Title: Supervised Learning based Multimodal MRI Brain Tumour Segmentation using Texture Features from Supervoxels

Authors: Mohammadreza Soltaninejad [a], Guang Yang [b,c], Tryphon Lambrou [a], Nigel Allinson [a], Timothy L Jones [d], Thomas R Barrick [e], Franklyn A Howe [e], Xujiong Ye [a]

Affiliations:

[a] School of Computer Science, University of Lincoln, Lincoln, LN6 7TS, UK (emails: {msoltaninejad | tlambrou | nallinson | xye}@lincoln.ac.uk)

[b] National Heart and Lung Institute, Imperial College London, London, SW7 2AZ, UK (g.yang@imperial.ac.uk)

[c] Neurosciences Research Centre, Molecular and Clinical Sciences Institute, St. George's, University of London, London, SW17 0RE, U.K.

[d] Academic Neurosurgery Unit, St. George's, University of London, London, SW17 0RE, UK (timothy.jones@stgeorges.nhs.uk);

[e] Neurosciences Research Centre, Molecular and Clinical Sciences Institute, St. George's, University of London, London, SW17 0RE, U.K. (emails: {tbarrick | howefa} @sgul.ac.uk).

Corresponding Author: Mohammadreza Soltaninejad

PhD Student, School of Computer Science, University of Lincoln, Lincoln, LN6 7TS, UK

Email: msoltaninejad@lincoln.ac.uk , Mobile Phone: +447741059327 , Fax: +44 (0) 1522 886974

Funding: This research was supported by European FP7 collaborative Project "MyHealthAvatar" (600929). MRI data were obtained during the EU FP7 "eTUMOUR" project (LSHC-CT-2004-503094).

Abstract

Background: Accurate segmentation of brain tumour in magnetic resonance (MR) images is a difficult task due to various tumour types. Using information and features from multimodal MRI including structural MRI and isotropic (p) and anisotropic (q) components derived from the diffusion tensor imaging (DTI) may result in a more accurate analysis of the images.

Methods: We propose a novel 3D supervoxel based learning method for segmentation of tumour in multimodal MRI brain images. Supervoxels are generated using the information across the multimodal MRI data set. For each supervoxel, a variety of features including Gabor texton and statistical features are extracted. This is then followed by a random forests (RF) classifier to classify each supervoxel into tumour core, oedema or healthy brain tissue.

Results: The proposed method is evaluated on two datasets: 1) Our clinical dataset: 11 multimodal images of patients and 2) BRATS 2013 clinical dataset: 30 multimodal images. For our clinical dataset, the average detection sensitivity of tumour (including tumour core and oedema) using multimodal MRI is 86% with balanced error rate (BER) 7%; while the Dice score for automatic tumour segmentation against ground truth is 0.84. The corresponding results of the BRATS 2013 dataset are 96%, 2% and 0.89, respectively.

Conclusion: The proposed method provides promising results in the segmentation of brain tumour. Adding features from multimodal MRI images increases the segmentation accuracy.

Keywords: Brain tumour segmentation, Diffusion tensor imaging, Multimodal MRI, Random forests, Supervoxel, Textons

Highlights:

- Supervoxel segmentation using multimodal MRI to produce boundaries across multiple image protocols.
- Unified framework to classify each supervoxel using features calculated from multimodal MRI.
- Improved performance for classification of brain tumour supervoxels by using texton descriptors.
- Applying DTI with conventional MRI increases the segmentation accuracy for tumour structures.

1. Introduction

Brain tumours can arise from abnormal growth of the cells inside the brain or can develop from cells that have spread to the brain from a cancer elsewhere. There are a wide variety of brain tumour types that are classified according to their cell of origin, and can be categorised as low or high grade depending on their malignancy and growth characteristics.

Diagnosis of tumour grade and type is essential for optimum treatment. Medical imaging modalities are used for detection and assessment of tumours. Among these medical imaging methods, magnetic resonance imaging (MRI) is the most widely used for clinical diagnosis, treatment selection, prognosis and to aid surgery and radiotherapy planning [1]. Due to the multimodal nature of MRI there are a range of image types and contrasts that enable a subtle radiological assessment of tumour type.

Computer-aided procedures are being developed to aid conventional neuroradiological diagnosis and treatment planning. Image processing with pattern recognition and machine learning algorithms are widely used for analysis as an aid to interpretation of medical images. Segmentation techniques have been proposed for several clinical applications [2]. For brain tumours, image segmentation may aid the fast and objective measurement of tumour volume and also find patient-specific features that aid diagnosis and treatment planning [3].

A primary segmentation task in the case of brain tumours is to accurately label the tumour tissue and the normal brain regions. In many cases, the tumour region is visually distinct, but it is a challenge for accurate and reproducible, segmentation and characterisation of the abnormality that works across multiple tumour types and with different MR scanner types [3]. Even within one pathological class of tumour there is a large variety and complexity of tumour imaging characteristics such as signal intensity, image texture, and its size, shape, location with respect to other normal brain structures. Some tumours with high grades are quite heterogeneous having a necrotic core surrounded by viable tumour that infiltrates into the normal brain tissue. Adjacent non-tumour regions may also look abnormal due to an inflammatory response creating areas of oedema. Hence

it is a difficult task to develop a universal method to segment tumours accurately [4]. Clinical needs for tumour segmentation include dose-planning for radiotherapy, for assessing changes in tumour volume when monitoring low to high grade transformation of glial tumours, and to monitor the response to treatment.

Manual segmentation of tumours in MRI images is time-consuming and subjective since it is dependent on the operators' skill and experience, hence inter-operator reproducibility can be low. Automatic computer assisted procedures have the potential to provide more objective segmentation of tumours, and also allow large-scale multimodal MRI data to be analysed within a reasonable processing time. Nevertheless, manual segmentation by experts is commonly used as a gold standard for assessing the automatic or computer-aided segmentation techniques and also for training the systems.

1.1. Related Works

The research work for automatic brain tumour segmentation has increased in recent decades which represents the demand for this area of research and currently it is still in progress [5]. Several methods have been proposed in the literature for detection and segmentation of tumours in MR images [6]. The segmentation methods can be categorized into unsupervised and supervised learning based methods [3].

Unsupervised segmentation techniques use clustering methods for segmenting unlabelled images. Expectation maximization (EM) is one of the popular unsupervised methods which was utilized in [7] with the application on multimodal conventional MRI (C-MRI) data in which the prior-knowledge of the normal brain was obtained from atlas of normal brain and

the intensity model for tumour was estimated. Another popular unsupervised clustering method is fuzzy *c-Means* (FCM) and in [8] an improved approach was proposed for brain tumour segmentation which included the information from class centres to regularize the clusters. A nonparametric model-based method was proposed in [9]. The method was based on graph-cut distribution without involving the training procedure and has low computation time. A comparison of most recent unsupervised methods for brain tumour segmentation was presented in [10]. They also introduced an unsupervised method for segmentation of high grade gliomas (HGG). Their method was applied to multiparametric MRI data which combines other modalities than conventional T2-weighted and contrast enhanced MRI and included diffusion-weighted imaging (DWI). DWI will also be considered in our paper, but using parameters derived from diffusion tensor imaging (DTI). The advantage of unsupervised methods is that they do not require a large amount of training data. However, the methods are not able to automatically label segmentation results to different tissue types (e.g. tumour core, oedema, necrosis, or healthy brain tissue). Those tissue types are determined by users, making the methods inherently semi-automated. Furthermore, using unsupervised segmentation for brain tumours is challenging due to the lack of shape or intensity prior [3].

Supervised learning based algorithms use training data for segmentation of tumours, which are labelled by experts. Helen *et al.* developed a hybrid method for brain tumour segmentation based on clustering, classification and conventional segmentation methods [11]. Several works applied random forests (RF) classification and its variants to segment tumours [12–15]. In [12] several features including intensity, geometry and asymmetry from multiple protocols are applied to a random forests classifier. Extremely randomized

trees were used in [13] with high level features including appearance and context-based features calculated from nonlinear transformation of the images. The work in [14] used Gaussian mixture models for different individual protocols (i.e. T1-weighted, T2-weighted and FLAIR) separately. Goetz *et al.* [15] proposed a new random forest based method which uses domain adaptation to reduce sample selection errors.

Few studies have combined different MRI modalities for brain tumour segmentation. A number of advanced algorithms [16–20] were recently presented in [5] using the Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) dataset [21,22] organized in conjunction with the international conference on Medical Image Computing and Computer Assisted Interventions (MICCAI) 2012 and 2013 conferences. The methods were based on segmentation of different tumour tissues, i.e. tumour core, oedema, necrosis, using multimodal conventional MRI containing FLAIR, T1-weighted , T1-contrast and T2 protocols [23]. In [24] C-MRI and DTI are combined and fed to support vector machines (SVM) to segment different tumour tissue types. In another work [20], which used decision forests to segment HGG, the segmentation results by adding DTI were improved compared to using only C-MRI modalities. Combination of DWI with C-MRI has been also researched for unsupervised methods such as spatial fuzzy *c-Means* [25] to improve the segmentation results. The BRATS dataset [21,22] included C-MRI modalities, whereas our own clinical dataset contains both C-MRI and DTI modalities. In this paper, we have combined multimodal MRI for accurate segmentation and labelling of different tumour parts (e.g. tumour core and oedema).

Most previous studies are voxel-wise, in which a window or subarea around a voxel is normally used to extract features for labelling (classifying) the voxel. In the case of

multimodal MRI data, it is comprised of millions of voxels (i.e. the sum of all voxels across each image modality) and consequently voxel based methods usually require significant computational time. Few studies have used superpixel or supervoxel methods for segmentation. Wu *et al.* used supervoxel based features in a conditional random fields (CRF) framework to detect brain tumours [26]. In [17] Markov random fields are applied on supervoxels of the images to segment the tumours based on intensity probabilities.

In this paper, we aim to segment brain tumour parts (core and oedema) using a novel multimodal MRI supervoxel based method (combining DTI with conventional MRI modalities). Gabor texton based features; alongside first order intensity based statistical features are calculated for each supervoxel and used in a random forest classifier to label supervoxels into different tissue types.

