Research Highlights

- A new fully automatic framework (PLCSF) for MRI pharynx and larynx cancer segmentation.

- Validation of the performance of the proposed framework with approach used in current clinical practice using well-established quality metrics.

- Applicability to MR images obtained from different MRI scanners with different imaging protocols.

- This study might provide a support tool for the clinicians in tumour delineation for radiotherapy treatment planning.

List of abbreviations

BoT	Base of Tongue
Cov	Coefficient of Variation
DSC	Dice Similarity Coefficient
FCM/MFCM	Fuzzy c-means/Modified fuzzy c-means
IIH	Intensity Inhomogeneity
LSM	Level Set Method
MHD	Modified Hausdorff Distance
MR/MRI	Magnetic Resonance/ Magnetic Resonance Imaging
MS	Mean Shift
Ncut	Normalised Cut
PCC	Pearson Correlation Coefficient
PLCSF	Pharyngeal and Laryngeal Cancer Segmentation Framework
RO	Radiation Oncologist
ROI	Region of Interest
RTP	Radiotherapy Treatment Planning
SUSAN	Smallest Univalue Segmentation Assimilating Nucleus
T1+Gd	Gadolinium enhanced T1-weighted

Automatic Pharynx and Larynx Cancer Segmentation Framework (PLCSF) on Contrast Enhanced MR Images

Trushali Doshi[1], John Soraghan[1], Lykourgos Petropoulakis[1], Gaetano Di Caterina[1], Derek Grose[2], Kenneth MacKenzie[3], Christina Wilson[2]

1 Department of Electronic & Electrical Engineering, University of Strathclyde, Glasgow, UK,

2 Beatson West of Scotland Cancer Centre, Glasgow, UK,

3 Glasgow Royal Infirmary, Glasgow, UK.

ABSTRACT

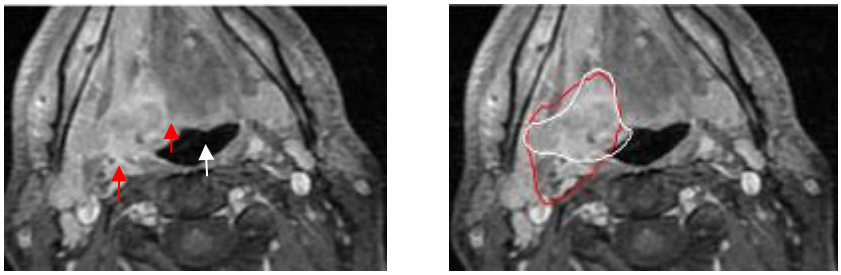
A novel and effective pharynx and larynx cancer segmentation framework (PLCSF) is presented for automatic base of tongue and larynx cancer segmentation from gadolinium-enhanced T1-weighted magnetic resonance images (MRI). The aim of the proposed PLCSF is to assist clinicians in radiotherapy treatment planning. The initial processing of MRI data in PLCSF includes cropping of region of interest; reduction of artefacts and detection of the throat region for the location prior. Further, modified fuzzy c-means clustering is developed to robustly separate candidate cancer pixels from other tissue types. In addition, region-based level set method is evolved to ensure spatial smoothness for the final segmentation boundary after noise removal using non-linear and morphological filtering. Validation study of PLCSF on 102 axial MRI slices demonstrate mean dice similarity coefficient of 0.79 and mean modified Hausdorff distance of 2.2mm when compared with manual segmentations. Comparison of PLCSF with other algorithms validates the robustness of the PLCSF. Inter- and intra-variability calculations from manual segmentations suggest that PLCSF can help to reduce the human subjectivity.

Index Terms - head and neck cancer, automatic segmentation, magnetic resonance imaging (MRI), fuzzy c-means clustering, fuzzy rules, level set method, radiotherapy

1. INTRODUCTION

1.1. Introduction

In recent years, the incidence of head and neck cancer (HNC), and in particular, pharyngeal and laryngeal cancer has increased dramatically due to the influence of the human papillomavirus and other related factors [1]. Approximately 13000 new cases for HNC with 46% of pharyngeal and laryngeal cases are reported each year in the United Kingdom with over 3,300 deaths per year [1]. The definitive treatments of these types of cancer are surgery, chemotherapy and/or radiation therapy (RT) with preferred RT treatment in an effort to preserve the organs [2]. Computed tomography (CT) has been the primary imaging modality in RT for localisation (segmentation) and staging of cancer and treatment planning with secondary information obtained from magnetic resonance imaging (MRI) and/or position emission tomography (PET). However, in recent years, dedicated MRI scanners [3] are being developed in radiation oncology to smoothly integrate MRI in RT planning (RTP) as it demonstrates excellent soft tissue characterization, and has superior diagnostic accuracy when compared to CT [3]. Other benefits of MRI include functional imaging for tumour segmentation and dynamic imaging techniques for motion assessment, all without adding a radiation dose. Thus, localizing (segmenting) the pharynx and larynx cancer from MRI for staging and treatment planning is important in spatially localized RTP. In current manual approach to segment cancer regions from axial MRI slices, radiation oncologists draw the boundary of the cancer regions. This manual segmentation process is time-consuming and subject to inter- (Fig. 1) and intra-observer variations, especially in presence of weak boundaries (Fig. 1(a)). Further, clinical work on pharynx and larynx cancer segmentation involves huge amount of data from different hospital centres. Thus, an automatic segmentation framework that produces quantifiable and repeatable segmentation results for the data obtained from different MRI scanners is highly desirable.



(a) (b)

Fig. 1: (a) Original T1+Gd (gadolinium enhanced T1-weighted) MRI slice with two red arrows to illustrate the weak and non-distinct boundary of the base of tongue cancer region, white arrow illustrate the throat region (b) Inter-variability in cancer segmentation drawn by two experts (red and white outline)

1.2. Related work

Many automatic and semi-automatic methods have been proposed for cancer segmentation from MRI. Some techniques [4-5] use multi-spectral MRI to segment cancer regions. However, obtaining multi-spectral MRI data is not always feasible and is expensive. Furthermore, multi-spectral MRI data may require registration step prior to segmentation due to misalignment and inconsistency. Some other techniques to segment cancer regions from single modality MRI include seed-growing [6], watershed method [7] and fuzzy connectedness [8]. These techniques do not consider spatial constraints and thus are sensitive to noise and other MRI artefacts such as intensity inhomogeneity (IIH) [9]. Further, semi-automatic approach in [6] requires manual-placing of seed points or drawing of a close loop outside the tumour from expert to segment the tongue cancer.

Active contour (AC) models [10] are also used in tumour segmentation for 2D and 3D datasets. AC is improved in [11] for concave shape brain tumours segmentation. Ho et al. [12] utilised cancer probability map as an initialization for the evolution of a level-set algorithm to segment blobby-shaped 3D cancer. Tongue cancer segmentation using manual initialization for level-set is proposed in our previous work [13]. AC models, however, are sensitive to initialization even when using 3D level set surface.

Clustering techniques, due to their robustness and efficiency, are prevalent for cancer segmentation task [14-18]. Fuzzy c-means (FCM) clustering [19] technique is modified in [14] by adding neighbourhood spatial information to correct IIH and segment brain cancer image. This technique only works with salt and pepper noise and cannot compensate for severe IIH. Prior information on cluster centres and uncertainty modelling is considered in FCM in [15] to improve FCM performance under noise and variation in data acquisition. This technique is good only if cluster centres information is known prior. FCM has been modified in our previous work [16] to segment base of tongue (BoT) cancer from magnetic resonance (MR) images. Non-parametric mean shift (MS) clustering [20] is employed in [17] as an initial step for clustering similar voxels in multidimensional feature space for breast lesion segmentation. Spectral graph clustering method notably normalised cuts (Ncut) algorithm [21] is used for cancer segmentation in [18].

In this paper we present a new pharynx and larynx cancer segmentation framework (PLCSF) for automatic segmentation of BoT and larynx (voice box) cancers from 2D (axial) contrast (gadolinium)-enhanced T1-weighted (T1+Gd) MRI slices. The aim of this work is to assist clinicians in RTP by obtaining quantifiable and repeatable segmentation results in an unbiased manner. T1+Gd MRI compared to unenhanced (normal) T1, proton-density and T2-weighted MRI is superior to define tumour spread for BoT and larynx cancers [22-23], as it significantly improves soft tissue contrast and cancer margin definition. However, even with T1+Gd MRI (Fig. 1), pharynx and larynx cancer segmentation is a challenging task due to variability in its geometry, presence of necrotic tissues in the cancer region, no distinct boundaries between tumourous and healthy tissues (Fig. 1), overlap of feature values of the cancer and non-cancer pixels, and the presence of MRI artefacts. Further, significant inter- and intra- intensity variations in MRI data across patients and highly anisotropic MRI slices (maximum slice spacing 6 mm) used in this work make it reasonable to process each axial slice separately to obtain satisfactory segmentation results.

