
UNIVERSITY OF
LINCOLN

# Supervised Learning-based Multimodal MRI Brain Image Analysis

M Soltaninejad

Doctor of Philosophy

2017

# Supervised Learning-based Multimodal MRI Brain Image Analysis

**Mohammadreza Soltaninejad**

A thesis submitted in partial fulfilment of the requirements of the University of Lincoln for the degree of Doctor of Philosophy

November 2017

# Acknowledgement

I wish to acknowledge MyHealthAvatar who have funded me during my research and study.

I would like to express my special thanks to my supervisor, Dr Xujiong Ye, for her kind guidance and support. I would also like to express my sincere thanks to my second supervisor Dr Tryphon Lambrou for his valuable support, guidance. I would also like to extend my gratitude to Prof Nigel Allinson, head of Laboratory of Vision Engineering.

I express my gratitude to Dr Guang Yang for making the clinical MRI data available. I am also grateful to Dr Adnan Qureshi for providing the clinical data and clinical assessments.

Special thanks to Prof Franklyn Howe, Dr Timothy Jones, Dr Tomas Barrick for their support and kind comments on my research.

I express my gratitude to my colleague Dr Lei Zhang for his guidance and support during my study.

Thanks to my mother and father, Farkhondeh and Rabi, my beloved brother, Mahdi, my dearest sisters, Zahra and Mansureh, for lots of love and their support of me.

# Abstract

Medical imaging plays an important role in clinical procedures related to cancer, such as diagnosis, treatment selection, and therapy response evaluation. Magnetic resonance imaging (MRI) is one of the most popular acquisition modalities which is widely used in brain tumour analysis and can be acquired with different acquisition protocols, e.g. conventional and advanced. Automated segmentation of brain tumours in MR images is a difficult task due to their high variation in size, shape and appearance. Although many studies have been conducted, it still remains a challenging task and improving accuracy of tumour segmentation is an ongoing field. The aim of this thesis is to develop a fully automated method for detection and segmentation of the abnormal tissue associated with brain tumour (tumour core and oedema) from multimodal MRI images.

In this thesis, firstly, the whole brain tumour is segmented from fluid attenuated inversion recovery (FLAIR) MRI, which is commonly acquired in clinics. The segmentation is achieved using region-wise classification, in which regions are derived from superpixels. Several image features including intensity-based, Gabor textons, fractal analysis and curvatures are calculated from each superpixel within the entire brain area in FLAIR MRI to ensure a robust classification. Extremely randomised trees (ERT) classifies each superpixel into tumour and non-tumour. Secondly, the method is extended to 3D supervoxel based learning for segmentation and classification of tumour tissue subtypes in multimodal MRI brain images. Supervoxels are generated using the information across the multimodal MRI data set. This is then followed by a random forests (RF) classifier to classify each supervoxel into tumour core, oedema or healthy brain tissue. The information from the advanced protocols of diffusion tensor imaging (DTI), i.e. isotropic ($p$) and anisotropic ($q$) components is also incorporated to the conventional MRI to improve segmentation accuracy. Thirdly, to further improve the segmentation of tumour tissue subtypes, the machine-learned features from fully convolutional neural network (FCN) are investigated and combined with hand-designed texton features to encode global information and local dependencies into feature representation. The score map with pixel-wise predictions is used as a feature map which is learned from multimodal MRI training dataset using the FCN. The machine-learned features, along with hand-designed texton features are then applied to random forests to classify each MRI image voxel into normal brain tissues and different parts of tumour.

The methods are evaluated on two datasets: 1) clinical dataset, and 2) publicly available Multimodal Brain Tumour Image Segmentation Benchmark (BRATS) 2013 and 2017 dataset. The experimental results demonstrate the high detection and segmentation performance of the

single modal (FLAIR) method. The average detection sensitivity, balanced error rate (BER) and the Dice overlap measure for the segmented tumour against the ground truth for the clinical data are 89.48%, 6% and 0.91, respectively; whilst, for the BRATS dataset, the corresponding evaluation results are 88.09%, 6% and 0.88, respectively. The corresponding results for the tumour (including tumour core and oedema) in the case of multimodal MRI method are 86%, 7%, 0.84, for the clinical dataset and 96%, 2% and 0.89 for the BRATS 2013 dataset. The results of the FCN based method show that the application of the RF classifier to multimodal MRI images using machine-learned features based on FCN and hand-designed features based on textons provides promising segmentations. The Dice overlap measure for automatic brain tumor segmentation against ground truth for the BRATS 2013 dataset is 0.88, 0.80 and 0.73 for complete tumor, core and enhancing tumor, respectively, which is competitive to the state-of-the-art methods. The corresponding results for BRATS 2017 dataset are 0.86, 0.78 and 0.66 respectively.

The methods demonstrate promising results in the segmentation of brain tumours. This provides a close match to expert delineation across all grades of glioma, leading to a faster and more reproducible method of brain tumour detection and delineation to aid patient management. In the experiments, texton has demonstrated its advantages of providing significant information to distinguish various patterns in both 2D and 3D spaces. The segmentation accuracy has also been largely increased by fusing information from multimodal MRI images. Moreover, a unified framework is present which complementarily integrates hand-designed features with machine-learned features to produce more accurate segmentation. The hand-designed features from shallow network (with designable filters) encode the prior-knowledge and context while the machine-learned features from a deep network (with trainable filters) learn the intrinsic features. Both global and local information are combined using these two types of networks that improve the segmentation accuracy.

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **2D** | Two dimensional |
| **3D** | Three dimensional |
| **AlexNet** | The CNN proposed in (Krizhevsky *et al.*, 2012) |
| **ANN** | Artificial neural networks |
| **BER** | Balanced error rate |
| **BFC** | Bias field correction |
| **BRATS** | Multimodal brain tumour segmentation challenge |
| **BraTumIA** | Brain tumour image analysis |
| **CART** | Classification and regression trees |
| **C-MRI** | Conventional magnetic resonance imaging |
| **CNN** | Convolutional neural networks |
| **CONV** | Convolutional layer |
| **CONV_ReLU** | Convolutional layer + rectified linear unit |
| **CRF** | Conditional random fields |
| **CSF** | Cerebral spinal cord |
| **CT** | Computed tomography |
| **DeCONV** | Deconvolutional layer |
| **DL** | Deep learning |
| **DSC** | Dice similarity score |
| **DTI** | Diffusion tensor imaging |
| **DWI** | Diffusion weighted imaging |
| **ERT** | Extremely randomised trees |
| **FC** | Fully connected layer |
| **FCBF** | Fast correlation-based filter |
| **FCM** | Fuzzy *c*-means |
| **FCN** | Fully convolutional neural network |
| **FCN_RF** | Fully convolutional network and random forests |
| **FCN_Texton_RF** | Fully convolutional network, texton, and random forests |
| **FLAIR** | Fluid attenuated inversion recovery |
| **FN** | False negative |
| **FP** | False positive |
| **GM** | Grey matter |
| **GoogleNet** | The CNN proposed by Google (Szegedy *et al.*, 2015) |
| **HGG** | High grade gliomas |
| **IID** | Independent and identically distributed |
| **ILSVRC** | ImageNet Large-Scale Visual Recognition Challenge |
| **IR** | Inversion time |
| **LeNet** | The CNN proposed by LeCun *et al.* (LeCun *et al.*, 1998) |
| **LGG** | Low grade gliomas |
| **LOO-CV** | Leave-one-out cross validation |
| **MAP** | Maximum a posteriori |
| **MICCAI** | Medical Image Computing and Computer Assisted Intervention |
| **MIUA** | Medical Image Understanding and Analysis |
| **ML** | Machine learning |