1.2. Our Contribution

Most of the existing studies on brain tumour segmentation are performed on conventional MRI protocols (i.e. FLAIR, T1-weighted (with contrast) and T2-weighted), which are based on qualitative image intensities. In this study, in addition to the conventional MRI sequences, we also consider the isotropic (p) and anisotropic (q) diffusion components derived from DTI [27], which provides parameters that relate to the average microscopic movement of water within tissue structure (p) and whether this movement has an anisotropic element of diffusion (q) such as for the water in white matter fibers. We hypothesize that combining DTI and C-MRI may provide quantitative features that increase the classification accuracy and improve tumour segmentation results.

Instead of applying voxel based techniques commonly used in classification-based segmentation of brain tumour in MR images, in this paper, a supervoxel based method is considered, which partitions an image into a number of small 3D patch volumes. The advantage of the supervoxel based method is that the required computation for classification in the new feature space can be significantly reduced. Feature vector size in the general case of supervoxels is less than those that are based on image voxels (i.e. moving window).

The main contributions of our method can be summarised as follows:

- The supervoxel is formed using multimodal MRI, including FLAIR, T1-weighted (with contrast), T2-weighted, p and q diffusion maps. Unlike existing methods [28] in which a supervoxel is calculated from one single MRI protocol, in this paper, information from multimodal images is combined to produce supervoxel boundaries across multiple image protocols.
- A unified framework is built to classify each supervoxel using features calculated from multimodal MRI for segmentation of each brain tumour.
- We have shown that our novel histogram of texton descriptors, calculated using a set of Gabor filters with different sizes and orientations provide improved performance for classification of brain tumour supervoxels. Since supervoxels are limited to clusters of similar intensities within each MRI modality, using the distribution of local textures inside each supervoxel improves further classification of supervoxels, Texton has demonstrated its advantages of providing significant information to distinguish various patterns.

The paper is organised as follows. Section II describes the proposed method, which consists of supervoxel segmentation, feature extraction, classification, and final segmentation. Section III presents experimental results and is followed by the discussion and discussion in Section IV and Section V.

2. Materials and Methods

2.1.Data Acquisition

Brain tumour patient data was acquired using a GE Signa Horizon LX 1.5T MRI system (GE Healthcare, Milwaukee, WI, USA) equipped with a maximum field gradient strength of 22mT/m and using a quadrature head coil. The multimodal MRI acquisition used in this study is described below.

FLAIR and T1-weighted images were acquired in the axial plane with a field of view (FOV) 240 x 240 mm^2, matrix size 256 x 256 and 5 mm slice thickness for FLAIR and 2.8 mm for T1 with no slice gap. The following acquisition parameters were used for FLAIR (TE = 133 ms, TR = 9000 ms, inversion time 2200 ms, band width = 61.04 Hz) and T1 weighted (TE = 14 ms, TR = 600 ms, band width = 122.1 Hz). T1-weighted images were acquired both with and without intravenously administered contrast agent (0.1 mmol/kg gadoterate meglumine, Dotarem).

T2-weighted images were acquired in the axial plane using a dual echo sequence with TR = 3500 ms and TE=14/98 ms and FOV of either 220 x 220 mm^2 or 240 x 240 mm^2, a 256 x 256 acquisition matrix, and 29 slices with 5 mm thickness [29].

DTI data were acquired using a diffusion-weighted spin-echo echo-planar imaging

sequence. A b_0 acquisition was made without diffusion gradients (b=0 s/mm^2) and diffusion weighted images were acquired using b=1000 s/mm^2 with 12 gradient directions [30]. The FOV was 240 x 240 mm^2 with a 96 x 96 acquisition matrix. In total 50 contiguous slices (2.5 mm in-plane resolution) were acquired with a slice thickness of 2.8 mm. TR and TE were 8 secs and 88 ms, respectively. The data was interpolated to a 256 x 256 matrix. The diffusion parameters p and q for isotropic and anisotropic diffusion respectively were calculated as proposed by Peña *et al.* [27].

A cohort consisting of 11 brain tumour patients (2 grade III, and 9 grade IV) retrospectively entered the study and were scanned using the multimodal MRI protocol. Histological diagnosis was available for all tumours. Patient ages at the time of scanning ranged from 33 to 73 years (mean age 53 and standard deviation 7). The ground truths are provided by a trained human expert.

2.2.Overview of the Method

Our method is comprised of four steps (preprocessing, supervoxel partitioning, feature extraction and classification) that are depicted in Fig. 1 and described below.

After image preprocessing, the supervoxel segmentation partitions the MRI data into equally sized patches with similar intensity ranges. Supervoxels are calculated based on a distance matrix which is formed using a combination of multimodal images. Use of different MRI modalities can enhance the supervoxel segmentation by identifying image boundaries simultaneously across all available images. For each supervoxel patch, a number of features including statistical and texture features are calculated. The supervoxels

are classified into tumour and nontumour using a random forests classifier. Tumour supervoxels are then grouped together to obtain tumour boundaries.

2.3.Preprocessing

DTI data were realigned to remove eddy current distortions using eddy correct (FSL Software Library by FMRIB [31]) prior to generating p and q maps. Images were skull stripped using Brain Extraction Tool in FSL. All conventional MRI data were then co-registered to the DTI b_0 data using an affine transformation with a mutual information based cost function using Statistical Parametric Mapping (SPM12 [32]) to avoid interpolation of quantitative diffusion characteristics.

The image intensities are normalised with a two-step procedure: histogram matching and dynamic range normalisation. First, one case (one patient data) is selected as reference and the histogram of each image protocol of other cases are matched to the corresponding protocol of the reference case. To eliminate the bias of the matched histogram to the reference case, another stage is added to the process according to [33]. The average of all the new histograms including the initial reference case is calculated for each protocol and the histograms are again matched to the new reference, e.g. the average histogram for each protocol. In the second stage, for each case, the intensity of new images of all the protocols obtained from the first step are linearly normalized to the dynamic range of the corresponding FLAIR related to that case. This is to ensure that, in the feature extraction stage, for each case, images from different protocols have similar intensity dynamic ranges.

2.4.Supervoxel Segmentation

The aim of supervoxel clustering is to group an image into a predefined number of portions, which have similar intensity range. In this paper, the simple linear iterative clustering (SLIC) superpixel method [34] is extended to extract 3D supervoxels for the segmentation of brain tumour. A brief description of SLIC is given below.

In our method, the initial grid height is chosen based on the slice thickness (spatial resolution in Z direction) of the MRI images and the spatial resolution ratio (R_s) between X and Y directions. For our own dataset, the resolutions in X and Y directions are the same (so R_s=1). Since all the data are co-registered in the preprocessing stage, the slice thickness for each dataset is consistent through all the slices in each image data set which is considered as R_t. It should be noted that registration of the data is very important to perform this multimodal supervoxel segmentation. If the supervoxel width is considered to be W_S voxels, its height, H_S is calculated from the ratio of slice spatial resolution to slice thickness

$$H_S = \left\| W_S \times \frac{R_S}{R_t} \right\|. \quad (1)$$

The operator ||…|| means the nearest rounding integer to the value. The minimum value for supervoxel height, H_S, is considered to be 3, whilst, H_S=1 results in 2D segments which are considered as superpixels.

In the first instance, the geometrical centres of the initial grids are considered as supervoxel region centres. The mean value of the voxel coordinates inside the supervoxel provides the centre of gravity of that supervoxel. The locations of the centres of gravity are updated during each iteration. The distance between each voxel in the dataset to the bounded cluster centres are calculated and then a label of the closest cluster centre is

assigned to that target voxel. The final distance is comprised of both intensity and location distances. The intensity distance, d_c, is calculated by defining the intensity difference between the *i*th and the *j*th voxel according to the following formula:

$$d_c = \sqrt{(I_j - I_i)^2}, \tag{2}$$

where, I_i and I_j are the normalized intensity values of the *i*th and the *j*th voxel, respectively. The location distance, d_s, between the two voxels is calculated as follows,

$$d_s = \sqrt{(R_x(x_j - x_i))^2 + (R_y(y_j - y_i))^2 + (R_z(z_j - z_i))^2}, \tag{3}$$

where, (x_i, y_i, z_i) is the coordinate of voxel *I* and R_x, R_y and R_z are the voxel resolutions.

The distance measure [34] is then defined as,

$$D = \sqrt{d_c^2 + \left(\frac{d_s}{w_s}\right)^2 m^2}, \tag{4}$$

where, *m*, is the compactness coefficient. A higher value of *m* results in more compact segments and a lower value creates more flexible boundaries.

Fig. 2 shows the supervoxel segmentation of a brain tumour using MRI FLAIR with two different initial grid sizes.

Supervoxel segmentation of multimodal MRI data is not straightforward as tissue boundaries apparent on one MRI modality, for example, on T1-weighted (with contrast) are not necessarily apparent on other MRI modalities such as DTI or FLAIR, and vice versa. Hence supervoxel boundaries determined independently for each MRI modality will not match, creating tissue partial volume effects at supervoxel boundaries. To solve this

problem, we adapt the supervoxel intensity distance equation (2) in a multidimensional formation and apply this across all MRI modalities, to determine a multimodal supervoxel cluster. Assuming that the multimodal MRI data is acquired with MRI protocols P_1, P_2, …, P_N, giving the images $\{I_{P1}, I_{P2}, \ldots, I_{PN}\}$ then the distance equation for multimodal MRI data is,

$$d_c = \sqrt{\left(I_{Voxel,P_1} - I_{Center,P_1}\right)^2 + \cdots + \left(I_{Voxel,P_N} - I_{Center,P_N}\right)^2}, \quad (5)$$

where, $I_{Voxel,Pi}$ is the grey-level intensity corresponding to the *voxel* in protocol P_i.

A framework for the multimodal supervoxel segmentation method is shown in Fig. 3. Combining all MRI modalities helps supervoxel segmentation by enhancing weak image boundaries that appear in any single modality. For example, weak edges may appear in one image but present strong in the remaining images. An example of this case is shown in Fig. 4. The calculated supervoxel map using the multimodal segmentation method is overlaid on both FLAIR (top row of the first column in Fig. 4) and p map (bottom row of the first column in Fig. 4). The middle and the last columns in Fig. 4 show two corresponding zoomed-in areas indicated in the FLAIR and p map images (yellow and orange rectangles). It is noted that, the middle column of the Fig. 4 shows strong edges in FLAIR image (shown by red ellipses), whereas corresponding edges in the *p* map are quite weak (shown by the blue ellipse). The opposite effect is apparent in the right column of Fig. 4. By using the multimodal clustering method, the extracted supervoxel map provides good image boundaries even when boundaries are not clear in one image modality.