The main objective of this work is to provide the robust cancer segmentation framework (PLCSF) that integrates spatial information in an unsupervised technique without requiring the use of complex statistical modelling, atlas or the training data. The novel contributions in this PLCSF are as follows: 1) to the best of our knowledge there is no computer-aided system in the literature that is focused on the automatic segmentation of BoT and larynx cancer from T1+Gd MRI slices; 2) the technique described in [9] is modified in this work in terms of spline distance (knot spacing) [24] parameter for bias field (IIH) estimation from a MRI slice 3) the algorithm developed for the throat region detection is novel in itself; 4) a novel technique based on FCM clustering is developed to robustly separate different tissue types in different clusters; 5) our approach does not require any manual interaction or different modalities of MRI.

The organisation of the remainder of the paper is as follows. The framework (PLCSF) developed to segment BoT and larynx cancer is presented in Section 2. Real dataset, comparison algorithms (MS clustering and NCut) and evaluation parameters used for comparison are described in Section 3. Experimental results are reported in Section 4. Finally, discussions and conclusion are presented in Section 5.

2. PHARYNX AND LARYNX CANCER SEGMENTATION FRAMEWORK (PLCSF)

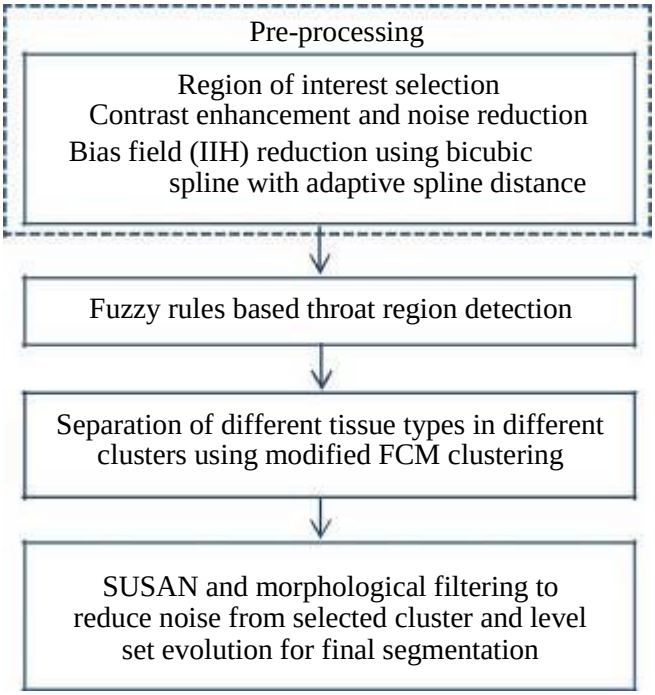


Fig. 2: Flowchart of the pharynx and larynx cancer segmentation framework (PLCSF).
IIH – Intensity Inhomogeneity, FCM – Fuzzy c-means, SUSAN - Smallest Univalve Segment Assimilating Nucleus

2.1. Overview

The flowchart in Fig. 2 illustrates the steps of the proposed PLCSF for BoT and larynx cancer segmentation from axial T1+Gd MRI slices. Each axial (2D) MRI slice is processed separately. The selected region of interest (ROI) is processed to increase the contrast between different tissues and to reduce background noise and IIH. Anatomical (Throat) regions which are close to the expected cancer location are detected from the ROI. Modified FCM which includes a squared Euclidean distance measure created from the average point of the throat region is used to separate different tissue types in different clusters. Further, (Smallest Univalve Segment Assimilating Nucleus) SUSAN technique [25] and morphological filtering [26] are applied to the selected cluster to reduce noise and wherever appropriate concavity checks are utilized for labelling of expected cancer regions. Finally, the labelled region is used as an initialization for the localized region based level set

evolution [13] which converges to obtain the final segmentation boundary.

2.2. Pre-processing

The ROI is selected automatically from the original MRI slice using our technique described in [16], to reduce computational time and complexity. The contrast (intensity difference) among different tissue types present in a ROI is increased using background brightness preserving histogram equalization (BBPHE) [27]. For BBPHE, in this work the ROI is divided into two sub-images using the first local minima of the ROI histogram. Further, background noise from ROI is reduced using log-exponential transformation [26].

(a) (b) (c)

Fig. 3: (a) MRI slice (b) estimated bias field (c) bias field reduced MRI slice

Additionally, MR images used in this work suffer from low frequency intensity variation also called bias field or IIH. An estimated bias field from Fig. 3(a) is shown in Fig. 3(b). Due to this artefact, pixels belonging to one tissue type have varying intensities. To classify pixels belonging to one tissue type in one cluster, it is necessary to correct this IIH. The method described in [9] is used as a basis to reduce this IIH.

An initial bias field (IIH) is estimated by fitting a third order polynomial function to the data excluding background pixels using least square approximation. The resulting bias field is then refined using a bicubic spline model. The important parameter in bicubic spline model is the ‘spline distance’ i.e. the distance between the knots (knot spacing) defining bicubic spline mesh [24] to correctly estimate the bias field. In [9], this parameter, spline distance, is set to a constant value

(21mm) for all patients.

In this work, spline distance is determined adaptively for each axial MRI slice using the technique described in [28] that is applied separately for rows and columns (Fig. 4(a)).

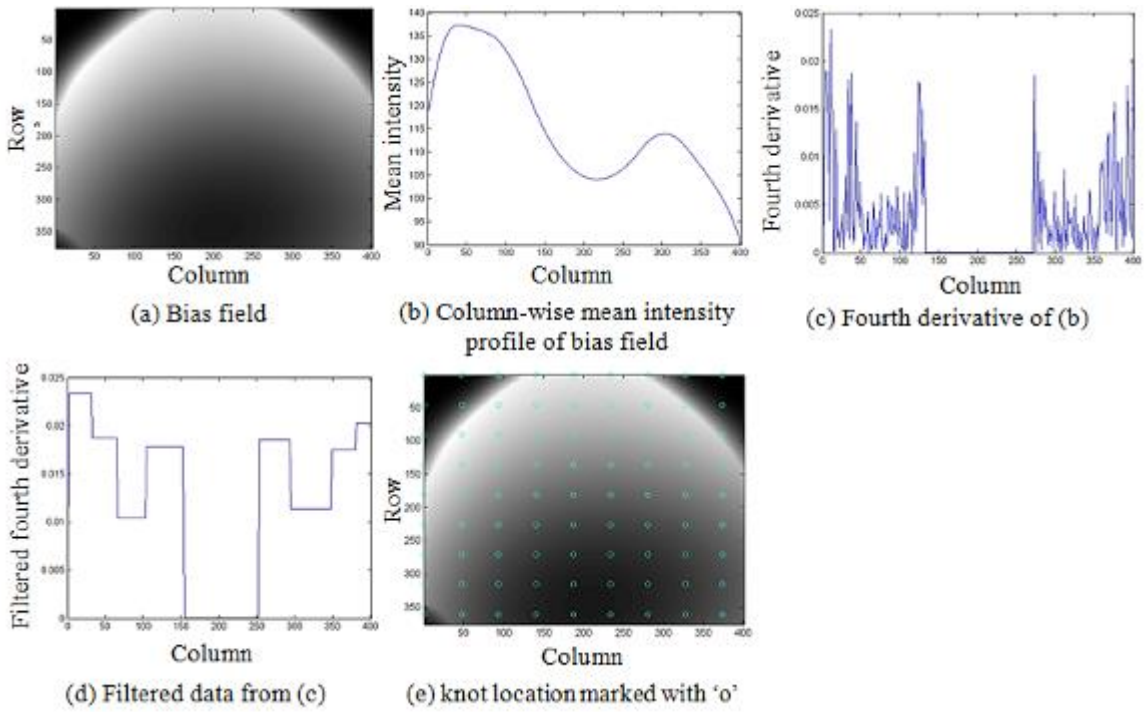


Fig. 4: The procedure to determine spline distance parameter (distance between consecutive knots) for bias field reduction

To determine spline distance, the fourth derivative of the data is calculated. This data is divided into distinct regions after passing it through a maximum filter [29] of window size 20. The value of window size is chosen so as to remove undesired high frequency components while avoiding excessive smoothing and merging of adjacent regions. The number of knots N_k in each distinct region is computed [28] as:

$$N_k = \left\lfloor \frac{r_{i+1} - r_i}{\sqrt[4]{\frac{\varepsilon}{KM_i}}} \right\rfloor \tag{1}$$

where r_i is the region under consideration, K is a constant and M_i is the fourth derivative value of a region under consideration. The error value ε is determined from the region with global maximum value of the fourth derivative. The total number of knots is obtained by summing the number of knots in each distinct region. The distance between consecutive knots (spline distance) is subsequently calculated, as the number of elements in data divided by the total number of knots. The process of determining knot spacing for column from single axial MRI slice is illustrated in Fig. 4.