| | |
|---|---|
| **MR** | Magnetic resonance |
| **MRF** | Markov random fields |
| **MRI** | Magnetic resonance imaging |
| **mRMR** | Minimum redundancy maximum relevance |
| **PD** | Proton density |
| **POOL** | Pooling layer |
| **ReliefF** | Feature selection method proposed in (Kononenko, 1994) |
| **ReLU** | Rectified linear unit |
| **ResNet** | Residual networks |
| **RF** | Random forests |
| **ROI** | Region of interest |
| **SBMLR** | Sparse logistic regression with Bayesian regularisation |
| **SFTA** | Segmentation based fractal texture analysis |
| **SGD** | Stochastic gradient descent |
| **SKIP** | Skip layer |
| **SLIC** | Simple linear iterative clustering |
| **SLogReg** | Sparse logistic regression |
| **SP** | Superpixel |
| **SP_ERT** | Superpixel and extremely randomised trees |
| **SP_SVM** | Superpixel and support vector machines |
| **SPEC** | Spectral feature selection |
| **SPM** | Statistical parametric mapping |
| **STD** | Standard deviation |
| **SV** | Supervoxel |
| **SVM** | Support vector machines |
| **TE** | Echo time |
| **TN** | True negative |
| **TP** | True positive |
| **TR** | Repetition time |
| **U-Net** | The CNN proposed in (Ronneberger *et al.*, 2015) |
| **VGGNet** | The CNN proposed in (Simonyan and Zisserman, 2014) |
| **VSD** | Virtual skeleton database |
| **WM** | White matter |
| **ZFNet** | The CNN proposed in (Zeiler and Fergus, 2014) |

# Chapter 1

# Introduction

## 1.1 Problem statement

The incidence rate of brain related tumours in the United Kingdom has been estimated to be approximately 11,000 cases in 2014 from which 46% were primary brain tumour ("Cancer Research UK," n.d.). Although the relative occurrence of brain cancer is low compared to other types of adult cancers[1], they affect significantly the lives of the people more than other types of cancer ("Cancer registration statistics, England Statistical bulletins - Office for National Statistics," n.d.).

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. The primary tumours are those started within the brain. The majority of primary brain tumours originate from glial cells (termed glioma) and are classified by their histopathological appearances using the World Health Organisation (WHO) system into low grade glioma (LGG) and high grade glioma (HGG).

Medical imaging modalities are used for detection and assessment of tumours. Among different imaging modalities, magnetic resonance imaging (MRI) is one of the most widely used modalities for clinical diagnosis, treatment selection, prognosis and to aid surgery and radiotherapy planning (Fink *et al.*, 2015). Due to the multimodal nature of MRI, which will be explained in Chapter 2, there are a range of image types and contrasts that enable a subtle radiological assessment of tumour type.

Delineation of the tumour boundary and assessment of tumour size are needed for patient management in terms of treatment planning and monitoring treatment response (Eisele *et al.*, 2016), and current guidelines incorporate the use of conventional MR images (C-MRI) (Niyazi *et al.*, 2016; Wen *et al.*, 2010). C-MRI can also be useful to help define the target volumes for radiotherapy planning of high-grade gliomas (Aslian *et al.*, 2013; Niyazi *et al.*, 2016). Tumour assessment requires accurate full 3D volume measurement of the tumour which is obtained by manually drawing around the region of interest (ROI). Manual segmentation around tumour margins on a slice-by-slice basis is time-consuming and can take 12 minutes or more per

[1] The percentage of brain tumour was 3% of total cancer cases in the UK in 2014 ("Cancer Research UK," n.d.).

tumour, with semi-automatic methods taking 3 to 5 minutes (Aslian *et al.*, 2013; Odland *et al.*, 2015). Furthermore, a human has limitations in detecting the visual features of the image which increases the risk of human error in manual segmentation. Therefore, an automated segmentation that is not subject to operator subjectivity may be beneficial (Aslian *et al.*, 2013), especially for the large-sized MRI data.

## 1.2 Motivations

Using computer-aided procedures for medical diagnosis and treatment tasks is a fast-growing field of research nowadays. Computer based analysis and measurements of medical images help the clinicians to obtain the measures and identifications faster and more accurate. Medical image analysis plays an important role in clinical procedures related to brain tumours by providing clinicians with automated (or semi-automated) computing tools to help them in diagnosis and treatment tasks. For brain tumours, accurate segmentation may aid the fast (approximately 5 minutes for each patient image) and objective measurement of tumour volume and also find patient-specific features that aid diagnosis and treatment planning (Gordillo *et al.*, 2013). However, automated segmentation of brain tumours is a very challenging task due to their high variation in size, shape and appearance (e.g. image uniformity and texture) (Patel and Tse, 2004).

Many segmentation methods have been proposed for brain tumour segmentation which will be reviewed in Chapter 3. Despite much effort being devoted to the segmentation problem, brain tumour segmentation remains an ongoing research topic.

Most of the existing studies on brain tumour segmentation are performed on conventional MRI protocols, which are based on qualitative image intensities. Considering the advanced MR acquisition protocols, i.e. diffusion tensor imaging (DTI), in the segmentation process may provide more useful information to increase the accuracy. The isotropic ($p$) and anisotropic ($q$) diffusion components derived from DTI (Peña *et al.*, 2006) provide parameters that relate to the microstructure of the brain tissues. The hypothesis of combining DTI and C-MRI is that they may provide quantitative features that increase the classification accuracy and improve tumour segmentation results. Furthermore, many LGG tumours do not show contrast enhancement hence conventional images are used to define the tumour extent and volume. A study has shown that LGG volume and growth rate can be used to assess whether patients are at risk with tumours likely to undergo an early malignant transformation (Rees *et al.*, 2009). Using advanced MR techniques may tackle this problem which should be investigated by combining and comparing to the conventional MR protocols.

Learning based segmentation techniques require a lot of training data which increases the complexity and computing time and memory. Since most segmentation algorithms are based on pixel classification in single/multiple images, in the case of multimodal MR images, the large number of voxels to be processed will significantly increase computational burden. Partitioning the images into small subregions with homogenous properties will decrease the data dimensionality by decreasing the feature space.

Due to the recent advances in deep neural networks (DNNs) in recognition of the patterns in the images, most of the recent tumour segmentations have focused on deep learning methods. Amongst DNNs, the methods based on deep convolutional neural networks (CNNs) has recently provided the best performance in the computer vision and brain tumour segmentation competitions. The CNNs can learn the image patterns in different levels of hierarchy and resolutions. However, the CNNs are, in fact, classifiers which have been used for whole-image classification (Krizhevsky *et al.*, 2012) or local tasks such as object detection (Sermanet *et al.*, 2013). For the task of image segmentation in the pixel resolution level, CNNs have been modified by adding pre- or post-processing blocks (Hariharan *et al.*, 2014). However, these approaches have limitations of lacking a whole end-to-end learning, since the additional blocks are independent from CNN training process. Recently, fully convolutional networks (FCN) have been suggested for dense (i.e. per-pixel) classification with the advantage of end-to-end learning (Long *et al.*, 2015), without requiring those additional blocks in CNN-based approaches. FCN can take an input image with any size, yielding the hierarchy of the features, and provide dense prediction with input-matching size. Despite the advantage of dense pixel classification, FCN-based methods still have limitations of considering the local dependencies in higher resolution (pixel) level. The loss of data, which occurs in the pooling layers, results in coarse segmentation. This limitation will be addressed in this thesis by incorporating high resolution hand-designed textural features which consider local dependencies of the pixel. Texton feature maps (Arbelaez *et al.*, 2011) provides significant information on multi-resolution image patterns in both spatial and frequency domains. This is an inspiration to combine texton features to a partially end-to-end learning process in order to improve the segmentation. The term “partially” is considered for end-to-end learning since the proposed method in this thesis is trained on a pretrained model, which will be discussed later in Chapter 6.