Fig. **5** shows a comparison of supervoxel segmentation of tumour core calculated from a single MRI modality (FLAIR) and from multimodal MRI (FLAIR, T1-weighted (with

contrast), T2-weighted, *p* and *q* maps). As it can be seen in Fig. 5, there are misalignments between supervoxels boundaries (computed from FLAIR) and the ground truth boundaries (see black ellipse in Fig. 5(f)), whilst multimodal supervoxels show improvement in boundary alignment to the tumour core (see black ellipse in Fig. 5(i)).

2.5.Feature Extraction

Grouping the supervoxels for final segmentation of the tumour is based on the feature sets that are extracted from each supervoxel. In this section, first order statistical features and texton features are considered.

2.5.1. First Order Statistical Features

First order intensity statistics [35] are also referred to as voxel-intensity based features. First order statistical features express the distribution of grey levels within selected regions of interest (ROI), represented by supervoxels in our case. We use 16 features including the average, standard deviation, variance, mean of the absolute deviation, median absolute deviation, coefficient of variance, skewness, kurtosis, maximum, minimum, median and mode of the intensity values, central moments, range, interquartile range and entropy.

2.5.2. *Texton Features*

Due to the complexity and heterogeneity of tumour tissue, first order intensity features are generally not sufficient for an accurate segmentation. In this paper, texton features are considered to improve segmentation. Textons are small image elements that can be

generated by convolution of the image with a set of image filters. We use the Gabor filter [31] defined in

$$G(x, y; \theta, , \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\,\sigma^2}\right)\exp\left(i\left(2\pi\frac{x'}{\psi}+\right)\right), \tag{6}$$

where

$$\begin{aligned} x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta. \end{aligned} \tag{7}$$

The Gabor filter parameters were chosen empirically. Six different filter directions were considered: [0^o, 30^o, 45^o, 60^o, 90^o, 120^o] with filter sizes from 0.3 to 1.5 at steps of 0.3. The wavelength of sinusoid coefficients of the Gabor filters were 0.8, 1.0, 1.2 and 1.5. This provided a filter bank of 120 filters.

Filter response images are the result of convolution of each filter with an MR image. For filters with the same size but different directions, the maximum response is considered, leading to a total of 20 filter responses (5 sizes, 4 wavelength coefficients). The texton map is then generated by applying 20-dimensional *k-Means* clustering to the 20 filter responses with a predefined number of clusters of $k_t = 5$ to represent tumour core, oedema and normal brain tissues. To reduce computation time for clustering, the lowest number of clusters which are capable of separating tumour core and oedema from normal brain in the training set was chosen. Histograms of the texton parameter were then calculated for each supervoxel using the generated texton map. The distribution of the local textures (descriptor) used to characterize the local object patterns, is one of the main features used in

our tumour classification.

Table 1 summarises all extracted features. In total, there are 21 features for each MR image, so there are 105 features across the multimodal MRI data (FLAIR, T1-weighted (with contrast), T2-weighted, p and q maps). All feature calculations are performed on supervoxels and the extracted features for each MR image are concatenated to form the final multimodal feature vector.

2.6. Random Forests Classification

Random forests (RF) is one of the best among classification algorithms [36]. It is an ensemble learning method that uses multiple decision trees. During the bagging process and at each attribute split, a random subset of features is used. After generating a large number of trees, a vote for the most popular class is made [37]. The structures of randomized trees are independent of training sample outputs.

In this study, all supervoxels within the brain are considered for classification. This not only represents a large amount of data, but this data is also unbalanced, as the number of supervoxels related to normal brain is in the range of 6 to 30 times more than the number of tumour supervoxels (average ratio of 12:1). Therefore, the use of a robust classifier is essential to achieve accurate segmentation. Due to the many advantages of the RF classifier, (e.g. accuracy, efficiency in application to large datasets, and ability to handle unbalanced datasets) we use RF to classify each supervoxel into three tissue classes: normal brain tissue, tumour core and oedema.

The main parameters used in RF, i.e. the number of trees, the number of attributes, and

tree depth, are chosen as follows: number of trees is 50 with depth of 15, and number of attributes (k_a) selected to perform the random splits for a specific number of features N_f is $k_a = \sqrt{N_f}$. For single modality and multimodal experiments, 5 and 10 attributes are selected, respectively. Further discussions are given in the Experimental Results Section (Section III-A)

In the training stage, the supervoxels are split into three classes: normal brain tissue, tumour core and oedema. Supervoxels which have at least 50% overlap with tumour core or oedema regions (ground truth according to manual labelling) are labelled as the appropriate corresponding classes. The remaining supervoxels are labelled as normal. The RF classifier is trained based on these three labels. In the testing stage, the trained classifier is applied and labels are assigned to each supervoxel inside the brain. The tumour area is then obtained by grouping the supervoxels classified as either tumour core or the oedema class.

3. Experimental Results

Two datasets were analysed: (i) our clinical dataset described in Section II-A for training and validation of the algorithm, and (ii) the publicly available MICCAI BRATS 2013 dataset [21,22] for further comparison and assessment of the robustness of the method. For both datasets, quantitative evaluations (e.g. supervoxel classification accuracy and Dice score overlap measures for segmented tumour vs ground truth) of the proposed method have been conducted using different imaging protocols (e.g. single modality or multimodal images). The leave-one-out approach is used to train and test the model. Subsections (A-C) are focused on our clinical data cohort; whist subsection D evaluates results of our technique to the MICCAI BRATS 2013 dataset [5,21,22]. Subsection E presents statistical

analysis on the two datasets.

3.1. Parameter Selection

For 2D superpixel calculation presented in our previous work [38], an optimal initial superpixel size of 5 was obtained. In the case of 3D supervoxels, the z direction is determined based on Equation (1) from the slice thickness and image resolutions. Due to the different resolutions used in our clinical data (all multimodal MRI data were co-registered to DTI with voxel dimensions 0.9375 mm × 0.9375 mm × 2.8 mm) and the BRATS dataset (isotropic voxel dimensions: 1 mm^3), the supervoxel initial sizes were chosen to be 8 × 8 × 3 for our clinical data, and 5 × 5 × 5 for the BRATS data. By visually inspecting the supervoxel boundaries and area, the value of m = 0.05 (in Equation (4)) was chosen, which presents coherent boundaries.

Implementation of the RF was performed in MATLAB 2016b based on the open source code provided in [39]. To select the optimum RF parameters, different ranges of number of trees and depth were assessed on our clinical data. 4-fold validation was used to select the optimal RF parameters (i.e. number of trees and depth). Classification accuracy was calculated for the testing fold in each iteration with different tree depth and number of trees. Values were averaged over all folds to determine the effects of number of trees and depth, and are presented in Fig. 6. It can be seen in Fig. 6 that, 50 trees with depth 15 give an optimum generalization and accuracy. These optimal parameters were also directly used in the analysis of the BRATS dataset (in Section D).

Table 2 presents the proportion of features selected from each acquisition protocol using the RF from two experiments: conventional MRI data only (C-MRI, namely, FLAIR,

T1-weighted (with contrast) and T2-weighted) and conventional MRI plus DTI (C-MRI+DTI). It can be seen that, for C-MRI, most of the features (61%) are selected from the FLAIR, which shows the importance of FLAIR for tumour segmentation. When DTI is added it has 24% of features selected from it (i.e. p (16%) and q (8%) maps); the presence of DTI also slightly reduces the proportion of corresponding features from the C-MRI modalities alone. Our experimental results in the next section show that p and q maps improve the overall segmentation of tumour core.

3.2. Supervoxel Classification Results

For the standard four classification measures (accuracy, precision, sensitivity, specificity), both accuracy and specificity will give very high values due to the highly imbalanced nature of our data. Therefore, to properly evaluate the classification performance, only precision and sensitivity are considered. Consequently, in this paper, evaluation of the performance of the supervoxel classification method was performed using precision, sensitivity and balanced error rate (BER).

To compare the supervoxel classification performances of our method using different MRI modalities for the whole tumour including core and oedema, three experiments are performed: 1) FLAIR only; 2) C-MRI data; 3) C-MRI+DTI.

In the first experiment, supervoxels are calculated based on FLAIR image only; whereas in the second and third experiments, supervoxels are calculated using Equation (4) based on different MRI modalities, i.e. C-MRI data in experiment 2 and C-MRI+DTI in experiment 3, respectively. The generated supervoxel map using different MRI modalities is then

applied to each modality image to extract features. As discussed in Section II-E and shown in Table 1, for each supervoxel, there are 21 features extracted from each modality, so in total 21 features for FLAIR only, 63 features for C-MRI data (e.g. FLAIR, T2 and T1-contrast), and 105 features for C-MRI+DTI (p and q maps). The random forests classification is then performed in each experiment to classify each supervoxel into normal brain tissue and tumour.

Table 3 shows the average results of supervoxel classification for the three experiments, using our clinical dataset. Results show significant improvement for classification of tumour core, oedema and the whole tumour using C-MRI+DTI, compared to use of the FLAIR image or the conventional MRI data alone.

3.3. Segmentation Results

The Dice score is used to evaluate the overlap ratio between the segmentation results and the manual segmented gold standard:

$$\mathrm{DC} = \frac{2|M \cap S|}{|M|+|S|}, \tag{8}$$

where, M and S are the manual and proposed segmentation masks, respectively.

Range of dice scores are 0 to 1 with closer to 1 representing better segmentation. Table 4 shows Dice scores comparing the ground truth with our automated method using the three experiment sets. Results show significant improvement in the segmentation of tumour core using the C-MRI+DTI approach with a Dice score (DC) of 0.78 compared to C-MRI (DC= 0.67) and the single FLAIR image (DC=0.54). This demonstrates that adding DTI increases

the tumour segmentation accuracy for multimodal approach.

Fig. 7 shows examples of the segmentation of tumour core and oedema with three grade IV tumours using FLAIR only, C-MRI and C-MRI+DTI. It can be seen that several supervoxels are wrongly classified, e.g. false positive regions (FPs), in the segmented masks when using FLAIR and C-MRI images (see Fig. 7(c2) and (c3)) whereas adding DTI image modalities reduces these FPs, leading to a more accurate segmentation. For example, in Fig. 7(e1) and (e3), there are areas of tumour core which are missed by the C-MRI protocol, but these tumour areas can be detected by adding DTI modalities as shown in Fig. 7(d1) and (d3). This demonstrates an improvement in segmentation accuracy by the use of both C-MRI and DTI.