With knots spaced evenly across the image, the estimated bias field value at each knot location is used for an optimization. The optimization is carried out by minimizing the local entropy of the corrected ROI, starting with knot with the highest estimated bias field value, and merging areas of lower values in sequential fashion [9]. Fig. 3(c) shows a typical IIH reduced image.

2.3. Fuzzy rules based throat region detection

In this work it is known that the BoT and larynx cancer will be adjacent to the throat region, as illustrated in Fig. 1(a). This information is used in FCM extraction of the cancerous regions. A combination of thresholding and fuzzy rule-based methods is used for the extraction of the throat region. The ROI is converted to a binary image B_I using a thresholding method described by Otsu [30]. A signature [26] of each connected component is calculated from the binary image. A signature is a 1-D representation of the boundary of an object, calculated as distance between pixels on the boundary of an object and its centroid. Any object with a maximum signature radius greater than 30mm is removed from the upper half of the binary image to reduce false positives. The value 30 is decided with empirical tests in which a range of values were systematically tested in turn to assess the effect on detection results. Furthermore, the knowledge of the throat position is incorporated in two fuzzy rules to classify pixels as the candidates for the throat region:

1) The throat region is present in the central part of the binary image. Pixels are associated with the throat region using a fuzzy vertical line membership function F_{VL} . F_{VL} is calculated from the column (Fig. 4(a)) centre (c_v) and it is given as:

$$F_{VL} = \{(i, j), \mu_{VL}(i, j)\} \quad i = 1..r_n, j = 1..c_n \tag{2}$$

with

$$\mu_{VL}(i, j) = f_{VL}(j) = \begin{cases} 1, & |c_v - j| \le a \\ 1 - \dfrac{|(c_v - j) - a|}{b - a}, & |c_v - j| > a \,\&\&\, |c_v - j| \le b \\ 0, & Otherwise \end{cases}$$

where r_n is the number of rows and c_n is the number of columns, a is set to 5 and b to 20. Thus, the highest membership value of 1 is given to c_v and its immediate neighbours gradually decreasing the membership values on either side as distance increases from the centre.

2) The throat region is close to the first row compared to the other objects in the middle part of the binary image. Thus, fuzzy height membership rule F_H is used to associate the closeness of pixels to the first row of the binary image.

Therefore, F_H is given as:

$$F_H = \{(i, j), \mu_H(i, j)\} \quad i = 1..r_n, j = 1..c_n \tag{3}$$

with

$$\mu_H(i, j) = f_H(i) = \begin{cases} 1, & i == c \\ \dfrac{|d - i|}{d - c}, & i > c \,\&\&\, i \le d \\ 0, & Otherwise \end{cases}$$

where is set to 1 and d to r_n .

The fuzzy ‘min’ operator is used to select the minimum of these two membership values for each element. It is given by:

$$F_{VLH} = F_{VL} \cap F_H \tag{4}$$

$$\mu_{F_{VLH}}(i,j) = min(w_1 * \mu_{VL}(i,j), (1-w_1) * \mu_H(i,j))$$
$$i = 1..r_n, j = 1..c_n$$

The weights w_1 and $1-w_1$ are assigned to the membership values to vary the relative importance between $\mu_H(i,j)$ and $\mu_{VL}(i,j)$. Finally, candidate pixels for the throat region are chosen as:

$$F_{Intersect} = F_{VLH} \cap B_I \tag{5}$$

$$\mu_{F_{Intersect}}(i,j) = \min(\mu_{VLH}(i,j), B_I(i,j))$$
$$i = 1...r_n, j = 1...c_n$$

$F_{Intersect}$ of these two fuzzy rules is demonstrated in Fig. 5(a). From $F_{Intersect}$ pixels with highest membership values are chosen as pixels belonging to the throat region. The detected throat region from representative MRI slice is illustrated in Fig. 5(b).

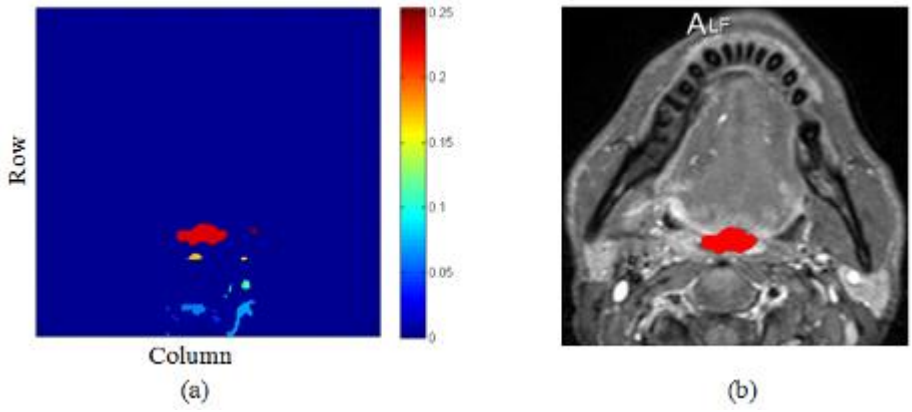


Fig. 5: (a) $F_{Intersect}$ (b) detected throat region (in red) from MRI slice

The squared Euclidean distance measure is created from the centre (t_r, t_c) of the detected throat region and is given as:

$$d_s^2(i,j) = \sum_{j=1}^{c_n} \sum_{i=1}^{r_n} ((i-t_r)^2 + (j-t_c)^2) \tag{6}$$

where c_n and r_n are number of columns and number of rows in the ROI respectively. This distance matrix is normalized and converted into one a dimensional feature vector.

2.4. Expected cancer region extraction using modified fuzzy c-means (MFCM) clustering

The standard FCM [19] is an iterative process that produces optimal C clusters of the grayscale image $\{I_k\}_{k=1}^{N}$ where N is number of pixels, by minimizing the following objective function:

$$J_T = \sum_{i=1}^{C}\sum_{k=1}^{N}\mu_{ik}^{m} d^2(v_i, I_k) \tag{7}$$

where v_i is the cluster centre of cluster i . Here C is the number of tissue types in the image. $\mu_{ik} \in [0,1]$ is the membership degree of k^{th} pixel to i^{th} cluster. m (> 1) controls the fuzziness of membership function μ_{ik} . $d(v_i, I_k)$ is the Euclidean distance between v_i and pixel intensity I_k . The membership functions are subject to the following constraints:

$$\sum_{k=1}^{N}\mu_{ik} > 0 \quad \forall i \in \{1,\ldots,C\}; \qquad \sum_{i=1}^{C}\mu_{ik} = 1 \quad \forall k \in \{1,\ldots,N\} \tag{8}$$

The squared Euclidean distance measure (6) is added to the objective function of the standard FCM. The intuition of this additional distance measure is that the probability of being cancer pixel is proportional to its distance to the throat region pixels. This leads to assigning a high membership values to the pixels when the distance to the throat region pixel is small and vice versa. The modified objective function is given as [16]:

$$J_{ST} = \sum_{i=1}^{C}\sum_{k=1}^{N}\mu_{ik}^{m} d^2(v_i, I_k) + \sum_{i=1}^{C}\sum_{k=1}^{N}\mu_{ik}^{m} {d_s}^2(i,k) \tag{9}$$

In (9) J_{ST} can be minimized under the same constraint of μ_{ik} as in (8).The membership functions μ_{ik} and cluster centres v_i are updated iteratively as follows:

$$\mu_{ik} = \frac{(d(v_i, I_k) + d_s(i,k))^{-2/(m-1)}}{\sum_{j=1}^{C} (d(v_j, I_k) + d_s(j,k))^{-2/(m-1)}} \quad \text{and} \quad v_i = \frac{\sum_{k=1}^{N} \mu_{ik}^m I_k}{\sum_{k=1}^{N} \mu_{ik}^m} \tag{10}$$

To reduce the sensitivity to noise, local spatial information described in [31] is employed in this membership function.