Most classification-based techniques have proposed and/or optimised the hand-designed features, while deep learning based methods automatically learn the features from the images. A hypothesis is that combining both hand-designed and machine-learned features encodes global information and local dependencies into feature representation, which results in more accurate segmentation.

## 1.3 Aims and objectives

The aim of this research is to develop automatic image processing techniques to accurately detect and segment the brain tumour tissue subtypes from multimodal MR images, including conventional and advanced acquisition techniques. This thesis will focus on statistical learning based medical image segmentation using hand-designed and machine-learned features.

To achieve this aim, the following objectives are considered:

- Developing and validating an automated method for a single MRI modality to segment the abnormal tumour part from the normal brain tissues.
- Building a generic framework which combine both C-MRI and DTI to incorporate information from multimodal clinical MRI images. Since each imaging protocol contains specific features from the tissues, merging them together may provide more accurate segmentation of tumour tissue subtypes.
- Exploring a new feature representation which combines hand-designed features (e.g. texton) considering local dependencies, and machine-learned features (from FCN) which provides better object localisation, for accurate segmentation of brain tumours.
- Evaluating the proposed algorithms by conducting experiments on different datasets, i.e. a clinical dataset, which are acquired from St George's Hospital Trust London, and a publicly available dataset of Multimodal Brain Tumour Image Segmentation Benchmark (BRATS) ("BRATS :: The Virtual Skeleton Database Project," n.d.; Kistler *et al.*, 2013; Menze *et al.*, 2015).

## 1.4 Contributions

The main contribution of this thesis can be summarised as follows

- Developing a fully-automated learning based method for detection and segmentation of the abnormal tissue associated with brain tumours as defined by the $T_2$ hyperintensity from Fluid Attenuated Inversion Recovery (FLAIR) MRI as a single protocol (Chapter 4). The previous methods have used multi-protocols to perform the automatic segmentation (Davy *et al.*, 2014; Havaei *et al.*, 2017), while this thesis introduced a method that uses one single protocol to segment the tumour. However, this single-modality method is suitable for segmentation of the whole tumour. Further segmentation of the tumour tissue subtypes requires more protocols, which will be discussed later chapters.
- Incorporating advanced MR acquisition protocols, i.e. DTI alongside with C-MRI for accurate segmentation of brain tumours by fusing the image intensities and features into a unified framework. The isotropic ($p$) and anisotropic ($q$) diffusion components derived

from DTI (Peña *et al.*, 2006), which are related to the microscopic structures of the tissues, provide more information about tumour structure which improves the multi-class tumour segmentation (Chapter 5). Using DTI protocols in segmentation of brain tumour via superpixel analysis and random forest classification is one of the novelties of this thesis.

- Proposing a unified framework for partitioning the multimodal MR images into small clusters (e.g. supervoxels) by incorporating the MR volumetric characteristics, i.e. voxel dimension and slice thickness. The previous methods (Su *et al.*, 2013) have used the pixels in the raw slices without considering the voxel characteristics. The information from multimodal images is combined to produce supervoxel boundaries across multiple image protocols. The advantage of the supervoxel based method is that the required computation for classification in the new feature space can be significantly reduced (Chapter 5).
- Proposing the histogram of texton descriptors particularly for superpixels/supervoxels using Gabor filters as one of the main features, since they are able to distinguish various textural patterns in the image. The previous methods based on superpixel histogram (Fulkerson *et al.*, 2009) have not used texton at a superpixel level. The previous texton-based method (Yu *et al.*, 2012) has used Gaussian filter banks. The related works, which used the histogram of Gabor filter responses as the representation of the features (Yi and Su, 2014), suggested using fixed-sized non-overlapping blocks of the images. In this thesis, flexible superpixel patches are used instead of the fixed blocks. Textons were calculated using a set of Gabor filters with different sizes and orientations, to increase the performance for classification of brain tumour superpixels (Chapter 4) and supervoxels (Chapter 5). Also, another novelty of texton features in Chapter 5 is using 3D Gabor filter banks for MRI volumetric datasets.
- Proposing a novel fully automatic learning based segmentation method, by applying hand-designed and machine-learned features to the state-of-the-art random forest (RF) classifier. The machine-learned FCN based features detect the coarse region of the tumour while the hand-designed texton descriptors consider the spatial features and local dependencies to improve the segmentation accuracy (Chapter 6). The previous Gabor-based CNN methods either fused the Gabor filters to the architecture of CNN (Luan *et al.*, 2017) or used the feature maps from the Gabor filters as an input to the network (Yao *et al.*, 2016). In this thesis, the Gabor-based textons features are considered as a neighbourhood system in the pixel level to compensate the loss of information that occurs in the pooling layers of the FCN.

The current research has resulted in 5 papers (one published journal, one journal under revision, one magazine, and two conferences) that are listed in Appendix 1.

## 1.5 Thesis Structure

Chapter 2 describes the clinical background of the brain tumour segmentation in MRI images. The focus will be on MRI acquisition technique which is common for brain tumour clinical tasks. Different MRI modalities will be explained. The datasets which are used for evaluation will be described followed by the evaluation protocols for brain tumour segmentation.

Chapter 3 presents the technical literature review on brain tumour segmentation using different MRI modalities. The chapter will also analyse specifically the related research work which use the most common publicly available MRI dataset specialised for the field of brain tumour segmentation, i.e. Brain Tumor Image Segmentation Benchmark (BRATS), and the relevant challenge, i.e. Medical Image Computing and Computer Assisted Intervention (MICCAI).

Chapter 4 investigates single modality learning based brain tumour segmentation using hand-designed features. FLAIR is used to detect the tumour since it is the most common clinically acquired protocol to detect and segment complete tumour structure. The texton map will be generated from the FLAIR image, from which texton histogram is calculated for each superpixel and will then be used as one of the main features. Extremely randomised trees (ERT) classifier will be investigated which is a powerful classifier that can deal with high dimensional features and large-sized unbalanced data.

Chapter 5 investigates multimodality brain tumour segmentation in three-dimensional space using hand-designed features. The texton histogram and feature maps are calculated for the supervoxels from the 3D image volume. The segmentation of brain tumour is developed further to its tissue subtypes, i.e. core and oedema, by defining multi-object classification problem. Effect of advanced MR imaging techniques (DTI) is also investigated and compared to conventional MRI protocols.

Chapter 6 introduces using the combination of hand-designed and machine-learned feature for learning based segmentation of brain tumours. The machine-learned features are extracted from the FCN. The main idea is to overcome the drawbacks of machine-learned features by considering the local dependencies which are obtained by hand-designed textural features. The algorithm is also further extended to the segmentation of more details from the tumour structures, i.e. oedema, necrosis, enhancing and non-enhancing tumour cores. This will also make the method comparable with other state-of-the-art work which are using the public datasets.

Chapter 7 summarises the thesis and provides the discussion and conclusion, and presents the future directions.

# Chapter 2

# Clinical Background

## 2.1 Introduction

Application of medical imaging for brain tumour diagnosis has developed over the past decades. The aim of brain tumour imaging is to identify the location and size of the tumour. This will help the clinical tasks such as diagnosis, surgical and radiotherapy planning. It is also used to evaluate the treatment results, e.g. follow up study after treatment.

Regarding the developments in MRI systems, they are now widely used for brain tumour patient evaluation tasks (Jenkinson *et al.*, 2007). The advantages of MRI compared to other techniques, such as computed tomography (CT) images can be summarised as: high contrast between the soft tissues, higher resolution than CT, and non-ionising radiation.