3.4. Evaluation on BRATS 2013 Dataset

To evaluate the robustness of our proposed method, it is also applied to the BRATS 2013 [21,22] patient dataset, which consists of 20 high grade and 10 low grade tumour types. In this dataset conventional FLAIR, T1-weighted, T2-weighted and T1-weighted with contrast image modalities are available. Data were acquired from different centres using different MR systems with field strengths of both 1.5T and 3T. The ground truth segmentations are manually provided by a human expert [5]. In this study, due to no DTI data available in the BRATS dataset, we evaluate the multimodal aspect of our proposed method, by calculating tumour segmentation performances using C-MRI (FLAIR, T1, T2 and T1+contrast), compared with that using the single imaging modality (FLAIR). The RF parameters selected by 4-fold cross-validation experiment on our clinical dataset (discussed in section

III-A) were used to assess robustness of parameter selection.

The parameters used for feature extraction are similar to those we used for our clinical datasets. For the supervoxel segmentation the only parameter that is different from analysis of our own data is the initial superpixel size.

This is due to the different voxel dimensions of the two datasets. The voxel dimension for all BRATS data is 1 mm × 1 mm × 1 mm. Therefore, the initial subvolumes are cubes with the same dimensions. The supervoxel size for segmenting both oedema and tumour core is defined as 5 mm × 5 mm × 5 mm considering small tumours in some images. Table 5 presents the average evaluation results using RF for supervoxel classification of tumour core, oedema against the rest of tissues and also classification of whole tumour against the healthy tissue using single modality of FLAIR and multimodal approach on conventional MRI protocols (C-MRI) including FLAIR, T1-weighted, T1-weighted (with contrast) and T2-weighted imaging. Table 5 shows that the classification performances for different tumour regions (e.g. core, oedema, whole tumour) using C-MRI have been significantly improved compared to that using the single FLAIR imaging.

Table 6 shows the Dice score computed between ground truth segmentation and our automated segmentation using both FLAIR and C-MRI, on the 30 tumours of the BRATS dataset. This demonstrates that using a multimodal approach presents better overlap measures for tumour core, oedema, and whole tumour, compared to the use of FLAIR only.

Fig. 8 shows comparison results of our automated method with the ground truth for both tumour core and oedema. Segmentation results are presented in axial slices overlaid on the FLAIR image (Fig. 8(e1), (e3), and (e3)). It can be seen that the segmentation from C-MRI results in better and more accurate tumour segmentation compared to the FLAIR imaging

alone. For the tumour core segmentation, comparing Fig. 8(d2) (yellow) to Fig. 8(c2) (yellow), using a multimodal approach achieved accurate segmentation compared to that using single modality. In particular, Fig. 8(c1) shows that several regions of normal brain are detected as tumour core and Fig. 8(c2) and (c3) show some regions of oedema that are wrongly classified as tumour core. Those regions have been improved in Fig. 8(d1), (d2), and (d3) using C-MRI data.

3.5. Statistical Analysis on the Two Datasets

The Wilcoxon signed-rank test was used on both our clinical dataset and the BRATS 2013 dataset to investigate if there were significant differences in both Dice scores and classification measures of precision, sensitivity and BER, from tumour segmentations obtained using the different imaging protocols, at a 95% confidence level.

Table 7 shows Wilcoxon signed-ranks test statistical results for whole tumour segmentation for the Dice scores and classification measures using the different imaging protocols on our clinical dataset (N=11). Results suggest that there is a statistically significant improvement in Dice scores and in classification measures of precision, sensitivity, BER, when using the C-MRI + DTI multimodal data compared to C-MRI or FLAIR alone.

Table 8 shows the corresponding Wilcoxon signed-ranks test statistical parameters for the BRATS 2013 dataset (N=30). These results also demonstrate a statistically significant improvement in Dice scores and all classification measures when using multimodal C-MRI data compared to FLAIR only. It is noted that there is no DTI available in the BRATS

dataset.

Finally, we combined our results from the two different datasets (i.e. our clinical data and the BRATS data) in a single group containing either FLAIR or C-MRI (N=41). Table 9 shows the corresponding Wilcoxon signed-ranks test statistical results, which also indicate a statistically significant improvement in Dice scores and all classification measures when using the C-MRI protocol, instead of the FLAIR image alone.

4. Discussion

Our supervoxel calculation is based on SLIC [34] which was originally developed for natural images using 2D regular arrays without considering pixel resolutions. Whilst, our 3D clinical dataset is anisotropic, with different voxel resolutions along each dimension. To address this problem, we adapt the distance formulation in the supervoxel calculation from MR data with different acquisition parameters as shown in Equation (3). In this study, two different sets of data with different voxel dimension and slice thickness were used to evaluate our supervoxel method. Our clinical dataset has slice thickness three times more than the in-plane voxel resolutions. Therefore, the initial supervoxel is chosen to be rectangular shape (e.g. 8x8x3). Whilst, the BRATS dataset has been interpolated to 1mm isotropic resolution, so initial supervoxels are defined to be cubic. The supervoxel segmentation boundary for BRATS data has better resolution in the Z direction. This is the main reason why the segmentation results from BRATS data are in general better than that from our clinical data. The results in Table 4 and Table 6 confirm this and show the overall segmentation of tumour for our dataset has average of 0.84 with standard deviation 0.06, whereas for the BRATS dataset they are 0.89 and 0.04 respectively.

One limitation of supervoxel segmentation is that there is a minimum size for supervoxels regarding its parameters and image characteristics. For this reason, the method has a limitation in segmenting very small volumes. The overall Dice score for larger tumour cores is more than 80%; whereas for smaller tumour cores the overlap measure decreases due to the initial supervoxel size. For example, the Dice scores for patient numbers 8 to 11 in Table 4 are relatively low. This is due to very small tumour cores for those data, which only contain a limited number of supervoxels.

To evaluate the robustness and generality of our proposed supervoxel method, it was applied to the BRATS 2013 multimodal dataset. However, this dataset doesn't contain DTI protocols p and q. So we only compare the single modal (FLAIR) against the multimodality (conventional) MRI. The supervoxel map generated from multimodality is different from single imaging modality based on FLAIR. The results show the improvement in segmentation of the tumour core. A zoomed-in image of the overlay of the tumour cores (shown in Fig. 8) is depicted in Fig. 9. To show the comparison between single modal and multimodal approaches, the segmentation results of both methods are overlaid on 2 different protocols, FLAIR and T1+C. As can be seen in Fig. 9, the information from protocol T1+C improves the segmentation of tumour core, as the tumour core has more clear boundaries in this protocol. The homogenous region in the FLAIR image (Fig. 9(a)) causes a wandering boundary (red dent in the figure) during single modality supervoxel segmentation, whereas using multimodal approach with the help of clear tumour core boundary in protocol T2 improves the segmentation accuracy (blue contour in Fig. 9(d)). The false positive region (shown in red in Fig. 9(b)) is the continuing of a supervoxel from adjacent slices. Using multimodal approach, the false positive regions can be successfully

removed from the tumour core.

The results of our proposed method on the BRATS 2013 clinical dataset and the best scores in 2012 and 2013 challenges from other groups [5] are presented in Table 10. The method proposed by Tustison *et al.* [12] was the winner of the on-site BRATS 2013 challenge. Although our testing dataset is different with their dataset, it provides a comparable scale to our method. To fairly evaluate our proposed method, we also provide comparison with the best scores for analyses that used the clinical training dataset for evaluation their results. Reza et al [18] used the training clinical data to evaluate their method and obtained the best results for the same data as we used in this study. The average of the top 10 best results which used the same training dataset of BRATS 2013 according to their website [21] is also presented in Table 10. The comparison results in Table 10 demonstrate a good performance of our method for segmentation both of tumour core and whole tumour, with Dice scores of 0.80 and 0.89, respectively.

5. Conclusion

A supervised learning based method is proposed for segmentation of tumour in multimodal MRI brain tumour images. Supervoxels are calculated using information fusion from multimodal MRI images. A novel histogram of texton descriptors, calculated using a set of 3D Gabor filters with different sizes and orientations, are extracted on each supervoxel from different MRI imaging modalities. A random forests classifier is then used to classify each supervoxel into tumour (including tumour core and oedema) or normal brain tissue. The multimodal supervoxel segmentation method results in inclusion of information from multimodal MRI, which improves multiple tissue boundary segmentation; whilst using the

distribution of local textures inside each supervoxel helps improving the further classification of supervoxel.

The experimental results show that the proposed method achieves promising results in the segmentation of brain tumour core and oedema. Adding features from different MRI imaging protocols increases the classification accuracy of the supervoxels in relation to a manually defined gold standard. Table 2 shows the proportion of the features selected from each protocol using the RF for the segmentation and classification of the tumour. It can be seen that, for our clinical dataset, features extracted from the DTI protocols are included (e.g. 16% from p map and 8% from q map) and our final results show the further improvement of the segmentation and classification performance by combining the p and q protocols into the conventional MRI images. In addition, the proposed supervoxel method has also been evaluated on the BRATS 2013 dataset which also presents the accurate and robust results.

Acknowledgements

This research was supported by European FP7 collaborative Project “MyHealthAvatar” (600929). MRI data were obtained during the EU FP7 “eTUMOUR” project (LSHC-CT-2004-503094).

References

[1] J.R. Fink, M. Muzi, M. Peck, K.A. Krohn, Continuing Education: Multi-modality Brain Tumor Imaging – MRI, PET, and PET/MRI, J Nucl Med. 56 (2015) 1554–1561. doi:10.2967/jnumed.113.131516.

[2] L.P. Clarke, R.P. Velthuizen, M.A. Camacho, J.J. Heine, M. Vaidyanathan, L.O. Hall, R.W. Thatcher, M.L. Silbiger, MRI segmentation: Methods and applications, Magnetic Resonance Imaging. 13 (1995) 343–368. doi:10.1016/0730-725X(94)00124-L.

[3] N. Gordillo, E. Montseny, P. Sobrevilla, State of the art survey on MRI brain tumor segmentation, Magn Reson Imaging. 31 (2013) 1426–1438. doi:10.1016/j.mri.2013.05.002.

[4] X. Xuan, Q. Liao, Statistical Structure Analysis in MRI Brain Tumor Segmentation, in: Fourth International Conference on Image and Graphics, 2007. ICIG 2007, 2007: pp. 421–426. doi:10.1109/ICIG.2007.181.