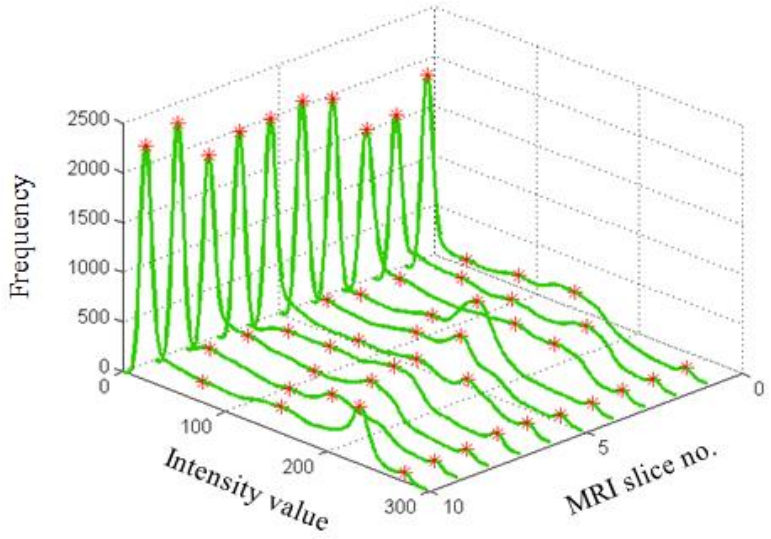
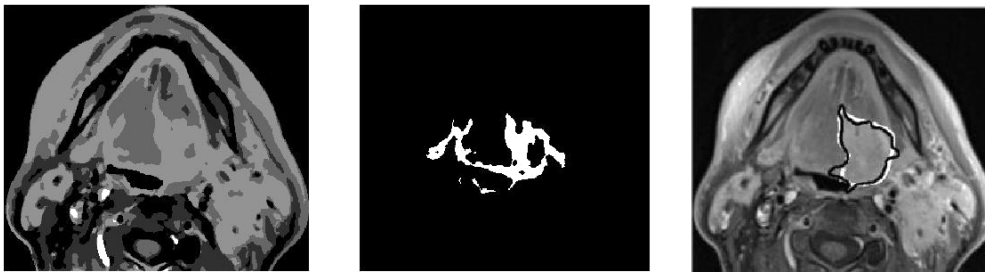


Fig. 6: Histogram of representative MRI slice from 10 patients showing 5 peaks (red stars) demonstrating five clusters for each slice

Using the MFCM, the IIH corrected ROI is divided into five clusters as it is known that the ROI consists of four main tissue types (fatty tissues, cancer tissues, normal tongue/ larynx tissues, and normal muscle tissues) and background (Fig. 6). Five clusters of the pre-processed slice are shown in Fig. 7(a). It is known that gadolinium-enhanced cancer pixels occupy the higher end of the histogram. This is utilized to select the cluster with the expected cancer regions. From selected cluster, pixels with membership value ≥ 0.8 are considered (labelled) as the candidates for the cancer pixels (Fig. 7(b)).



(a) (b) (c)

Fig. 7: (a) Five clusters represented with different grey-levels (b) cluster with expected cancer region (c) smooth boundary of the labelled cancer region (black outline), manual outline (white)

2.5. Labelled region boundary refinement

Further, noise and isolated speckles present in the selected cluster are reduced using SUSAN technique [25] and morphological filtering [26]. The SUSAN algorithm uses non- linear filtering to reduce noise while preserving edges. A mask of 3x3 and brightness threshold of 10 was used to discriminate noise and underlying image features. Morphological filtering here is a closing operation followed by an opening process with disk shape structuring element of radius 5. This filtering operation removes small regions while it can preserve topology of large regions.

For some MRI slices, the labelled cancer region is around the throat region. To decide the throat region inclusion in a detected region, a concavity measure [26] of the detected region is computed. Concavity measure is obtained by dividing the area of object with its convex area. If the concavity of the detected region is less than certain value, the throat region is included in the detected cancer region, otherwise it is excluded.

Finally, the level set function (signed distance function) where zero (initial) level set represents the boundary of the estimated cancer region is evolved to obtain the final smooth boundary of the labelled cancer region. In this work the statistics (mean and area) of the local interior

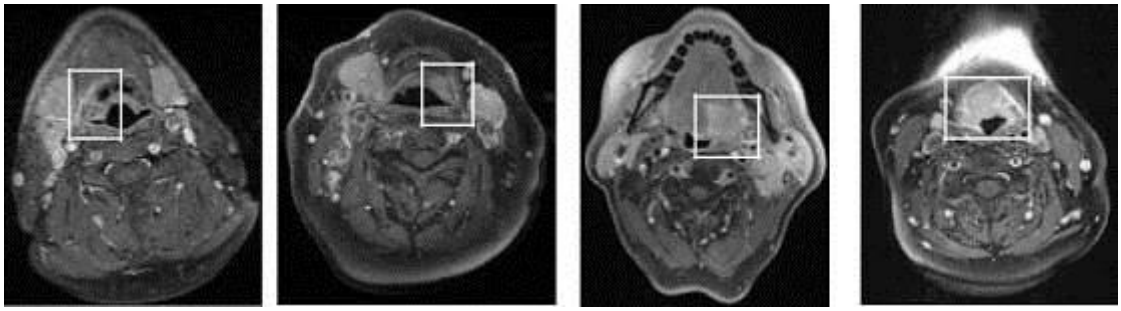
and exterior region around each point on the zero level set [13] is considered as a force for evolution. The mean separation energy where optimization of energy occurs when foreground and background have maximally separate mean intensities is used to stop the evolution [13]. When the level set evolution stops after 500 iterations, the final tumour segmentation boundary is obtained. Obtaining the smooth boundary of the cancer region using level set is important in radiation oncology as clinical expert outlines the cancer borders using contouring for radiotherapy planning as opposed to pixel by pixel labelling as done in clustering algorithms. This smoothing process is expected to obtain a cancer boundary close to the expert's manual contouring. The qualitative result of level set evolution is shown in Fig. 7(c).

3. DATASETS AND EVALUATION METHODS

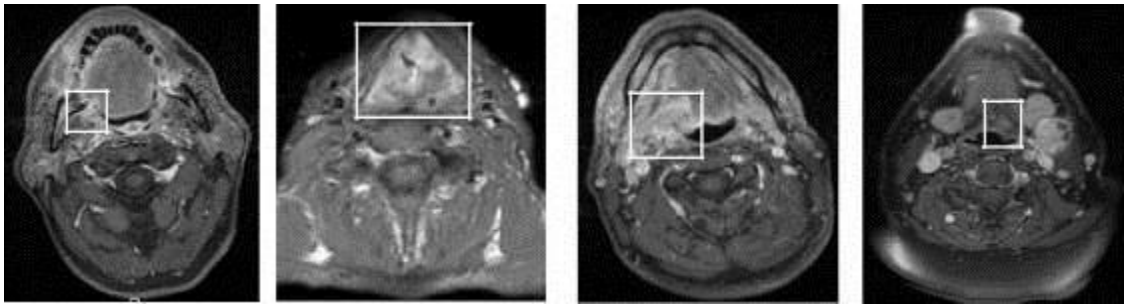
3.1. Real MRI data set

The performance of the PLCSF algorithm was evaluated on real axial T1+Gd MRI slices from 10 patients (6 base of tongue (BoT) and 4 laryngeal cancers). All MRI scans were obtained before radiotherapy treatment from three different 1.5Tesla MRI scanners namely Magnetom Avanto from Siemens, Intera Neuro coils from Philips Medical Systems, and Signa HDxt from GE Medical Systems.

T1+Gd MRI scans were acquired after 15-20 minutes of intravenous injection of 0.1ml/kg, with typical 3-5mm slice thickness. The range of other imaging parameters were, 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 in-plane resolution, 256x256-512x512 acquisition matrix and 97.65-221 Hz/pixel bandwidth. All images were obtained in standard digital imaging and communications in medicine (DICOM) format. In total 102 (55 BoT and 47 laryngeal) axial T1+Gd MRI slices from 10 patient with visible cancer regions were selected to validate the proposed PLCSF algorithm. Eight axial T1+Gd MRI slices from the dataset used to validate the PLCSF algorithm is shown in Fig. 8.