MRI sequences are generally classified into "conventional" and "advanced". Conventional MRI (C-MRI) techniques provide qualitative images of the tissues. Advanced MRI images provide quantitative or semi-quantitative measurements of the brain tissues. In this chapter, firstly the conventional MRI will be introduced followed by its application in brain tumour diagnosis. Then, the advanced MRI techniques on diffusion imaging will be explained.

## 2.2 Brain Tissues

The brain is the most complex organ of the body and it consists of many parts. The most distinguished parts in MR images are Grey Matter (GM), White Matter (WM) and Cerebrospinal Fluid (CSF). GM is the major component of the brain. It consists of mostly body cells and few myelinated axons. WM contains myelinated axons and glial cells. CSF is a clear fluid exists in ventricular system inside and around the brain and spinal cord.

## 2.3 Conventional MRI

Magnetic resonance (MR) is defined as the result of interaction between the magnetic moment of a nucleus spin and an external magnetic field. Three types of magnets are used to create the MR signal and are electromagnet, permanent magnet, and superconducting magnet. Superconducting magnets are extensively used in modern MR scanners which can produce very strong fields of up to 8 Tesla. In current clinical application, the strength of 1.5 to 3 Tesla are used.

### 2.3.1 The Physics behind MRI

MRI utilises the properties of spin to acquire images. The atom elements, i.e. protons, neutrons and electrons, spin around a central axis. Nuclei with an odd mass number (MR active nuclei), such as hydrogen ($H_1$), create a net nonzero spin which acquire a magnetic moment. Their magnetic moment will align their axis of rotation when they are exposed to an external magnetic field.

When no magnetic field is applied to a MR active nuclei, the magnetic moment of the nuclei is oriented in random directions. Therefore, the net magnetic field will be zero. By applying an external magnetic field of $B_0$, the nuclei will align along the flux lines of field.

When a hydrogen nucleus is exposed to an external magnetic field, a secondary spin will be added to its normal spin which is like wobbling around its magnetic moment around $B_0$. This spin is called "precession" and forces the magnetic moments to have a circular precessional path at precessional frequency speed which is called "Larmor frequency".

### 2.3.2 Resonance

Resonance is occurred when a radiofrequency (RF) pulse ("excitation" pulse) is applied at the same energy of precessing hydrogen nuclei at angle of 90 degrees to the field $B_0$. At the resonance, two phenomena will occur which are energy absorption and phase coherence.

In the presence of external magnetic field ($B_0$), the number of spin-up and spin-down nuclei are equal (Figure 2-1-(a)). The net magnetisation vector (NMV) lies in the transverse ($X$-$Y$) plane (90), which is known as "flip angle". The hydrogen nuclei absorb energy from the RF pulse. This will increase the number of high energy (spin-down) nuclei (Figure 2-1-(b)). The magnitude and duration of the RF pulse affect the magnitude of the flip angle. By increasing the magnetic field $B_0$ the required energy for generating resonance will also increase. It should be noted that, the Figure 2-1 shows a schematic representation of NMV before and after applying the RF pulse to the MR active nuclei.

Phase coherence occurs at resonance when the magnetic moment of hydrogen is aligned in the same position of the processional route around the magnetic field $B_0$. This is also called in-phase (coherent). It results in a superimposed magnetic vector in the $X$-$Y$ plane which is depicted in Figure 2-1-(d) which is called transverse magnetisation.

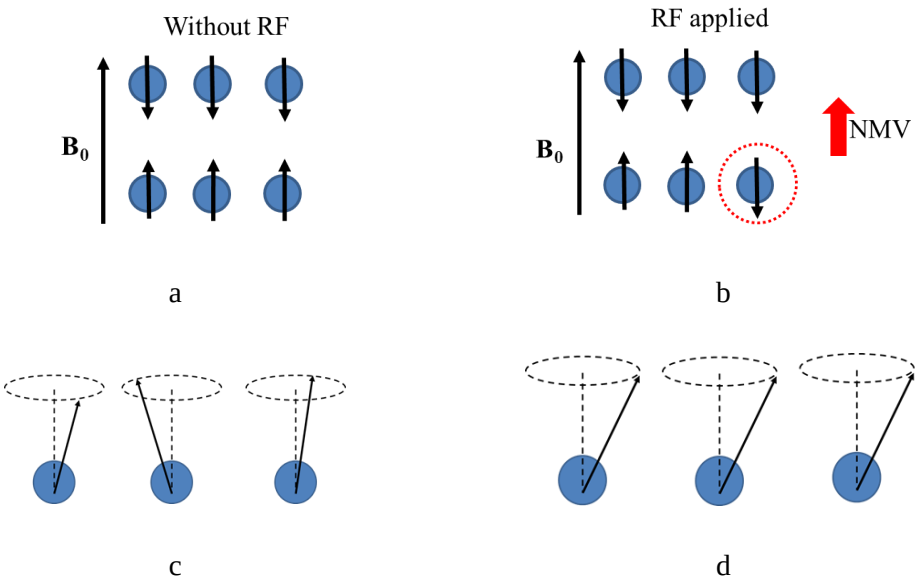



Figure 2-1 Effect of applying RF pulse when the nuclei is exposed to the external magnetic field. NMV which is related to the spin-up and spin-down nuclei: a) without application of RF pulse, and b) when a RF pulse is applied the number of spin-down (high energy) nuclei increases which results in a NMV. Phase coherence: c) out of phase or incoherent in the absence of RF pulse, and d) in-phase or coherent when applying the RF pulse at Larmor frequency.

### 2.3.3 MR Signal Generation

To create a MR signal, a strong and constant magnetic field is applied to the target sample or tissue. As explained in Section 2.3.2, resonance produces in-phase magnetisation precessing in the transverse plane. This magnetisation can induce a voltage when it cuts across the receive coil which creates the MR signal. The magnitude of the signal is related to the amount of magnetisation in the transverse plane and the frequency of the signal is equal to Larmor frequency.

After termination of the RF pulse, the nuclei will lose the energy obtained from the RF pulse and the NMV tries to realign with the external magnetic field $B_0$. The magnetisation in longitudinal plane increases which is known as recovery and has exponential properties. Meanwhile, the magnetisation in the transverse plane decreases exponentially which is known as “decay”. The induced voltage magnitude in the receiver coil will decrease during the decay which is called free induction (FID) signal. The effect of applying RF pulse and creating the FID signal is depicted in Figure 2-2.

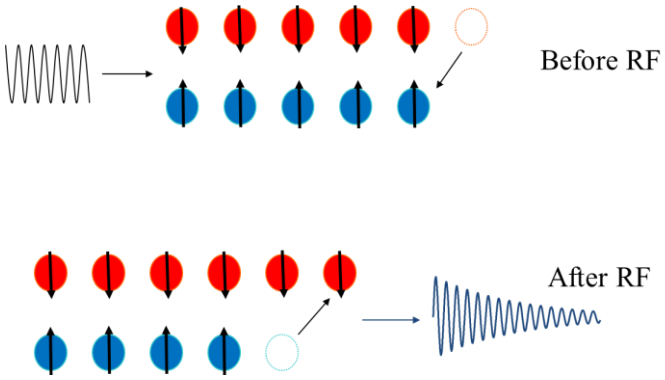
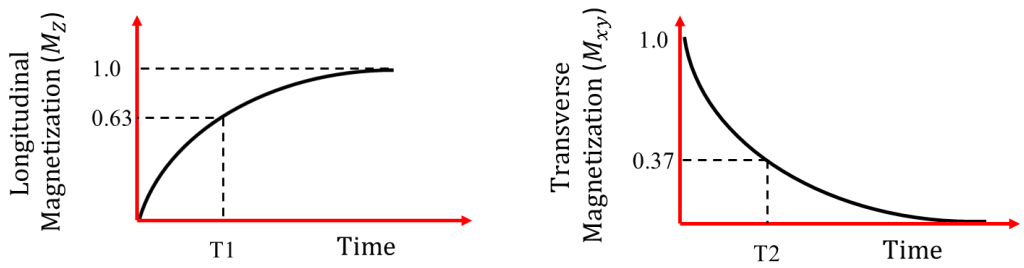



Figure 2-2 Effect of applying RF pulse and generating the FID signal, high energy or spin-down (red) and low energy or spin-up (blue).

