14] Huang W, Chan KL, Gao Y, Zhou J, Chong V. Semi-supervised nasopharyngeal carcinoma lesion extraction from magnetic resonance images using online spectral clustering with a learned metric. Med Image Comput Comput Assist Interv. 2008; 11:51-58.

[15] Lodder WL, Gilhuijs KG, Lange CA, Pameijer FA, Balm AJ, van den Brekel MW. Semi-automated primary tumor volume measurements by dynamic contrast-enhanced MRI in patients with head and neck cancer. Head Neck. 2013 Apr; 35(4):521-526

[16] Vincent F, Chong H, Zhou J, et al. Tongue carcinoma: tumor volume measurement, Int J Radiat Oncol Biol Phys 2004; 59 (1): 59–66

[17] Zhou J, Krishnan S, Chong V, et al., Extraction of Tongue Carcinoma Using Genetic Algorithm-induced Fuzzy Clustering and Artificial Neural Network from MR Images, International Conference of the IEEE EMBS 2004;1790 – 1793

[18] Doshi T, Soraghan J, Petropoulakis L et al., Semi-Automatic Segmentation of Tongue Tumors from Magnetic Resonance Imaging, In Proc. of 20th IEEE IWSSIP conference 2013;143-146.

[19] Doshi T, Soraghan J, Grose D, et al., Modified fuzzy c-means clustering for automatic tongue base tumour extraction from MRI data, In Proc. of European Signal Processing Conference (EUSIPCO) 2014; 2460 – 2464.

[20] Park J, Jung S., Joo Y, Jung C, Cho K, and Kim M, Diagnostic accuracy of magnetic resonance imaging (MRI) in the assessment of tumor invasion depth in oral/oropharyngeal cancer, Oral Oncol. 2011; 47(5):381–386.

[21] Sakai F, Sone S, Kiyono K, et al., MR evaluation of laryngohypopharyngeal cancer: value of gadopentetate dimeglumine enhancement, AJNR Am. J. Neuroradiol. 1993; 14(5): 1059–1069.

[22] Zacharaki EI, Hogea CS, Shen D, Biros G, Davatzikos C. Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth. Neuroimage. 2009 Jul; 46(3):762-74.

[23] Clark CH, Miles EA, Urbano MT, et al. Pre-trial quality assurance processes for an intensity-modulated radiation therapy (IMRT) trial: PARSPORT, a UK multicentre Phase III trial comparing conventional radiotherapy and parotid-sparing IMRT for locally advanced head and neck cancer. Br. J. Radiol. 2009; 82(979):585-594

[24] DAHANCA, EORTC, GORTEC, NCIC and RTOG endorsed consensus guidelines for the delineation of the CTV in the N0 neck of patients with head & neck squamous cell carcinoma. [Online] Available: www.RTOG.org [Recent accessed: July 5, 2016]

[25] Guerrero Urbano MT, Clark CH, Kong C, et al., PARSPORT Trial Management Group, Target volume definition for head and neck intensity modulated radiotherapy: pre-clinical evaluation of PARSPORT trial guidelines. Clinical Oncology (Royal College of Radiologists (Great Britain)), 2007; 19(8): 604–613.

[26] Warfield S, Zou K, and Wells W, Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation, IEEE Transactions on Medical Imaging. 2004; 23(7): 903-921.

[27] Dice L, Measures of the amount of ecologic association between species. Ecology 1945; 26:297–302.

[28] Gerig G, Jomier M, Chakos M, A new validation tool for assessing and improving 3D object segmentation. In Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI '01). Vol. 2208 of Lecture Notes in Computer Science 2001; 516–523.

[29] Sims R, Isambert A, Grégoire V. A pre-clinical assessment of an atlas-based automatic segmentation tool for the head and neck. Radiother Oncol. 2009 Dec; 93(3):474-478.

[30] Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans Med Imaging. 1994; 13(4):716-724.

[31] Chong V. Tumour volume measurement in head and neck cancer. Cancer Imaging. 2007; 7(Special issue A): S47–S49.