(a)



(b)

Fig. 8: Typical real MRI dataset used to validate PLCSF (pharyngeal and laryngeal cancer segmentation framework). Different contrast uptake, shape and size of the cancer regions can be observed (a) MRI slice with cancer region that is not fully enhanced. (b) MRI slice with cancer region diffused in normal tissues with no distinct edge.

3.2. Manual segmentation procedure and Evaluation parameters

To obtain the reference segmentation for comparison with the automatically obtained results from the PLCSF, the MRI scans were peer reviewed in the oncology centre at the head-and-neck radiotherapy weekly team meeting. From that meeting, expert general consensus on the tumour outline on 2D (axial) slices for all 10 patients, which is considered as a gold standard in current clinical practice was obtained. In order to analyse inter-variability in manual segmentation, two independent radiation oncologists' (RO1 and RO2), sub-specializing in head-and-neck cancer and with the experience of approximately 10 years, manually outlined the cancer region in all axial slices according to the published guidelines [32]. For intra-variability evaluation both RO1 and RO2 repeated this procedure on the same dataset approximately one month later. The PLCSF results were also compared to RO1 and RO2 outlines.

For PLCSF performance assessment, two metrics were utilized; Dice Similarity Coefficient (DSC) [33] and Modified Hausdroff Distance (MHD) [34]. Spatial overlap between two segmentation results was measured using DSC. A high value of DSC (i.e.1) indicates good agreement between two segmentation results. Compared to original Hausdroff distance, MHD reduces impact of outliers and noise and it was used for shape variation evaluation between segmentation results. In addition, relative area difference (RAD) was calculated as:

$$RAD = \frac{\left|Area(manual) - Area(a\lg orithm)\right|}{\left|Area(manual)\right|} x100\% \tag{11}$$

where $Area(.)$ is the area of either manual or automatically segmented cancer region. The Pearson correlation coefficient (PCC) with p-value is used to test the statistical significance of the PLCSF. The PCC value of -1 indicates negative correlation, 0 indicates no correlation and 1 indicates positive correlation.

3.3. Other algorithms

The MFCM clustering method proposed in PLCSF was compared to two other clustering algorithms in the literature used for cancer segmentation [17-18]. One was mean-shift (MS) clustering [20] and other was spectral clustering (Ncut) [21] technique. The comparison was performed on MRI dataset described in Section 3.1.

MS clustering algorithm was chosen due to its unsupervised non-parametric nature that does not require prior knowledge on number of clusters or cluster positions. The method uses a gradient ascent technique to detect local maxima of data density in feature space. The data points associated with same local maxima are considered as a member of the same cluster.

Ncut [21] is a graph based segmentation method. Ncut is included in comparison due to its popularity among spectral clustering methods. Ncut views an image as a graph, where image pixels represent nodes and edges between nodes are weighted according to the similarity between pixels. In this work, brightness (intensity) and spatial location was used to assign weights to the edges. The aim of Ncut is to find a partition of a graph that minimizes the normalised cut between two parts. The normalised cut avoids bias for partition of small set of points normally observed in graph-cut approach [21]. The partition of the image is obtained by thresholding the resultant eigenvectors.

For both algorithms (MS and Ncut), the input was the pre- processed MRI slice obtained from Section 2.2. From the clusters (partition) obtained from both algorithms only the bright (intensity-wise) cluster near the throat region was further processed using the techniques described in Section 2.5. Parameters for both algorithms were chosen so as to give the best performances across the MRI dataset. For implementation of MS clustering and Ncut, we used publicly available Matlab codes [35-36].

4. EXPERIMENTAL RESULTS

The proposed PLCSF was implemented in Matlab 2014a (MathWorks, Natick, MA), on a Windows 7 system. The software was executed on Dell U2412M with Processor Intel Xeon E5-1607 3.00 GHz and 8GB RAM.

4.1. Spline distance comparison: constant vs. adaptive

As described in section 2.2, in this work spline distance, i.e. the distance between the knots defining bicubic spline mesh, is determined adaptively as opposed to the literature [9] where this distance is set to a constant value of 21mm for each MRI slice. The spline distance in this work ranged from 11mm to 38mm for 102 axial MRI slices from three different MRI scanners. Experiments were performed by comparing a set of constant spline distances (21mm, 30mm, 60mm) against adaptive distance for each MRI axial slice for 10 patients. Coefficient of variation (CoV) for cancer region segmented using manual consensus outline from corrected MRI slice was calculated and is given as:

$$CoV = \frac{\sigma_{corrected}}{\mu_{corrected}} \tag{12}$$

where σ and μ are standard deviation and mean respectively. From Fig. 9 it can be noticed that CoV value for adaptive distance is less compared to other constant values except for one patient (Patient 2). The requirement for adaptive spline distance for the MRI dataset in this work can be attributed to different range of imaging parameters, variability in cancer area (range: 89.97-3361.46 mm^2) and significant inter- and intra- intensity variations in MRI data across patients.

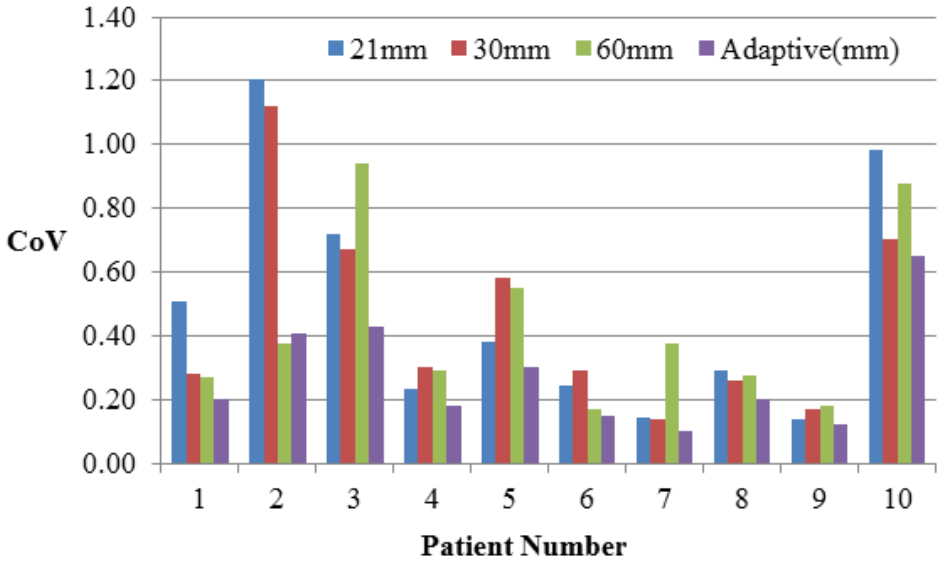


Fig. 9: Coefficient of variation (CoV) for different knot spacing (spline distance) parameter to reduce the bias field.

4.2. Comparison with manual segmentation results

The visual comparison of the segmentation results between the PLCSF and the gold standard (consensus manual outline) for each patient for single axial MRI slice is demonstrated in Fig. 10. In Fig. 10: Patients 1 and 2 demonstrate that the PLCSF system can effectively segment small cancer regions (area: 89.97-246.67mm^2). Other examples in Fig. 10 such as Patients 3 to 8 demonstrate the variability in cancer segmentation outlines. For some MRI slices (Fig. 10: Patients 3 and 4), where the cancer region depicts intensity range close to surrounding tissues and the non-distinct boundary between surrounding and cancerous tissues leads to a slight over segmentation by PLCSF. Inclusion or exclusion of the throat region in the outline may also influence the segmentation results as illustrated in Fig. 10: Patient 5. There is no particular guideline as to include or exclude throat region from segmentation results. A variation in the segmentation results (DSC: 0.65) between manual and PLCSF outline is shown in Fig. 10: Patient 6. This variation is mostly observed in MRI slices where the cancer region is small and often fused into nearby normal structures. For example in Fig. 10: Patient 6, PLCSF overestimated cancer region by including normal palatine tonsils in the outline.