### 2.3.4 Relaxation

During the relaxation phase, the hydrogen nuclei will discard the energy that was absorbed when RF was applied. As a result, the NMV returns back to the initial $B_0$ and the magnetic moments of hydrogen nuclei lose their coherency. The longitudinal magnetisation is recovered by the recovery process which is called T1 recovery. And the transverse magnetisation is decayed by the process T2 decay. Figure 2-3 presents the recovery and decay processes.




a b

Figure 2-3 a) T1 recovery curve which represents exponentially increasing longitudinal magnetisation. T1 recovery is the time taken for 63% of the longitudinal magnetisation ($M_z$) to recover b) T2 decay curve which represents the decay of magnetisation in transverse plane ($M_{xy}$) after switching off the RF pulse. T2 relaxation is the time taken for 63% of the transverse magnetisation to be faded.

#### T1 Recovery

When nuclei release their energy, which is called spin-lattice relaxation, it results in T1 recovery. This leads the magnetic moments of nuclei to recover longitudinal magnetisation.

The recovery is exponential with T1 relaxation time which is the recovery time constant (Figure 2-3 (a)). The relaxation time is defined as the time taken for 63% of the longitudinal magnetisation to recover. T1 depends on the inherent characteristics of the tissue and the magnetic field strength.

**T2 Decay**

When the nuclei transfer the energy to the surrounding nuclei the result is a T2 decay which is the interaction of nuclear magnetic fields between the corresponding nucleus. The decay process will cause loss of magnetisation in the transverse plane which is also called spin-spin relaxation. The decay is exponential with T2 relaxation time which is the decay time constant (Figure 2-3 (b)). T2 decay time is defined as the time taken for 63% of the transverse magnetisation to be faded. T2 also depends on the inherent characteristic of the tissue and the magnetic field strength and is usually 5 to 10 times faster than T1 recovery time.

### 2.3.5 Tissue Contrast

By using a specific pulse sequence, the relaxation properties can be controlled which results in tissue contrast. The pulse sequence itself can be controlled by a combination of RF pulses, signal sampling and the time periods (Sprawls, 2000). Figure 2-4 shows a RF sequence in the upper area.

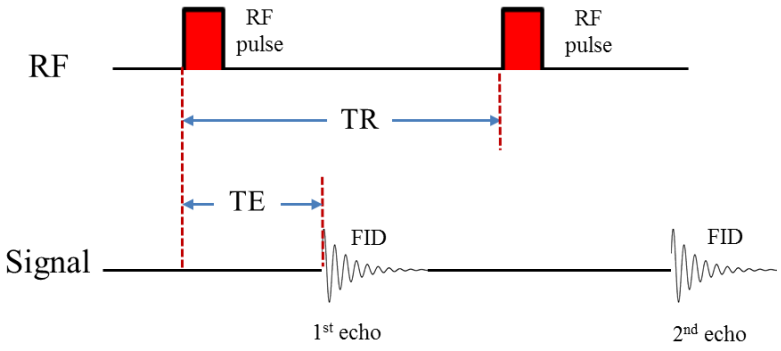



Figure 2-4 RF sequence and the FID induced signal. TR and TE are related to the repetition time and echo time, respectively.

RF pulses generate spin excitation of the nuclei and can manipulate them and a signal echo can be obtained. Magnetic field gradients can select the slice to be imaged and spatially encode the induced signal in the receiver coil. The frequency and phase of the received signal are encoded to obtain the location information. The intensity of the voxel in the MR image is

determined by three components which are: the number of protons in the voxel or "proton density", T1 recovery, T2 decay.

**Repetition time (TR)**

TR presents the time from the beginning of applying the RF pulse to the beginning of the next pulse which is measured in milliseconds. It also specifies the relaxation time which is applicable between the end of the first RF pulse and the next one which determines the amount of T1 relaxation (Figure 2-4).

**Echo time (TE)**

Echo time is the time from the beginning of the RF pulse to the peak time of the FID signal and is measured in milliseconds (Figure 2-4). It also specifies the decay time of transverse magnetisation which determines the amount of T2 relaxation. T1 and T2 properties depend on the tissue's inherent energy, the density of the molecules, and how the molecular wobbling rate matches the Larmor frequency of hydrogen. Optimal tissue contrast can be determined by adjusting the TE and TR properties of the MR sequence for specific imaging protocols and clinical applications. Table 1 presents these properties. Table 1 presents T1 and T2 relaxation times for brain tissues at a magnetic field of 1 Tesla. Figure 2-5 presents the weighting of magnetic resonance images for different MR protocols. Figure 2-6 shows the effect of TE and TR on signal intensity for brain tissues. Different protocols can be obtained by determining these times.

Table 2-1 T1 and T2 times for the brain tissues at magnetic field of $B_0$ = 1 Tesla.

| **Tissue** | **T1 (ms)** | **T2 (ms)** |
|---|---|---|
| Water | 2500 | 2500 |
| Grey matter | 900 | 90 |
| White matter | 500 | 100 |
| Fat | 200 | 100 |
| CSF | 2000 | 300 |

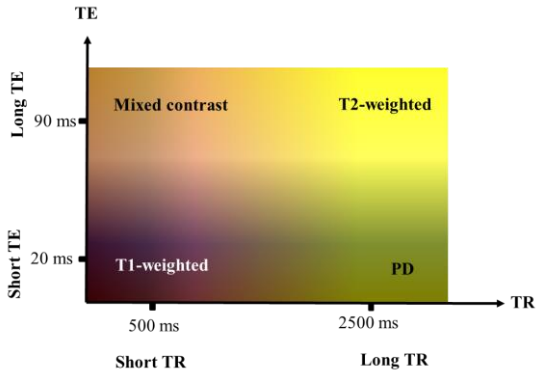
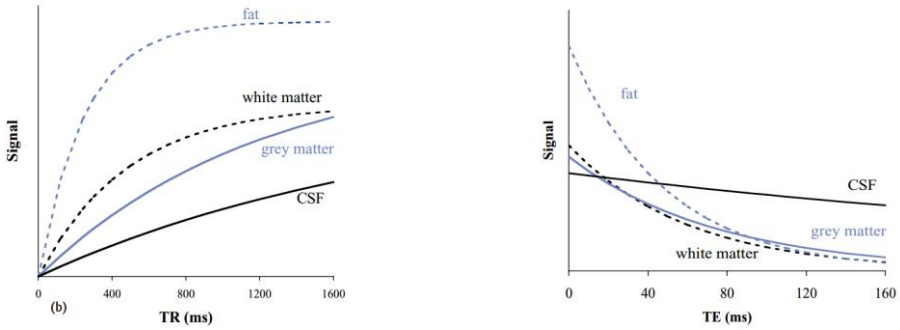



Figure 2-5 Weighting of MR imaging modalities. Mixed contrast is not used The image is recreated from ("MRI Signal weighting (T1, T2, PD) and sequences parameters," n.d.).




a b

Figure 2-6 Signal intensity of brain tissues against: a) repetition time (TR), b) echo time (TE). The plots are from the Reference (McRobbie *et al.*, 2006).

### 2.3.6 Conventional MRI Protocols

The conventional MRI protocols are generated using the tissue properties and TE and TR which were explained in Section 2.3.5 and Figure 2-5. The following sections will explain the C-MRI acquisition details.