Supplementary data

Table S1: Quantitative comparison of automatic contours (C-auto) tool against manual contours from two expert Clinical oncologists (C1 and C2)

Patient No.	DSC (C1 vs. C-auto)	DSC (C2 vs. C-auto)	HD (C1 vs. C-auto) (mm)	HD (C2 vs. C-auto) (mm)	Volume Concordance (C1 vs. C-auto) (%)	Volume Concordance (C2 vs. C-auto) (%)
1	0.80	0.85	5.23	5.83	76.77	88.32
2	0.81	0.80	6.68	8.56	82.48	84.78
3	0.79	0.80	7.69	6.65	80.14	83.29
4	0.71	0.77	6.37	6.01	75.36	94.45
5	0.76	0.82	4.11	3.65	91.09	97.35
6	0.79	0.79	10.17	8.22	67.65	72.41
7	0.79	0.83	9.50	6.02	93.87	88.99
8	0.84	0.82	5.43	5.63	97.63	76.58
9	0.78	0.77	7.13	7.94	98.91	96.85
10	0.67	0.85	10.62	6.72	72.38	92.07
Absolute mean	**0.77**	**0.81**	**7.29**	**6.52**	**83.63**	**87.51**
Coefficient of Variation (%)	**5.96**	**3.43**	**28.59**	**21.22**	**12.55**	**9.06**

Underline values in Table S1 illustrate a positive concordance, i.e. greater C-auto volume.

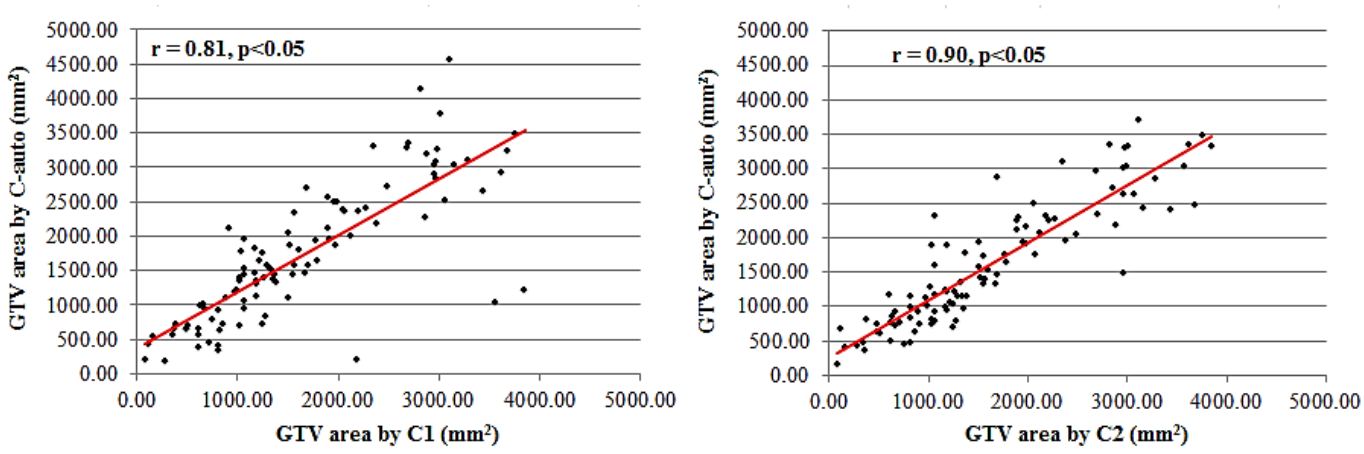


Fig. S1: Correlation analysis of GTV on 102 MRI axial slices
(a) C1 vs. C-auto **(b) C2 vs. C-auto**