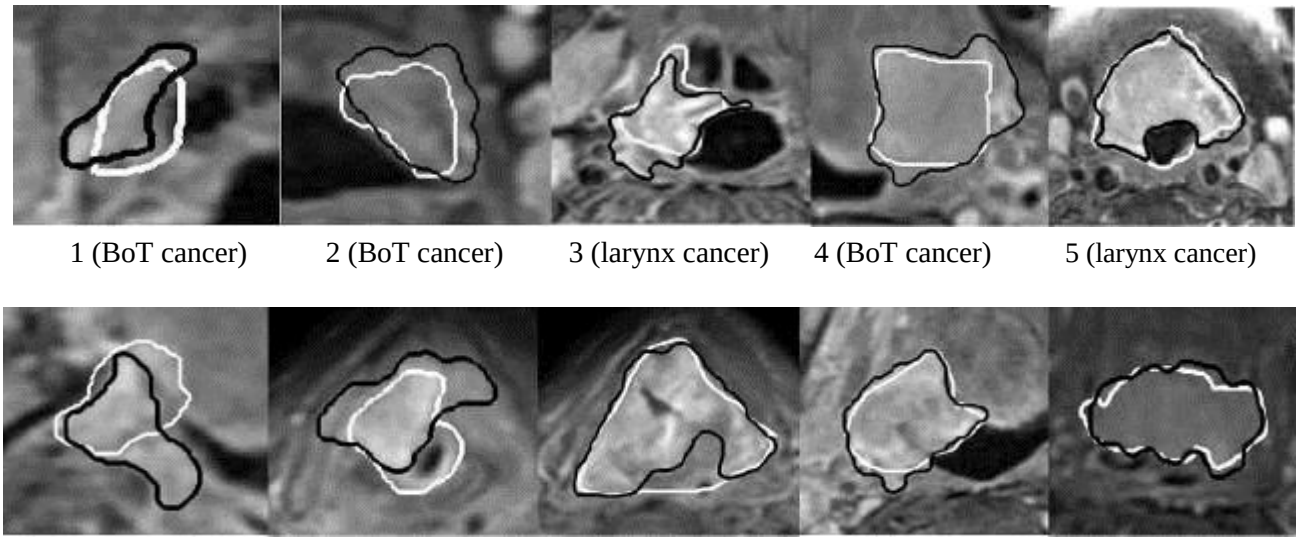


Fig. 10: Examples of PLCSF segmentation results (black outline) on small cancer regions or challenging cases, superimposed with the gold standard (consensus manual outline) (white outline) from real MRI dataset.

The results for laryngeal cancers, Fig. 10: Patients 7 and 8, demonstrate less agreement (DSC: 0.7-0.8) between manual and PLCSF results as compared to BoT cancer, Fig. 10: Patient 9 and 10, which has a DSC: >0.9. The possible reason for this bias is the distinct characteristics of cancer tissues as compared to surrounding tissues in BoT cancer than laryngeal cancer.

The performance measures calculated between the manual outlines from different experts and PLCSF are shown in Table 1.

Table 1: Quantitative measures between manual outlines from experts and PLCSF outlines (mean (standard deviation))

Comparison	DSC	RAD (%)	MHD (mm)
Consensus manual outline vs. proposed method	0.79 (0.05)	39.51 (17.04)	2.15 (0.64)
RO1 vs. proposed Method	0.77 (0.05)	38.07 (15.79)	2.48 (0.90)
RO2 vs. proposed method	0.81 (0.03)	24.50 (8.24)	1.97 (0.46)

The mean DSC of 0.79 and RAD of 34.03% for all 10 patients indicates acceptable [11, 17] agreement between manual segmentation results and PLCSF. The standard deviations show the range of performance of PLCSF for different patients. In some MRI slices, small DSC and large area difference are due to the shape variation between manual and automatic results, as the PLCSF segments concavity region in detail than manual results (Fig. 10: Patient 8).

The box plot demonstrating distribution of DSC is shown in Fig. 11. The MHD of average 2.2 mm indicates that the estimated cancer boundaries are comparable to the expert outlined boundaries.

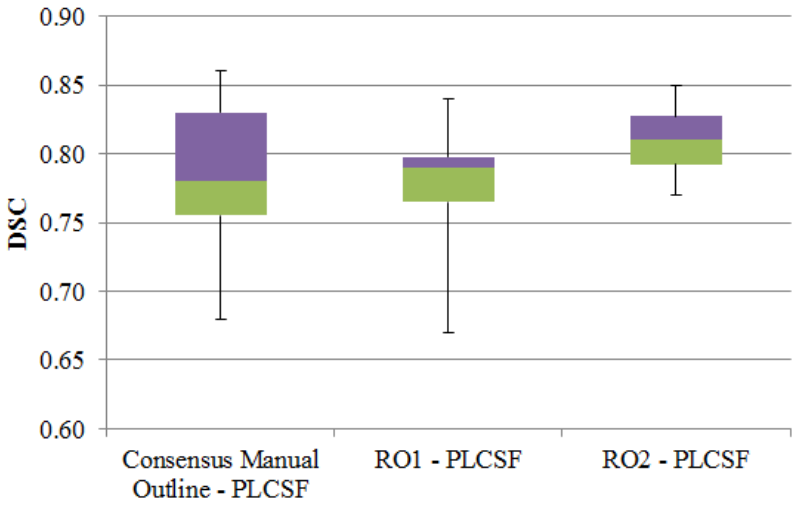


Fig. 11: Box plot showing DSC (dice similarity coefficient) distribution between manual and PLCSF (pharyngeal and laryngeal cancer segmentation framework) results.
RO1 – Radiation Oncologist 1, RO2 – Radiation Oncologist 2

The Pearson correlation coefficient (PCC) between PLCSF results and the gold standard (consensus manual outline), is 0.89 with $p<0.05$ (Fig. 12) verifying statistically significant results in terms of cancer area. Similarly, PCC between PLCSF cancer area and RO1 and RO2 cancer area is 0.81 and 0.90 respectively with $p<0.05$. The PCC between PLCSF and consensus manual outline for BoT cancer is 0.95 as compared to 0.83 ($p<0.05$) for laryngeal cancer indicating that PLCSF performs better for BoT cancer than for larynx cancer.

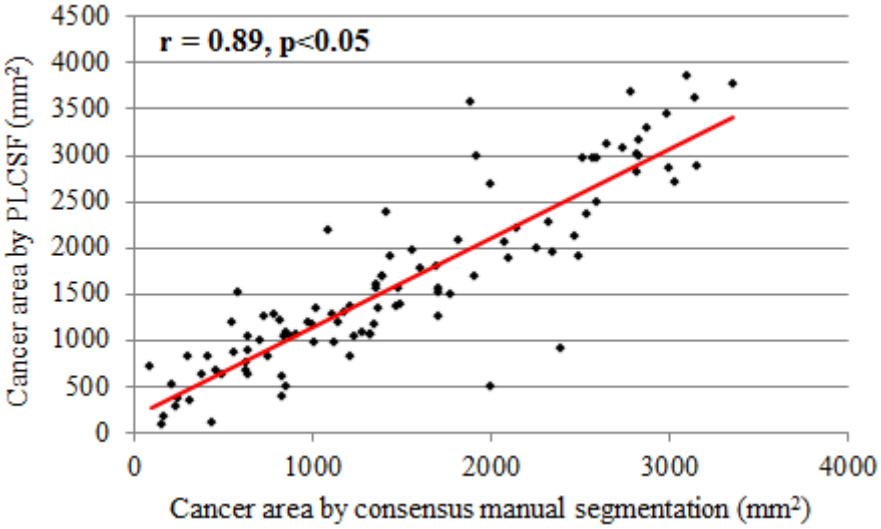


Fig. 12: Correlation analysis between PLCSF and consensus manual outline cancer areas.

4.3. Inter- and Intra-operator variation calculations

Inter-operator variation was calculated between manual segmentations from RO1 and RO2 for the same dataset. The variation in the area was calculated in percentage (%) as the difference between cancer areas from RO1 and RO2 divided by the average of two areas. The average variation in area is 25.28%, mean DSC is 0.80 and mean MHD is 2.24mm. These values are consistent with comparison of PLCSF and manual segmentation results. The PCC between RO1 and RO2 in terms of cancer area is 0.88 ($p<0.05$). This PCC value suggests that there is no significant difference in RO1 and RO2 outline. However, RO1 over-estimates the cancer region as compared to RO2 due the practice to draw the cancer boundary just outside the detected cancer edges rather than drawing them on the cancer edge.

The intra-variability between first and second manual segmentation from RO1 in terms of mean DSC, MHD is 0.85 and 1.67mm respectively and for RO2, DSC is 0.86 and MHD is 1.39mm.

4.4. Comparison of PLCSF with other algorithms

Qualitative results in Fig. 13 shows comparable performance of PLCSF with manual outline (Fig. 13(e), (f)) as compared to Ncut and MS clustering algorithms.