#### T1-weighted

T1-weighted protocol is obtained when both TE and TR are short. This protocol usually provides excellent contrast between fluids, water-based tissues and fat-based tissues. In the case of brain images, it presents good contrast between GM and WM. In the brain tumour microstructures, the nuclei spins are influenced greatly by relaxation after 90 degree RF excitation. For the tissues without those microstructures, such as free fluids, the relaxation is slower. By selecting a low TR value (around 500 ms), three tissue types, i.e. CSF, brain tissue and fat will be clearly separable. In a T1-weighted image, the CSF appears hypointense, while

the brain tissues (GM and WM) appear with medium intensity and fat has considerable hyperintense values. Contrast agents can be injected to the patients to help increase the specificity by providing an extra set of images with different contrast which will be explained in the following section.

**Contrast-enhanced T1-weighted**

The contrast of the T1-weighted image can be improved by applying a contrast agent which is based on low molecular weight such as gadolinium (Gd). The electron configuration of the $Gd^{3+}$ ion makes the molecule highly paramagnetic, and therefore enhances the relaxation. When the water molecules interact with the Gd ion, the T1 value is shorter which results in hyperintense appearance in the T1-weighted images.

The idea of using contrast-enhancing agents for brain tumour imaging is that their molecules do not pass the blood vessel barriers inside the healthy brain tissues and remain inside the vessels. Malignant primary brain tumours cause damage to the blood brain barriers, thus the contrast agent comes out of vessels and leaks into the tissue space. The affected region will have shorter T1 which results in outstanding hyperintensity appearance on the T1-weighted images.

The evaluation of both contrast-enhanced and non-contrast T1-weighted images is very important to distinguish between the lesions with bright T1 properties, i.e. fat, blood, and protein-containing lesions, from the contrast-enhanced ones (Ginat and Meyers, 2012). Contrast-enhanced T1-weighted protocol will be referred as T1-contrast in the following sections.

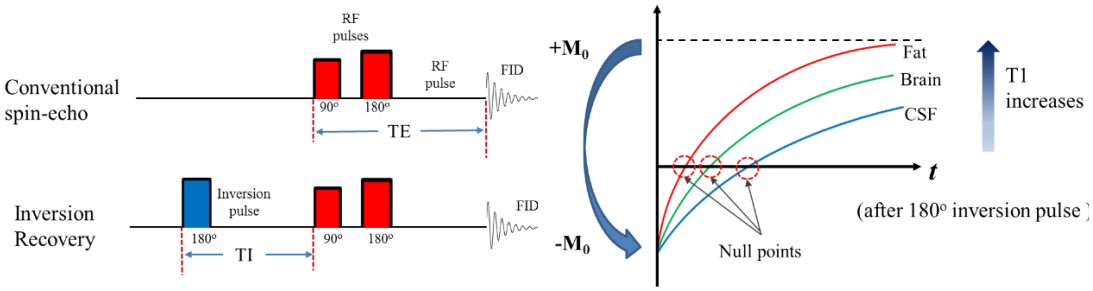
**T2-weighted**

T2-weighted protocol is obtained when both TE and TR are long. Regarding the T2 values of brain tissues, this protocol provides a good distinction between those brain parts, i.e. GM, WM, CSF, and scalp fat (Roberts and Mikulis, 2007).

In the fluid spaces of tissues with higher mobility, the spin interactions occur in shorter times with slower loss of transverse coherence which leads to longer T2 time. On the other hand, for more constrained structures such as dense population of cells, the spin interactions are longer with faster loss of transverse coherence, which leads to shorter T2 time. Generally, in T2-weighted images, fluids appear with high intensity, whilst water- and fat-based tissues are mid-grey. Most of the tumours damage the microstructures of brain which prolong the T2 values of the affected tissue.

**FLAIR**

A FLAIR image is created by applying a 180° RF pulse instead of 90° one, to increase the dynamic range of T1-weighted images (Figure 2-7). The 180° pulse will fully invert the longitudinal magnetisation ($M_z$) of all the underlying tissue opposite the $B_0$, and therefore is called an "inverting pulse". Therefore, $M_z$ starts at a negative initial value and passes the zero value, which is also known as "null point", and occurs at the time $t_{null} = 0.69$ T1. A standard imaging sequence starting with a 90° RF excitation pulse will be applied at the time after the inversion pulse which is known as the recovery time (TI). The sequence of these pulses is shown in Figure 2-7-(a). At a specific time *t*, the different tissue intensities will be scaled by the value of $M_z$. By fixing the *t* to the null point of a specific tissue, e.g. CSF, the signal intensity of that tissue will be "suppressed" or scaled to zero. Since CSF has the longest T1, the FLAIR image will have the properties of a T2-weighted image while improving the separation of lesions with similar properties to CSF in a T2-weighted image. In other words, FLAIR protocol produces a T2-weighted image with a suppressed CSF signal, which makes it the most commonly protocol used for diagnosis of the non-enhancing lesions in brain tumours. TI controls the extent of separation in T1 (Figure 2-7-(b)) and hence the image contrast.




a b

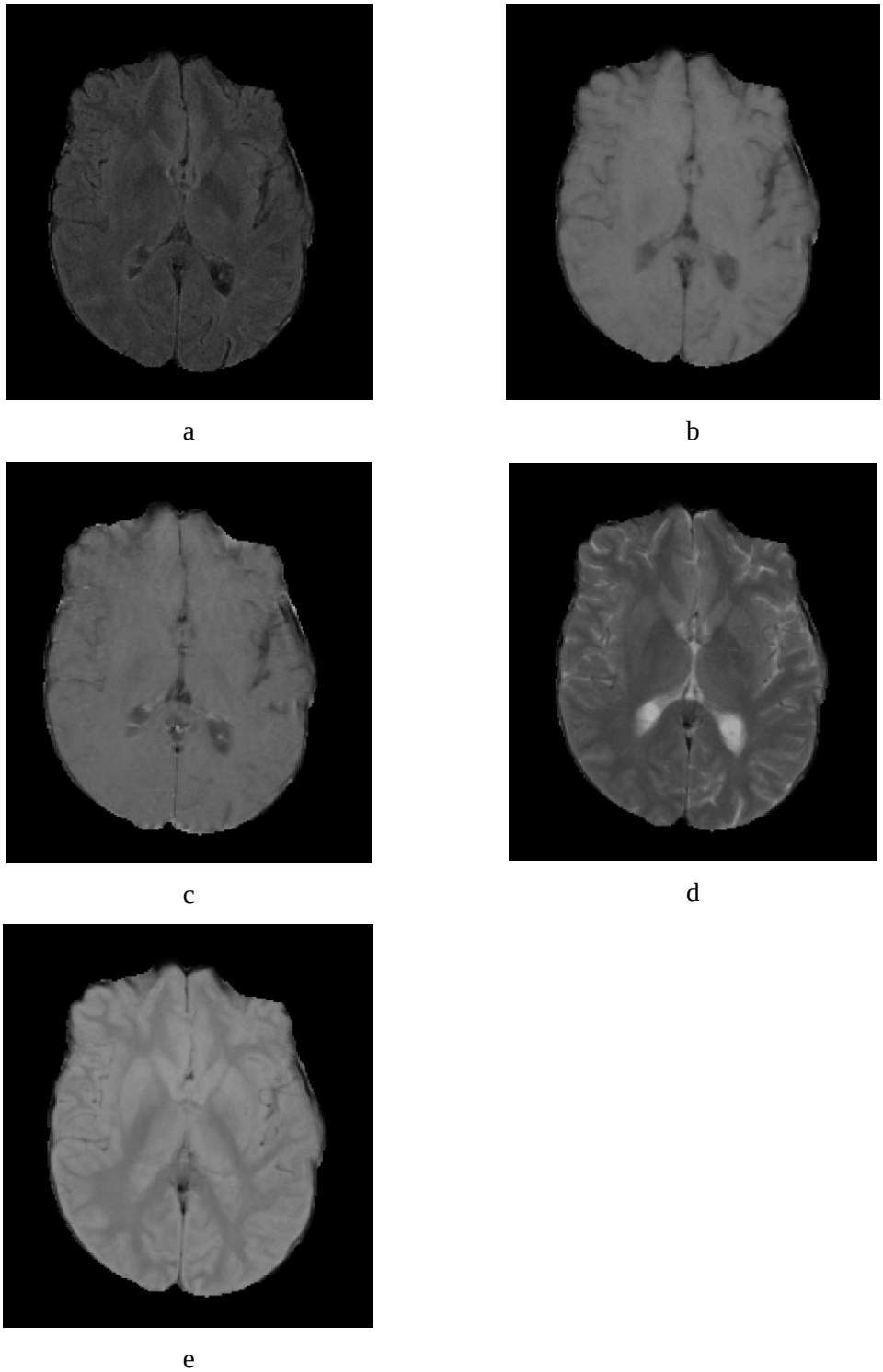
Figure 2-7 Illustration of FLAIR imaging pulses and T1 recovery. a) the echo-pulse sequence of inversion recovery, b) The longitudinal magnetisation and its effect on the T1 recovery of different tissues. The images are recreated from ("Inversion recovery," n.d.).