[5] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.A. Weber, T. Arbel, B.B. Avants, N. Ayache, P. Buendia, D.L. Collins, N. Cordier, J.J. Corso, A. Criminisi, T. Das, H. Delingette, Ç. Demiralp, C.R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K.M. Iftekharuddin, R. Jena, N.M. John, E. Konukoglu, D. Lashkari, J.A. Mariz, R. Meier, S. Pereira, D. Precup, S.J. Price, T.R. Raviv, S.M.S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H.C. Shin, J. Shotton, C.A. Silva, N. Sousa, N.K. Subbanna, G. Szekely, T.J. Taylor, O.M. Thomas, N.J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D.H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, K.V. Leemput, The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Transactions on Medical Imaging. 34 (2015) 1993–2024. doi:10.1109/TMI.2014.2377694.

[6] S. Bauer, R. Wiest, L.-P. Nolte, M. Reyes, A survey of MRI-based medical image analysis for brain tumor studies, Phys Med Biol. 58 (2013) R97-129. doi:10.1088/0031-9155/58/13/R97.

[7] T. Haeck, F. Maes, P. Suetens, An untrained and unsupervised method for MRI brain tumor segmentation, in: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), 2016: pp. 265–268. doi:10.1109/ISBI.2016.7493260.

[8] M.T. El-Melegy, H.M. Mokhtar, Tumor segmentation in brain MRI using a fuzzy approach with class center priors, EURASIP Journal on Image and Video Processing. 2014 (2014) 21. doi:10.1186/1687-5281-2014-21.

[9] I. Njeh, L. Sallemi, I.B. Ayed, K. Chtourou, S. Lehericy, D. Galanaud, A.B. Hamida, 3D multimodal MRI brain glioma tumor and edema segmentation: A graph cut distribution matching approach, Computerized Medical Imaging and Graphics. 40 (2015) 108–119. doi:10.1016/j.compmedimag.2014.10.009.

[10] N. Sauwen, M. Acou, S. Van Cauter, D.M. Sima, J. Veraart, F. Maes, U. Himmelreich, E. Achten, S. Van Huffel, Comparison of unsupervised classification methods for brain tumor segmentation using multi-parametric MRI, NeuroImage: Clinical. 12 (2016) 753–764. doi:10.1016/j.nicl.2016.09.021.

[11] R. Helen, N. Kamaraj, CAD scheme to detect brain tumour in MR images using active contour models and tree classifiers, Journal of Electrical Engineering and Technology. 10 (2015) 670–675. doi:10.5370/JEET.2015.10.2.670.

[12] N.J. Tustison, K.L. Shrinidhi, M. Wintermark, C.R. Durst, B.M. Kandel, J.C. Gee, M.C. Grossman, B.B. Avants, Optimal Symmetric Multimodal Templates and Concatenated Random Forests for Supervised Brain Tumor Segmentation (Simplified) with ANTsR, Neuroinform. 13 (2014) 209–225. doi:10.1007/s12021-014-9245-2.

[13] A. Pinto, S. Pereira, H. Correia, J. Oliveira, D.M.L.D. Rasteiro, C.A. Silva, Brain Tumour Segmentation based on Extremely Randomized Forest with high-level features, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015: pp. 3037–3040. doi:10.1109/EMBC.2015.7319032.

[14] A. Chaddad, Automated feature extraction in brain tumor by magnetic resonance imaging using gaussian mixture models, International Journal of Biomedical Imaging. 2015 (2015). doi:10.1155/2015/868031.

[15] M. Goetz, C. Weber, F. Binczyk, J. Polanska, R. Tarnawski, B. Bobek-Billewicz, U. Koethe, J. Kleesiek, B. Stieltjes, K.H. Maier-Hein, DALSA: Domain Adaptation for Supervised Learning From Sparsely Annotated MR Images, IEEE Transactions on Medical Imaging. 35 (2016) 184–196. doi:10.1109/TMI.2015.2463078.

[16] N. Tustison, M. Wintermark, C. Durst, B. Avants, ANTs and Arboles, in: Proceedings of NCI-MICCAI BRATS, 2013: pp. 47–50.

[17] L. Zhao, D. Sarikaya, J.J. Corso, Automatic Brain Tumor Segmentation with MRF on Supervoxels, in: Proceedings of NCI-MICCAI BRATS, 2013: pp. 51–54.

[18] S. Reza, K.M. Iftekharuddin, Multi-class Abnormal Brain Tissue Segmentation Using Texture Features, in: Proceedings of NCI-MICCAI BRATS, 2013: pp. 38–42.

[19] J. Festa, S. Pereira, J.A. Mariz, N. Sousa, C.A. Silva, Automatic brain tumor segmentation of multi-sequence mr images using random decision forests, in: Proceedings of NCI-MICCAI BRATS, 2013: pp. 23–26.

[20] D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O.M. Thomas, T. Das, R. Jena, S.J. Price, Decision Forests for Tissue-Specific Segmentation of High-Grade Gliomas in Multi-channel MR, in: N. Ayache, H. Delingette, P. Golland, K. Mori (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012, Springer Berlin Heidelberg, 2012: pp. 369–376. doi:10.1007/978-3-642-33454-2_46.

[21] BRATS :: The Virtual Skeleton Database Project, (n.d.). https://www.smir.ch/BRATS/Start2012 (accessed July 3, 2016).

[22] M. Kistler, S. Bonaretti, M. Pfahrer, R. Niklaus, P. Büchler, The virtual skeleton database: an open access repository for biomedical research and collaboration, J. Med. Internet Res. 15 (2013) e245. doi:10.2196/jmir.2930.

[23] A. Crimi, B. Menze, O. Maier, M. Reyes, H. Handels, eds., Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer International Publishing, Cham, 2016. http://link.springer.com/10.1007/978-3-319-30858-6 (accessed July 17, 2016).

[24] R. Verma, E.I. Zacharaki, Y. Ou, H. Cai, S. Chawla, S.-K. Lee, E.R. Melhem, R. Wolf, C. Davatzikos, Multiparametric Tissue Characterization of Brain Neoplasms and Their Recurrence Using Pattern Classification of MR Images, Academic Radiology. 15 (2008) 966–977. doi:10.1016/j.acra.2008.01.029.

[25] A.F. Kazerooni, M. Mohseni, S. Rezaei, G. Bakhshandehpour, H.S. Rad, Multi-parametric (ADC/PWI/T2-w) image fusion approach for accurate semi-automatic segmentation of tumorous regions in glioblastoma multiforme, Magn Reson Mater Phy. 28 (2015) 13–22. doi:10.1007/s10334-014-0442-7.

[26] A.Y.C.C. Wei Wu, Brain tumor detection and segmentation in a CRF (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features, International Journal of Computer Assisted Radiology and Surgery. 9 (2013). doi:10.1007/s11548-013-0922-7.

[27] A. Peña, H. a. L. Green, T.A. Carpenter, S.J. Price, J.D. Pickard, J.H. Gillard, Enhanced visualization and quantification of magnetic resonance diffusion tensor imaging using the p:q tensor decomposition, Br J Radiol. 79 (2006) 101–109. doi:10.1259/bjr/24908512.

[28] W. Wu, A.Y.C. Chen, L. Zhao, J.J. Corso, Brain tumor detection and segmentation in a CRF (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features, Int J Comput Assist Radiol Surg. 9 (2014) 241–253. doi:10.1007/s11548-013-0922-7.

[29] T. Jones, B. Bell, T. Barrick, A novel whole-brain DTI segmentation technique for brain tumour delineation and diagnosis, in: Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM), 2012: p. 188.

[30] T.R. Barrick, C.A. Clark, Singularities in diffusion tensor fields and their relevance in white matter fiber tractography, NeuroImage. 22 (2004) 481–491. doi:10.1016/j.neuroimage.2004.02.001.

[31] FSL, (n.d.). https://fsl.fmrib.ox.ac.uk/fsl/fslwiki (accessed February 28, 2017).

[32] SPM - Statistical Parametric Mapping, (n.d.). http://www.fil.ion.ucl.ac.uk/spm/ (accessed February 28, 2017).

[33] L.G. Nyúl, J.K. Udupa, X. Zhang, New variants of a method of MRI scale standardization, IEEE Trans Med Imaging. 19 (2000) 143–150. doi:10.1109/42.836373.

[34] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, SLIC Superpixels Compared to State-of-the-Art Superpixel Methods, IEEE Transactions on Pattern Analysis and Machine Intelligence. 34 (2012) 2274–2282. doi:10.1109/TPAMI.2012.120.

[35] A.K. Jain, Fundamentals of Digital Image Processing, 1 edition, Prentice Hall, Englewood Cliffs, NJ, 1988.

[36] A. Liaw, M. Wiener, Classification and regression by randomForest, 2 (2002) 18–22.

[37] L. Breiman, Random Forests, Machine Learning. 45 (2001) 5–32. doi:10.1023/A:1010933404324.

[38] M. Soltaninejad, G. Yang, T. Lambrou, N. Allinson, T.L. Jones, T.R. Barrick, F.A. Howe, X. Ye, Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI, Int J CARS. (2016) 1–21. doi:10.1007/s11548-016-1483-3.

[39] R. Taormina, MATLAB_ExtraTrees - File Exchange - MATLAB Central, (n.d.). http://uk.mathworks.com/matlabcentral/fileexchange/47372-rtaormina-matlab-extratrees (accessed February 16, 2016).

Table 1. Number of features which are used for our learning based method.

Features calculated from each supervoxel	One Protocol	Multimodal (e.g. 5 protocols)
Statistical 1st order	16	80
Texton Histogram	5	25
Total	21	105

Table 2. Ranking based on their repetition in nodes of the forests of a RF with 50 number of trees and depth 15.

Experiment	FLAIR	T1C	T2	*p*	*q*
C-MRI	0.61	0.15	0.24	-	-
C-MRI+DTI	0.49	0.09	0.18	0.16	0.08

Table 3. Classification results for supervoxels using single MRI modality (FLAIR).C-MRI (FLAIR, T1-C and T2) and C-MRI+DTI (C-MRI + p and Q MAPS)

		Precision	Sensitivity	BER
Core	Single	69.49 ± 13.05	65.39 ± 8.38	0.18 ± 0.04
	C-MRI	73.64 ± 13.14	69.67 ± 7.59	0.15 ± 0.04
	C-MRI +DTI	83.44 ± 12.36	74.62 ± 18.95	0.13 ± 0.09
Oedema	Single	84.17 ± 7.93	79.28 ± 8.18	0.11 ± 0.04
	C-MRI	85.63 ± 8.24	80.59 ± 8.44	0.10 ± 0.04
	C-MRI +DTI	88.53 ± 7.37	84.57 ± 8.21	0.08 ± 0.04
Whole	Single	88.16 ± 6.38	81.88 ± 9.81	0.09 ± 0.05
	C-MRI	89.54 ± 6.18	83.66 ± 9.16	0.09 ± 0.05
	C-MRI +DTI	92.22 ± 5.80	86.25 ± 9.02	0.07 ± 0.05

Table 4. Dice score comparison for the segmentation of tumour core, oedema and whole tumour using single protocol (FLAIR), C-MRI (FLAIR, T1-Contrast, T2-weighted) and C-MRI+DTI (FLAIR, T1-Contrast, T2-weighted, p and q).