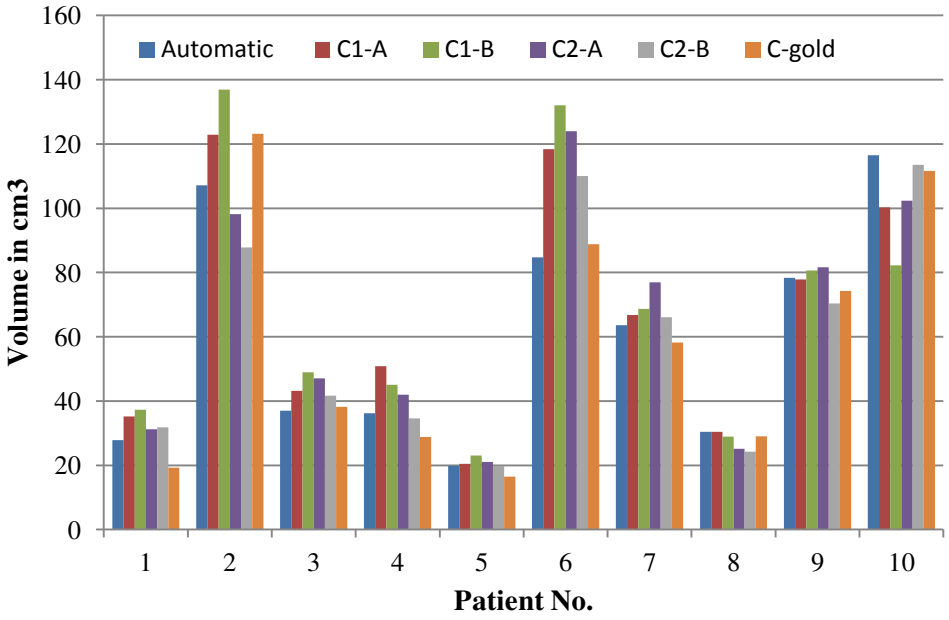


Fig S2: Volume of reconstructed GTV in cm^3 for all 10 patients

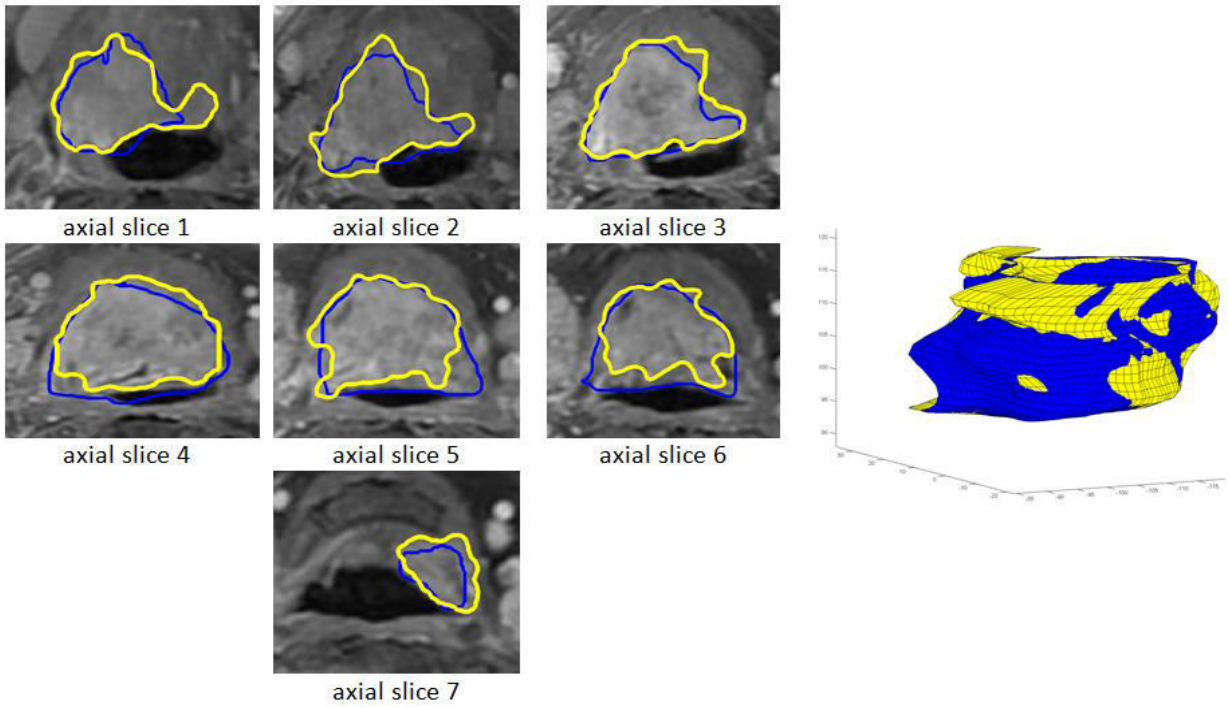


Fig. S3:

(a) Sequential MRI axial slices with automatic (yellow) and gold-standard (blue) GTV outline for one patient

(b) 3D reconstructed volume from GTV outlines shown in Fig. S2 (a) (automatic - yellow; manual - blue)