**Proton Density**

Proton density (PD) is obtained when TE is short and repetition time TR is long. This provides images based on minimising the impact of T1 and T2 differences. Therefore, PD can be considered as an intermediate sequence between T1- and T2-weighted which has less contrast compared to both protocols. The tissues with higher concentration of protons appear with

higher intensities in the image. For the case of brain imaging, the PD is not common nowadays and FLAIR is used instead (Simha *et al.*, 2012).

Figure 2-8 shows different C-MRI images acquired for a normal brain.



a b c d e

Figure 2-8 Example of C-MRI images for normal brain: a) FLAIR, b) T1-weighted, c) T2-weighted, d) T1-contrast, and e) PD.

### 2.3.7 Limitations of Conventional MRI

Conventional MRI modalities have two main limitations which are related to the spatial resolutions and contrast. During MR measurement time, which is 10-100 ms, the movement of water molecules is about 10 μm, which is considered as MR image resolution. This is not achievable in practice since water signal from this resolution is weak so cannot be distinguished from noise. A very long scanning time is required in order to detect such a weak signal and is not applicable in biological studies which restrict the actual resolution.

The signal intensity of MRI is a greyscale value for each voxel which is acquired from the detected signal from protons of water molecule. Considering two separate anatomical regions with different properties, if they appear in one voxel, they will not be distinguishable with this contrast limitation of conventional MRI protocols, analysis of the related biological interpretations will be difficult.

## 2.4 Diffusion MR Imaging

To overcome the limitations of C-MRI, several advanced MR modalities have been developed such as diffusion, functional MRI (fMRI), and spectroscopy. Diffusion imaging modalities are of the most important advanced MRI techniques in neuroradiology (Huisman, 2010), which will also be investigated in this study alongside with conventional MR techniques. The most important types of diffusion MRI are diffusion weighted (DWI) and diffusion tensor imaging (DTI). The following sections are dedicated to description of the concept of diffusion and how MR images are generated based on the diffusion of water molecules in the brain structures.

### 2.4.1 Magnetic Resonance Diffusion

Diffusion is defined as the result of microscopic random thermal movements, which are also known as Brownian motion. The principle of diffusion imaging is based on the diffusion of water molecules in the brain which is determined by several factors such as type of tissue, temperature and microenvironmental structure of the surrounding area (Huisman, 2010). The magnitude and direction of diffusion of a water molecule is determined by the geometry of the environment which includes the water mass. In the case of brain microstructure, this is useful since they have very small size boundaries which can be determined by calculating the water tensor. These boundaries cannot be identified by conventional MRI.

Diffusion has two different types: isotropic and anisotropic. The displacement of molecules can be modelled by a sphere when the water molecules move randomly in all directions. The

diffusion can be described by a single value, *D*, which is known as the isotropic constant (Figure 2-9 (a)). In the case the molecules are restricted by their surrounding environment, their diffusion occurs along axes (anisotropic). Therefore, the molecular displacements can be model by an ellipsoid (Figure 2-9 (b)).

Diffusion-weighted (DW) sequences are generated by adding two magnetic field gradients while varying the magnetic field linearly across the tissue. The precessional frequency is related to the magnetic field strength, therefore the gradient will impose a precessional frequency which is position dependent. The spins acquire a phase while they are precessing over a duration of time. By applying a gradient with the same size and duration but with 180$^{o}$ (refocusing RF pulse), the phase will be reversed. Figure 2-10 shows a diffusion pulse sequence which is used to detect the diffusion signal.

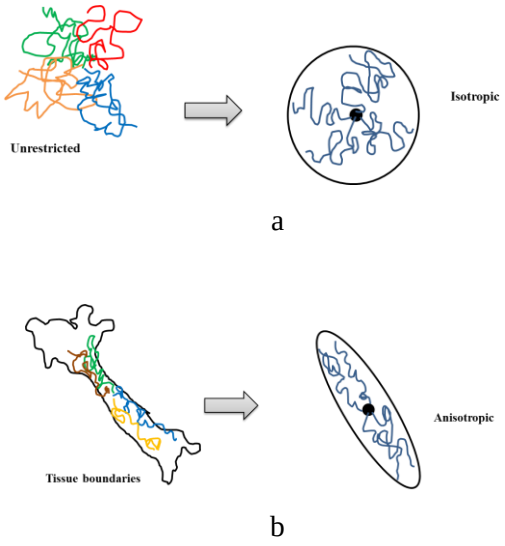



Figure 2-9 Schematic illustration for isotropic and anisotropic diffusion of water molecule in: a) the free space (isotropic), and b) restricted to tissue (anisotropic)

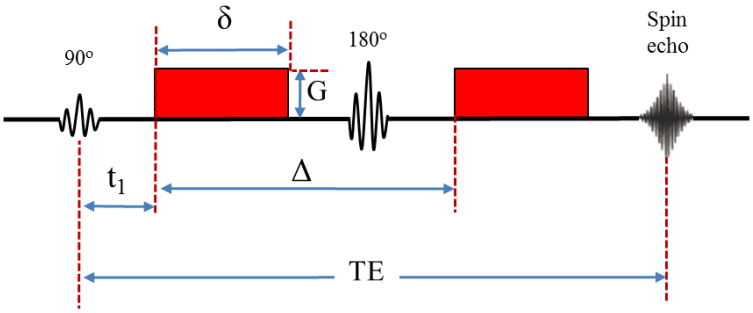



Figure 2-10 Spin-echo sequence in diffusion weighted imaging. The shaded rectangles show the gradient pulses which induce (left block) and reverse (right block) the phase shift. The image is recreated from (Winston, 2012).

Phase depression depends on the strength ($G$), duration ($\delta$), and time intervals ($\Delta$) of the gradient field which are presented in Figure 2-10. The measure of sensation to diffusion, *b*-value (in s/mm$^2$), is calculated using the diffusion gradient characteristics and the Stejskal-Tanner equation

$$b = (\gamma G \delta)^2 (\Delta - \frac{\delta}{3}) \tag{2-1}$$

where $\gamma$ is the gyromagnetic ratio. The diffusion is measured by comparing the voxel signal intensity between low and high DW which corresponds to low and high *b*-values, respectively. The following section will explain the details of diffusion-weighted imaging.