No	FLAIR			FLAIR, T1-Contrast, T2-weighted			FLAIR, T1-Contrast, T2-weighted, p and q		
	Core	Oedema	Whole	Core	Oedema	Whole	Core	Oedema	Whole
1	0.79	0.63	0.75	0.84	0.69	0.77	0.91	0.71	0.79
2	0.55	0.66	0.70	0.60	0.69	0.72	0.84	0.73	0.77
3	0.63	0.70	0.71	0.68	0.70	0.74	0.76	0.71	0.73
4	0.65	0.73	0.78	0.76	0.77	0.82	0.85	0.86	0.91
5	0.56	0.81	0.82	0.62	0.83	0.83	0.68	0.85	0.85
6	0.65	0.72	0.75	0.72	0.73	0.76	0.83	0.81	0.85
7	0.53	0.85	0.86	0.74	0.86	0.87	0.86	0.85	0.86
8	0.42	0.85	0.85	0.58	0.86	0.86	0.62	0.87	0.87
9	0.34	0.82	0.83	0.59	0.83	0.85	0.70	0.89	0.91
10	0.41	0.86	0.86	0.68	0.85	0.86	0.83	0.86	0.88
11	0.34	0.83	0.84	0.52	0.85	0.87	0.67	0.86	0.87
Mean	0.54	0.77	0.79	0.67	0.79	0.81	0.78	0.82	0.84
STD	0.14	0.08	0.06	0.10	0.07	0.06	0.09	0.07	0.06

Table 5. Average classification results for superpixels from Single modality (FLAIR) and multimodal C-MRI (FLAIR, T1, T1-Contrast and T2) of BRATS 2013 dataset (20 high grade and 10 low grade tumour)

		Precision	Sensitivity	BER
Core	Single modal	93.82 ± 5.08	90.69 ± 4.99	0.05 ± 0.02
	C-MRI	98.19 ± 1.90	94.75 ± 3.24	0.03 ± 0.02
Oedema	Single modal	94.01 ± 7.77	87.53 ± 5.91	0.06 ± 0.03
	C-MRI	98.31 ± 1.72	95.89 ± 4.49	0.02 ± 0.02
Whole	Single modal	98.25 ± 2.12	92.29 ± 4.68	0.04 ± 0.02
	C-MRI	99.46 ± 0.66	96.09 ± 3.00	0.02 ± 0.01

Table 6. Comparison results for Dice overlap ratio between manual annotation and the automated segmentation using single modality (FLAIR) and multimodal C-MRI (FLAIR, T1, T1-Contrast and T2) of BRATS 2013.

		Core	Oedema	Whole
Single modality	Mean	0.65	0.79	0.85
	STD	0.09	0.09	0.06
C-MRI	Mean	0.80	0.89	0.89
	STD	0.09	0.05	0.04

Table 7. Wilcoxon signed-ranks test statistical parameters results for the segmentation overlap measure of Dice and the classification measures using different protocols (i.e. FLAIR only, conventional MRI (C-MRI), and conventional MRI plus DTI (C-MRI plus DTI), on our own dataset (11 subjects).

Whole Tumour	FLAIR *vs* C-MRI		FLAIR *vs* C-MRI + DTI		C-MRI *vs* C-MRI + DTI	
	p	z	p	z	p	z
DICE	0.003	-2.956	0.003	-2.952	0.003	-2.940
Precision	0.010	-2.578	0.004	-2.845	0.006	-2.756
Sensitivity	0.003	-2.936	0.003	-2.934	0.008	-2.667
BER	0.024	-2.264	0.007	-2.680	0.008	-2.666

Table 8. Wilcoxon signed-ranks test statistical parameters results for the segmentation overlap measure of Dice and the classification measures using different protocols (i.e. FLAIR only, and Conventional MRI (C-MRI), on BRATS dataset (30 subjects).

Whole Tumour	FLAIR *vs* C-MRI	
	p	z
DICE	< 0.001	-4.723
Precision	< 0.001	-4.021
Sensitivity	< 0.001	-4.762
BER	< 0.001	-4.051

Table 9. Wilcoxon signed-ranks test statistical parameters results for the segmentation overlap measure of Dice and the classification measures using different protocols (i.e. FLAIR only, and Conventional MRI (C-MRI), on both our own dataset and BRATS 2013 (41 subjects).

Whole Tumour	FLAIR *vs* C-MRI	
	p	z
DICE	< 0.001	-5.531
Precision	< 0.001	-4.743
Sensitivity	< 0.001	-5.566
BER	< 0.001	-4.589

Table 10. Comparison with other methods which used BRATS 2013 dataset (MICCAI 2012 and 2013)

Work	Method	Toumor Core (Dice)	Whole (Dice)
Tutison [12]	RF (ANTsR package)	0.78	0.87
Reza [18]	RF+ texture features	0.91	0.92
Top 10 average		0.78	0.87
Our method	RF+ multimodal supervoxel	0.80	0.89

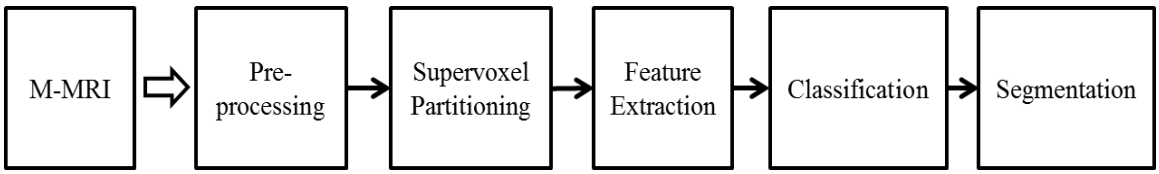
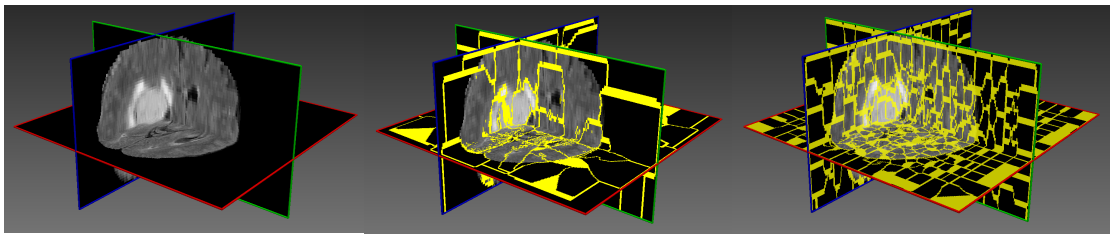


Fig. 1. Flowchart of the proposed multimodal MRI segmentation method for segmentation of brain tumour.



(a) (b) (c)

Fig. 2. Supervoxel segmentation of MRI FLAIR for different supervoxel sizes: a) original image, b) large supervoxel size (30 × 30 × 11), c) small supervoxel size (15 × 15 × 5).

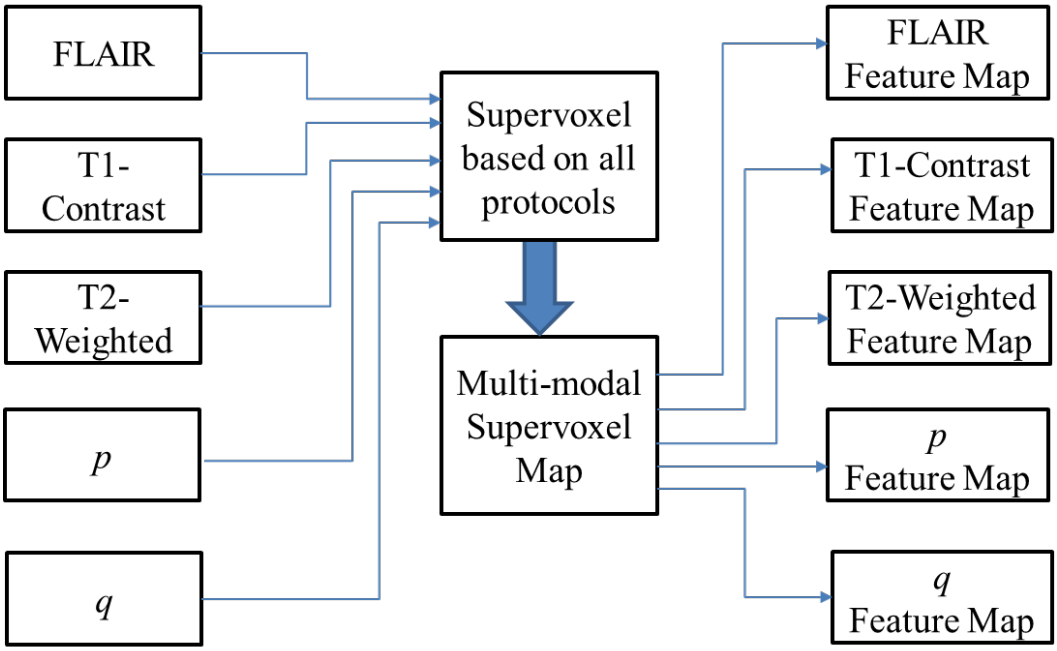


Fig. 3. Framework of multimodal supervoxel segmentation.