For Ncut algorithm, it is well known that this clustering is biased towards partitioning slice into equal segments. This bias affects the segmentation results specifically in case of small cancer regions (Fig. 13(a)). For MRI cases with low variation between cancer and surrounding tissues, Ncut shows low performance due to segmentation leakage (Fig. 13(b)). Furthermore, this algorithm is computationally intensive. The average time for cancer outline per axial slice using Ncut is 82 seconds compared to average of 45 seconds (in non-optimized Matlab code) per slice for PLCSF.

The results we observed with MS clustering shows that it tends to over-segment the MRI slice giving large number of clusters rather than finding the right object (Fig. 13(c), (d)). Furthermore, the varied number of clusters obtained by MS clustering for each MRI slice makes it

challenging to select cluster with candidate cancer pixels. The MFCM used in this work clustering lack this issue as number of clusters was fixed.

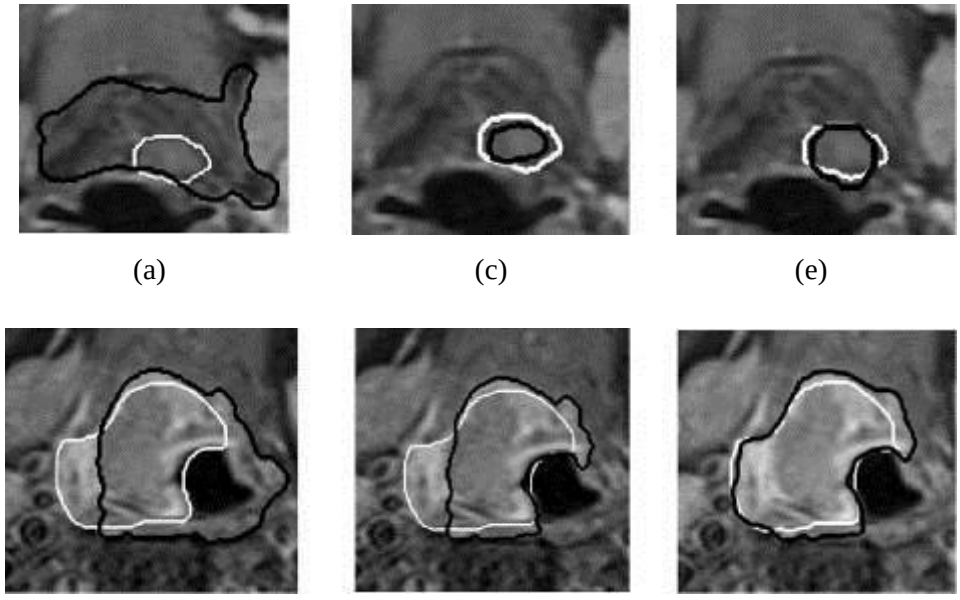


Fig. 13: Qualitative comparison of segmentation results. Results of proposed PLCSF framework (column 3 – black outline) are qualitatively more similar to the Gold Standard (consensus manual segmentation) (white outline) as compared to Ncut (Normalised Cut) (column 1 –black outline), and Mean-shift clustering (column 2 –black outline).

DSC comparison results for these three methods for all 10 patients to the gold standard (consensus manual outline) as a reference are shown in Fig. 14. This comparison yielded mean DSC of 0.71 for Ncut, 0.75 for Mean Shift and 0.79 for proposed PLCSF algorithm.

The PLCSF algorithm exhibits a lower CoV (0.07) on average for DSC compared to Ncut (0.11) and MS (0.09) clustering. Comparison of MS clustering and PLCSF using paired t-test shows statistically significant improvement in DSC (p-value: 1.44×10^{-3}) with the PLCSF method.

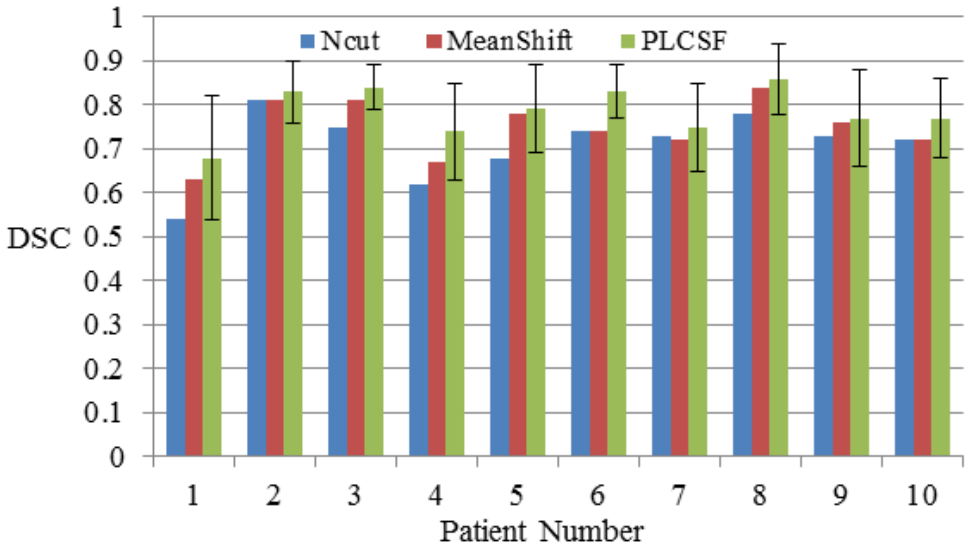


Fig. 14: Comparison of DSC (Dice Similarity Coefficient) for Ncut (Normalised Cut), Mean-shift and the proposed method on real MRI data set, demonstrates improved spatial overlap with the proposed method. Black vertical bar indicates ± standard deviation for MRI slices for each patient for the proposed method.

5. DISCUSSION AND CONCLUSIONS

5.1. Discussion

Segmentation of BoT and larynx cancer regions is particularly difficult due to the presence of MRI artefacts, enhancements of other non-cancer regions (blood vessels, salivary glands), geometric variability and weak edges of the cancer regions across the patients. An unsupervised segmentation framework (PLCSF) was presented in this paper for this task that does not require any manual intervention or training data. This framework makes no assumption about the shape or size of the cancer regions, thus can successfully segment the cancer regions with geometric variability. Also, the cases used in this study are representative of everyday clinical challenges.

In this framework, a novel adaptive determination of parameter spline distance (knot spacing) allowed the estimation of complex bias field (IIH) present in MRI slices used in this work. Detection of the throat region using fuzzy rule based technique allowed the knowledge of the approximate

cancerous position to be embedded in the system, particularly in MFCM, thus reducing further processing steps to eliminate healthy tissues from cancer detected clusters that are away from the throat region. Comparison of MFCM with the standard FCM [19] in [16] showed that MFCM achieved better results compared to the standard FCM. The continuity and spatial smoothness of the cancer boundary was ensured by evolving the level set surface on the detected cancer region.

Quantitative comparison with the Gold Standard (consensus manual outline) on 102 T1+Gd MRI axial slices from 10 patients, the system (PLCSF) shows no significant difference in performance (PCC: 0.89, $p<0.05$) with the method used in current clinical practice. The PLCSF result also demonstrated improved performance when compared to other algorithms (MS clustering and Ncut). Existing semi-automatic approach [6] for tongue cancer segmentation validated on 16 patients (78 axial slices) demonstrated mean correspondence ratio of 0.83 which is comparable to PLCSF DSC of 0.79. However, the semi-automatic approach in [6] required manual-placing of seed points in the tongue tumour region or drawing of close loop outside the tumour from expert and have no results to prove any validation on laryngeal cancer.

The limitation of the current framework (PLCSF) is over-segmentation of cancer region in case of similar characteristics of cancer tissues as compared to surrounding tissues.

One of the main purposes of the automatic cancer region segmentation of T1+Gd MRI slices is the reproducibility of the segmentation results that contain intra- and inter- variability from manual segmentation results. For this framework, if the parameters values are unchanged, the system obtains similar results for repeated number of times, indicating the reproducibility of the system. Further, using single modality (T1+Gd) in RTP can reduce scanning and processing time of MRI slices and increase the computational efficiency. Thus, this tool can assist RO in RTP to obtain pharynx and larynx cancer boundaries from T1+Gd MRI axial slices in time-effective and unbiased manner. To the best of our knowledge this is first automatic tool focused on segmentation of BoT and larynx

cancer from T1+Gd MRI. The system also demonstrated that it can perform robustly against variations caused by different MRI scanners protocols with different manufacturer and scanner models.