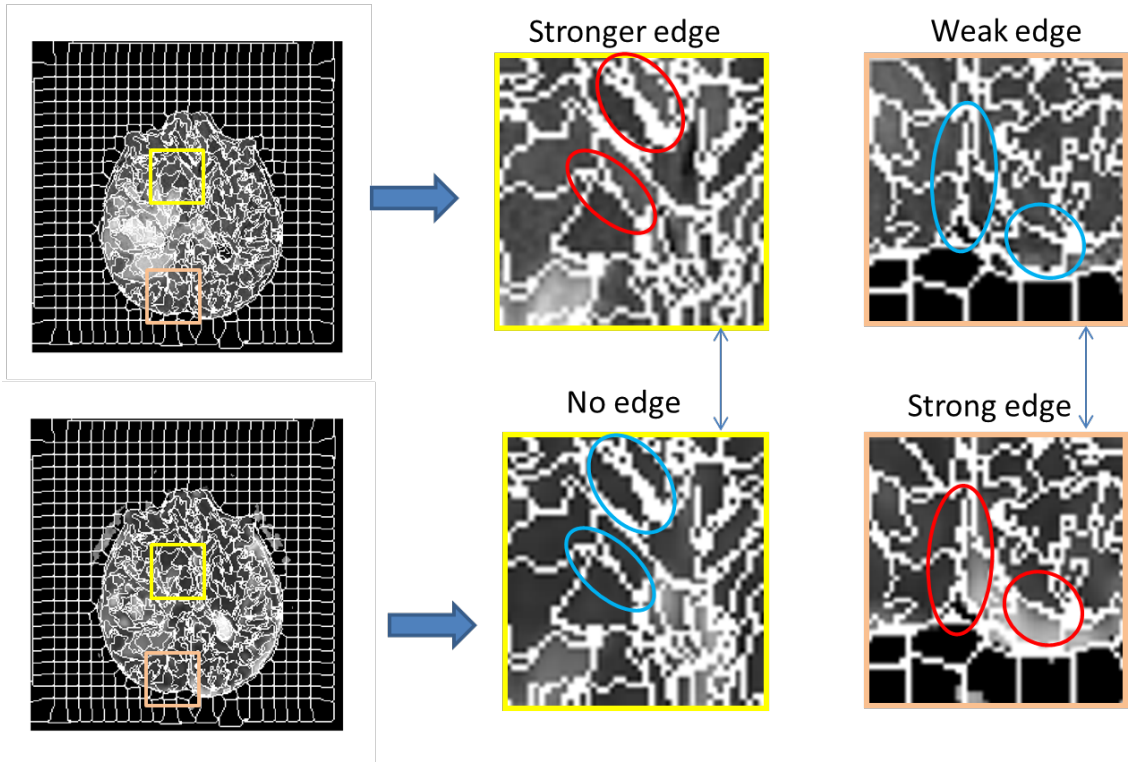


Fig. 4. An example of using a multimodal approach to improve supervoxel boundary by finding the edges which appear weak in one modality (blue ovals), but are apparent in the other modality (red ovals). (a) Upper image: FLAIR image overlaid by multimodal supervoxel segmentation, lower image: *p* map overlaid by the same multimodal supervoxel segmentation. (b) Close up of the region surrounded by the yellow box for both image modalities, (c) Close up of the region surrounded by the red box for both image modalities.

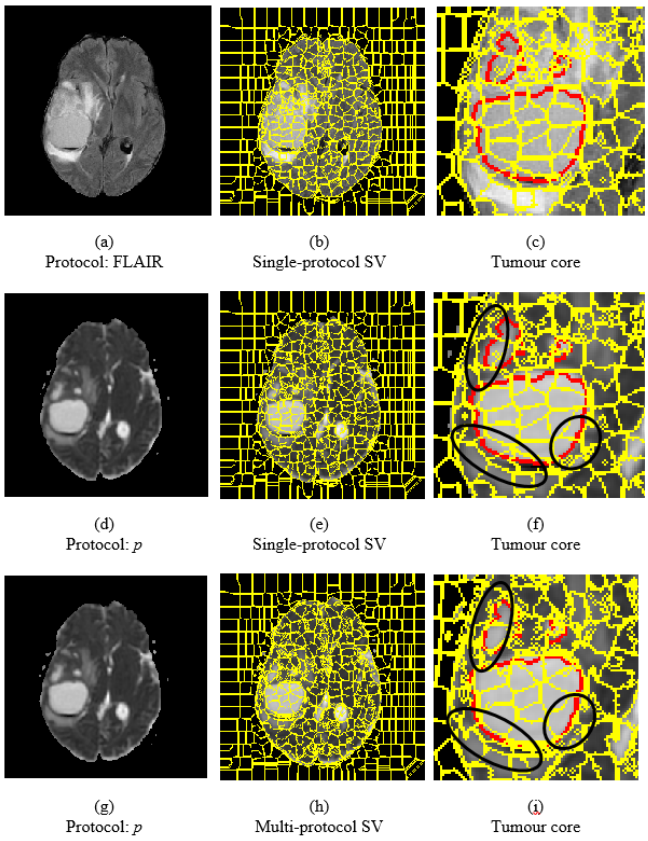


Fig. 5. One comparison example of tumour core supervoxel segmentation (SV) using single modality and multimodal MRI approaches. (a-1) FLAIR, (b-1): overlay of the corresponding supervoxels calculated using single modality (FLAIR), (c-1): zoomed-in of (b-1) on tumour area (to show the details of the SV boundaries) and overlay of tumour core (ground truth from manual delineation shown in red); (a-2): protocol *p* map, (b-2): Supervoxels calculated using single imaging modal (FLAIR) overlay on image protocol *p*, (c-2): zoomed-in of (b-2) on tumour area and overlay of tumour core (red). (a-3): protocol *p*, (b-3): Supervoxels calculated using multimodal (FLAIR, T1+contrast, T2, *p* and *q*) overlay on image protocol *p*. (c-3): zoomed-in of (b-3) on tumour area and overlay of tumour core (red). The boundaries surrounded by black ellipses in (c-2) and (c-3) highlighting the improvement of supervoxel boundary alignment with that of the tumour core using the proposed multimodal SV method. The supervoxels are initially sized 15 × 15 × 5 with $m = 0.2$ compactness.

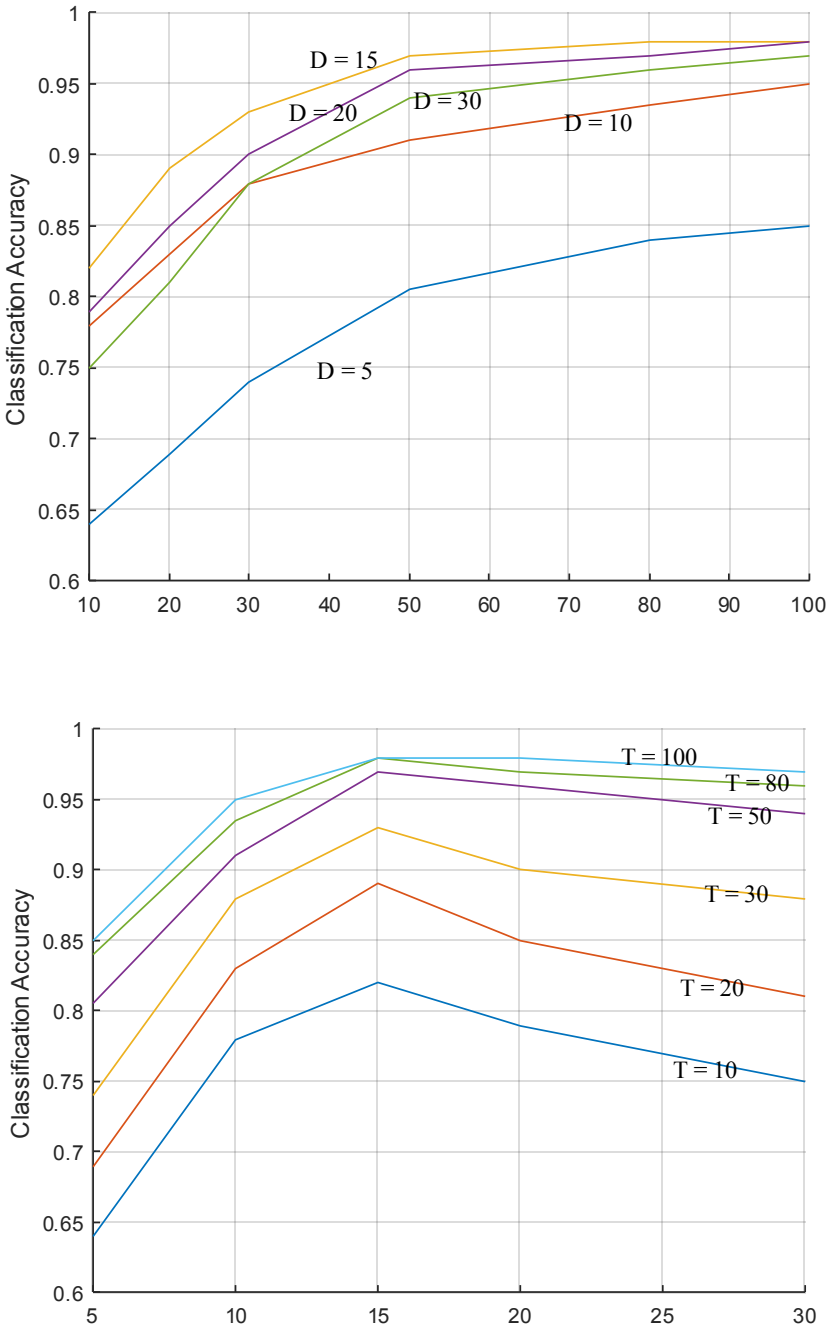


Fig. 6. Upper) Effect of number of trees on RF classification accuracy with different depths. Lower) effect of tree depth on RF classification accuracy with different numbers of trees.