5.2. Conclusions

This new and automatic pharynx and larynx cancer segmentation framework (PLCSF) shows initial acceptable statistically-validated results in BoT and larynx cancer segmentation from axial T1+Gd MRI slices. This framework may help reduce inter- and intra- variability and can assist radiation oncologists with time-consuming, complex radiotherapy planning.

Future work will focus on testing the system on large cohort of BoT and larynx cancer patients. Further, when labelling the MRI slice, during manual outline, knowledge from the previous and next slices is used. The current system does not make use of this information. Future work will make use of this information in 3D segmentation and visualization of the cancer.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Beatson Cancer Charity and the University of Strathclyde for their financial support with this study. The authors would like to acknowledge the work of the two Radiation Oncologists and members of head-and-neck radiotherapy weekly team meeting for manually segmenting the MRI data.

REFERENCES

[1] E. Ward and C. As-Brooks, *Head and Neck Cancer Treatment*, 2nd revised edition. San Diego, CA: Plural Publishing Inc, 2014.

[2] S. Yeh, "Radiotherapy for Head and Neck Cancer", *Semin Plast Surg*. vol. 24, no. 2, pp. 127–136, May 2010.

[3] E. Paulson, B. Erickson, C. Schultz and X. Allen, "Comprehensive MRI simulation methodology using a dedicated MRI scanner in radiation oncology for external beam radiation treatment planning." *Med Phys.* vol. 42, no. 1, pp.28-39, 2015.

[4] A. Ashraf, S. Gavenonis, D. Daye, C. Mies, M. Rosen, and D. Kontos, "A Multichannel Markov Random Field Framework for Tumor Segmentation with an Application to Classification of Gene Expression-Based Breast Cancer Recurrence Risk", IEEE Trans. on Medical Imaging, vol. 32, no. 4, pp. 637-648, April 2013.

[5] V. G. Kanas, E. I. Zacharaki, C. Davatzikos, K. N. Sgarbas, and V. Megalooikonomou, "A low cost approach for brain tumor segmentation based on intensity modeling and 3D Random Walker", Biomedical Signal Processing and Control, vol. 22, pp. 19–30, Sep. 2015.

[6] V. Chong, J. Zhou, J. Khoo, J Huang and T. Lim. "Tongue carcinoma: tumor volume measurement.", *Int J Radiat Oncol Biol Phys*, vol. 59, no. 1, pp. 59-66, 2004

[7] Y. Cui, Y. Tan, B. Zhao et al., "Malignant lesion segmentation in contrast-enhanced breast MR images based on the marker-controlled watershed.", *Med Phys*., vol. 36, no. 10, pp. 4359-4369, 2009.

[8] V. Harati, R. Khayati and A. Farzan, "Fully automated tumor segmentation based on improved fuzzy connectedness algorithm in brain MR images", *Computers in Biology and Medicine*, vol. 41, pp. 483–492, 2011

[9] O. Salvado, C. Hillenbrand, S. Zhang, and D. Wilson, "Method to Correct Intensity Inhomogeneity in MR Images for Atherosclerosis Characterization", *IEEE Trans. on Medical Imaging*, vol. 25, no. 5, pp. 539-552, May 2006

[10] S. Lankton and A. Tannenbaum, "Localizing region-based active contours," IEEE Transactions on Image Processing, vol. 17, no. 11, pp. 2029–2039, 2008

[11] T. Wang, I. Cheng, and A. Basu, "Fluid Vector Flow and Applications in Brain Tumor Segmentation", *IEEE Trans. Biomedical Engineering*, vol. 56, no. 3, pp. 781-789, March 2009

[12] S. Ho, E. Bullitt and G. Gerig, “Level set evolution with region competition: automatic 3-D segmentation of brain tumors”, *In Proc. of International Conference Pattern Recognition*, Quebec, Canada, Vol. I, pp.532–535, 2002

[13] T. Doshi, J. Soraghan, L. Petropoulakis, D. Grose and K. MacKenzie “Semi-Automatic Segmentation of Tongue Tumors from Magnetic Resonance Imaging”, *in Proc. of IEEE IWSSIP*, Bucharest, Romania, pp. 143-146, July 2013

[14] M. Ahmed, S. Yamany, N. Mohamed, A. Farag, and T. Moriarty, “A Modified Fuzzy C-Means Algorithm for Bias Field Estimation and Segmentation of MRI Data”, IEEE Trans. on Medical Imaging, vol. 21, no. 3, pp. 539-552, March 2002

[15] M. El-Melegy1 and H. Mokhtar, “Tumor segmentation in brain MRI using a fuzzy approach with class center priors”, *EURASIP Journal on Image and Video Processing*, vol. 21, pp. 1-14, 2014

[16] T. Doshi, J. Soraghan, L. Petropoulakis, D. Grose and K MacKenzie, “Modified Fuzzy C-means Clustering for Extraction of Tongue Base Tumor from MRI data”, *In Proc. of IEEE EUSIPCO*, Lisbon, Portugal, pp. 2460-2464, Sept 2014

[17] D. McClymont, A. Mehnert, A. Trakic, D. Kennedy and S. Crozier, “Fully automatic lesion segmentation in breast MRI using mean-shift and graph-cuts on a region adjacency graph”, *Magnetic Resonance Imaging*, vol. 39, no. 4, pp. 795–804, April 2014.

[18] J. Zhao, L. Zhang, W. Zheng, H. Tian, D.-M. Hao, and S.-H. Wu, J. He, X. Liu, E. Krupinski, and G. Xu, Eds., “Normalized cut segmentation of thyroid tumor image based on fractional derivatives,” *in Health Information Science. New York: Springer*, vol. 7231, LNCS, pp. 100–109, 2012

[19] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981

[20] D Comaniciu and P Meer, “Mean shift: A robust approach toward feature space analysis.” *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 24, no. 5, pp. 603–619, May 2002

[21] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 22, no. 8, pp. 888–905, Aug. 2000

[22] J. Park, S. Jung, Y. Joo, C. Jung, K. Cho and M. Kim, “Diagnostic accuracy of magnetic resonance imaging (MRI) in the assessment of tumor invasion depth in oral/oropharyngeal cancer”, *Oral Oncology*, vol. 47, pp. 381–386, 2011

[23] F. Sakai, S. Sone, K. Kiyono et al., “MR Evaluation of Laryngohypopharyngeal Cancer: Value of

Gadopentetate Dimeglumine Enhancement", *Am J Neuroradiol*. vol. 14, no. 5, pp.1059-1069,1993

[24] N. Tustison, B. Avants, P. Cook et al., "N4ITK: improved N3 bias correction." *IEEE Trans Med Imaging*. vol. 29, no.6, pp. 1310-20, 2010

[25] S. Smith and J. Brady, "SUSAN—A new approach to low level image processing", *International Journal Computer Vision*, vol. 23, no. 1, pp. 45–78, May 1997

[26] R. Gonzalez and R. Woods, *Digital image processing*, Prentice-Hall, 2000

[27] T. Tan, K. Sim and C. Tso, "Image enhancement using background brightness preserving histogram equalisation*", IEEE Electronics Letters,* vol. 48, no. 3, Feb 2012

[28] J. Mason, K. Hlavackova and K. Warwick, "*Approximation Using Cubic B-Splines with Improved Training Speed and Accuracy*", in Curse of Dimensionality, Computer-Intensive Methods in Control and Signal Processing. Birkh*f*auser Computer Science, Boston, pp. 295-303, 1997

[29] R. Haralick and L. Shapiro, Computer and Robot Vision, Volume I. Addison-Wesle, 1992

[30] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-9, no. 1, Jan 1979

[31] K. Chuang, H. Hzeng, S. Chen, J. Wu and T. Chen "Fuzzy c-means clustering with spatial information for image segmentation", *Computerized Medical Imaging and Graphics*, vol. 30, pp. 9–15, 2006

[32] C. Clark, E. Miles, M Urbano 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*. vol. 82, pp. 585-594, 2009

[33] L. Dice, "Measures of the amount of ecological association between species" *Ecology*, vol. 26, no. 3, pp. 297-302, July 1945

[34] M. Dubuisson and A. Jain, "A Modified Hausdorff Distance for Object Matching", *in Proc. of 12th IEEE IAPR Conference*, Jerusalem, Israel, pp. 566 - 568 vol.1, 1994

[35] Shai Bagon, Matlab interface for EDISON [Online].
Available: http://www.wisdom.weizmann.ac.il/~bagon/matlab.html

[36] T. Cour and K. Shi, Normalized Cuts Matlab Code [Online].
Available: http://www.cis.upenn.edu/~jshi/software/