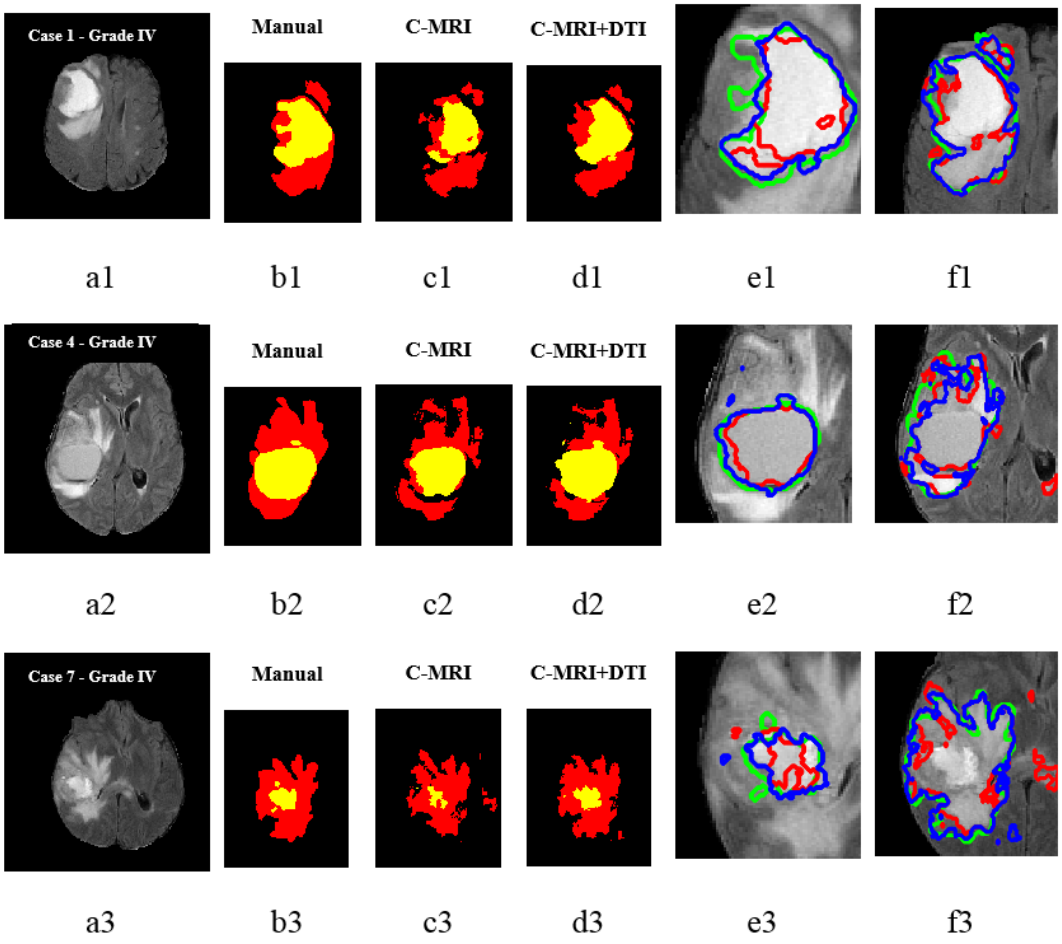


Fig. 7. Comparison examples of segmentation of tumour core and oedema using conventional MRI and conventional MRI plus DTI for three different cases with grade IV tumours. A) FLAIR image, B) manual segmentation of core (yellow region) and oedema (red region) C) segmentation using conventional MRI, D) segmentation using conventional MRI and DTI (M-MRI), E) comparison of both methods C-MRI (red), plus DTI (blue) and manual (green) segmentation for core (zoomed in), F) comparison of both methods C-MRI (red), plus DTI (blue) and manual (green) segmentation for oedema (zoomed in)

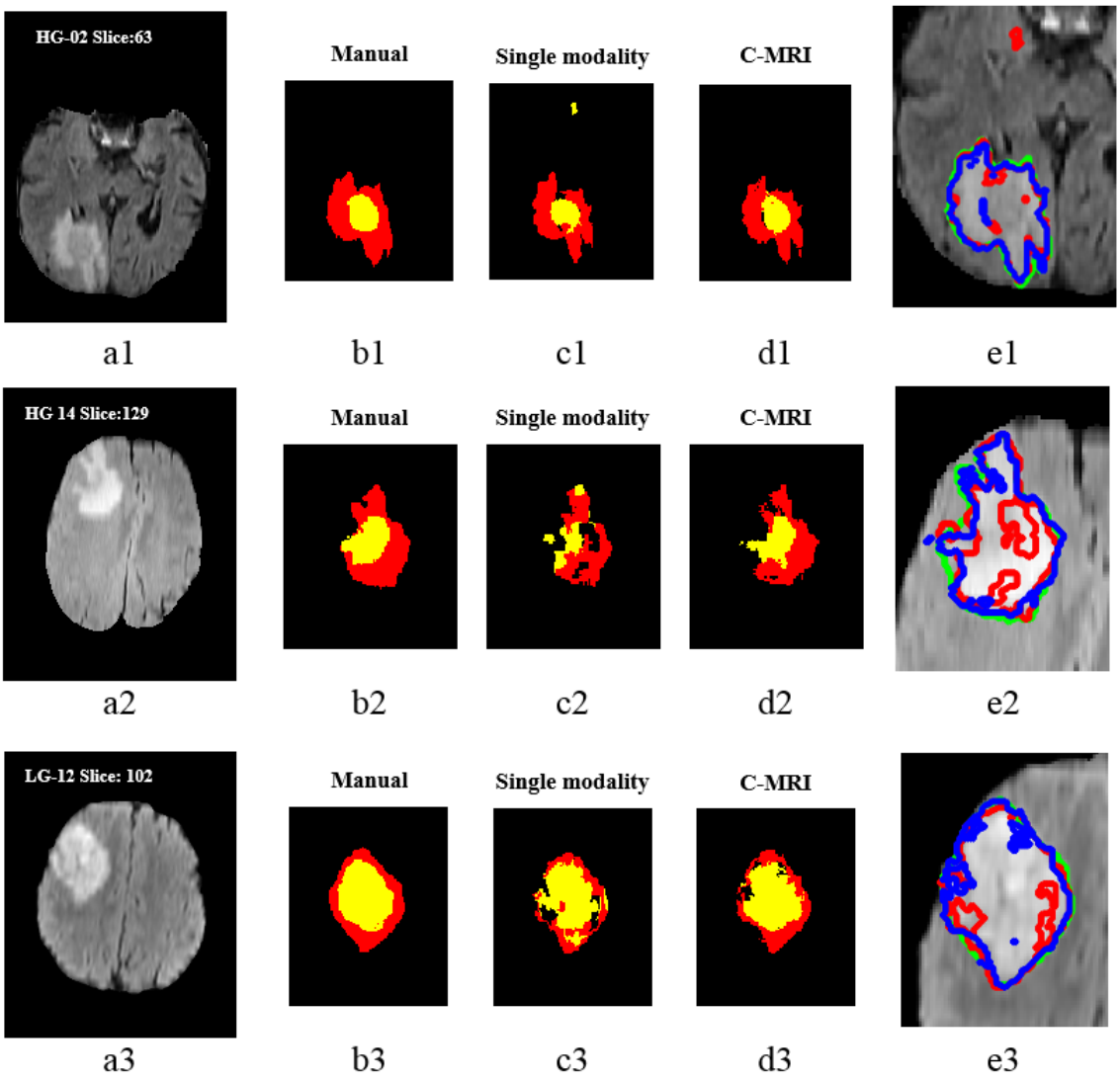


Fig. 8. Segmentation results overlay on the ground truth (whole tumour including oedema and core), using single (FLAIR) and multimodal (conventional MRI including FLAIR, T1, T1-contrast and T2), for three different cases with grade IV tumours; A) FLAIR image, B) manual segmentation of core (yellow region) and oedema (red region) C) segmentation using FLAIR D) segmentation using conventional MRI E) comparison of both methods: single modal (red), multimodal (blue) and manual (green) segmentation for whole tumour (zoomed in)

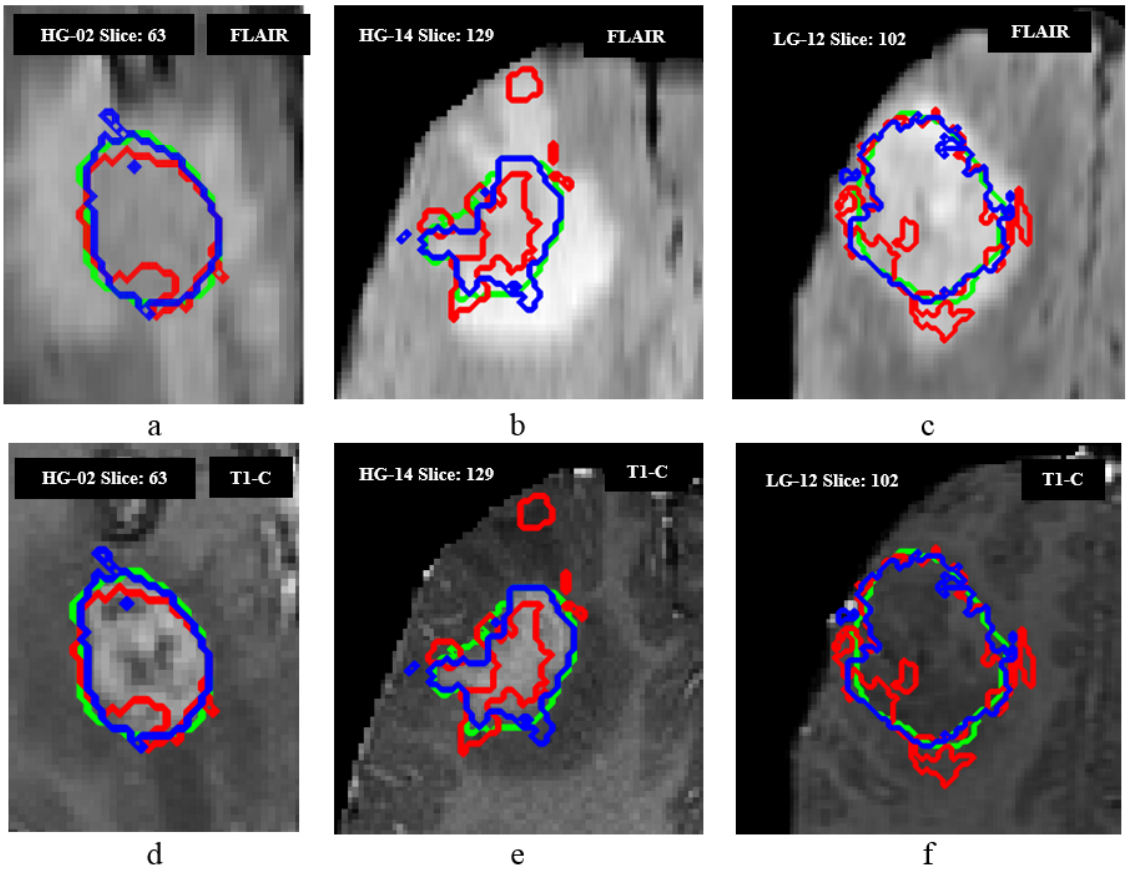


Fig. 9. Comparison between single modality and multimodal segmentation of core. a-c) FLAIR, d-f) T1-C.

Green: manual ground truth, red: single modal, blue: multimodal.