### 2.4.2 Diffusion Weighted Imaging

The diffusion images are generated by comparing the images acquired with low and high *b*-values, which are often 0 and 1000 s/mm$^2$, respectively (Huisman, 2010). The image acquired with $b$ = 0 s/mm$^2$ is not sensitised for diffusion and is termed $S_0$. This image is equivalent to a T2-weighted image without diffusion-weighting. The DW image, $S_b$, is acquired with a known *b*-value (e.g. $b$ = 1000 s/mm$^2$) and the same TE. The diffusion constant, $D$, of a voxel is then calculated using

$$\frac{S_b}{S_0} = e^{-bD} \tag{2-2}$$

The calculated value of $D$ in Equation (2-2) corresponds to diffusion in one direction. By measuring the diffusion in several directions, the signal loss versus $S_0$ is measured and reflects the displacement in the corresponding direction of gradient.

### 2.4.3 Imaging Using the Diffusion Tensor

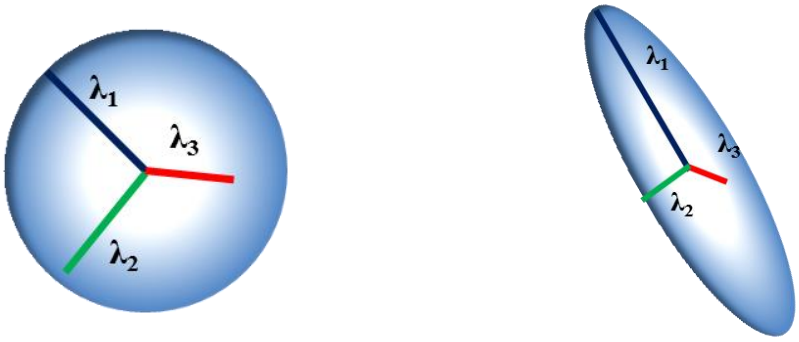
A tensor of size 3 × 3 is used to characterise the diffusion ellipsoid by modelling the coefficient along different directions and is called the diffusion tensor, $\boldsymbol{D}$.

$$\boldsymbol{D}_{i,j} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}. \tag{2-3}$$

The diffusion tenor is symmetric with diagonal elements representing the mobility rate in each direction and non-diagonal elements representing the correlation between the orthogonal directions.

When the water molecules are moving freely, their motion will be random in all directions. Therefore, the displacement of the molecules is modelled by a sphere with a diameter which is determined by *D*. The centre of the sphere remains in the same position without moving.

The anisotropic diffusion ellipsoid is described by three principal axes which are perpendicular to each other which are also called eigenvectors. The magnitudes and the directions of these axes are determined by the corresponding eigenvalues and eigenvectors, respectively. Therefore, six measurements are required to describe the anisotropic ellipsoid which are $\lambda_{i\ i\epsilon\{1,2,3\}}$ and $\boldsymbol{v}_{i,\ i\epsilon\{1,2,3\}}$.




a b

Figure 2-11 Isotropic and anisotropic diffusion, their eigenvalues and corresponding directions. a) isotropic diffusion, and b) anisotropic diffusion.

### 2.4.4 DTI Measures

The diffusion parameters of the voxels are represented by 3D ellipsoids, which are difficult to be visualised for the whole brain. For the purpose of visualisation and to make them comprehendible by human observers, 2D scalar isotropic and noninotropic maps are used.

**Isotropic Diffusion Maps**

The magnitude of diffusion can be measured by the summation of the diagonal elements of *D*, or trace of the tensor

$$Tr\ (\boldsymbol{D}) = D_{xx} + D_{yy} + D_{zz}\ . \qquad (2\text{-}4)$$

*Tr(D)* is orientation invariant which means that it is not sensitive to the orientations of microstructures or cells.

The most basic measure of the diffusion is the mean diffusivity (MD) which is also known as the apparent diffusion coefficient. MD is calculated using

$$MD = \frac{tr\,(\boldsymbol{D})}{3}, \tag{2-5}$$

which can be considered equivalent to

$$MD = \frac{\lambda_1+\lambda_2+\lambda_3}{3}. \tag{2-6}$$

The map of isotropic diffusion provides a measure for the content of diffusivity without considering the directional properties of the diffusion.

**Anisotropic Diffusion Maps**

Measuring the diffusion anisotropy provides the directional characteristics of DW images. Several formulations have been proposed for measuring the diffusion anisotropy (Pierpaoli *et al.*, 1996). A simple representation of the anisotropy is the ratio of the longest and shortest axes of the ellipsoid ($\lambda_1/\lambda_3$) which shows the elongation of the ellipsoid. The most common measure for anisotropy is fractional anisotropy (FA) which is calculated using

$$FA = \sqrt{\frac{3}{2}}\,\frac{\sqrt{(\lambda_1-\lambda_2)^2+(\lambda_2-\lambda_3)^2+(\lambda_1-\lambda_3)^2}}{\sqrt{\lambda_1{}^2+\lambda_2{}^2+\lambda_3{}^2}}. \tag{2-7}$$

which is a scalar value in the range [0, 1]. The zero value represents pure isotropic while 1 is related to pure anisotropic diffusion.

The scalar map of the whole brain can be generated by calculating MD and FA for each voxel. Therefore, the diffusion images are visualised similar to C-MRI with greyscale 2D slices.

### 2.4.5 Decomposition of Tensor

A tensor has nine elements in 3D space. To reduce the dimensionality of tensor, transformation is used to create a single scalar value. Also, more detailed measure of the tensor can be extracted by decomposing the diffusion tensor into its components, i.e. isotropic and anisotropic. An advanced decomposition technique for visualising more scalar measures for the tensor is proposed by Peña *et al.* (Peña *et al.*, 2006) which will be explained in the following. The tensor in Equation (2-3) is firstly decomposed using

$$D_{ij} = DI_{ij} + [D_{ij} - DI_{ij}], \tag{2-8}$$

where $I_{ij}$ is the identity tensor. Regarding to Equation (2-8), the diffusion tensor is decomposed into two separate tensors. By considering the components as *P* and *Q* terms, Equation (2-8) can be rewritten as

$$D_{ij} = P_{ij} + Q_{ij} \,, \tag{2-9}$$

where *P* and *Q* represents the isotropic tensor and the deviatoric tensor, respectively. In other word, the decomposed tensors are

$$P = DI_{ij} \,, \tag{2-10}$$

$$Q = D_{ij} - DI_{ij} \,. \tag{2-11}$$

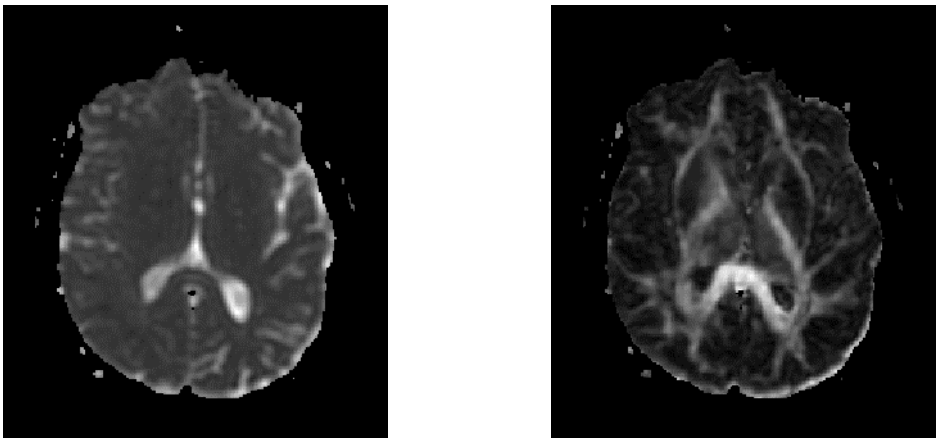
The magnitude of these tensors represents the isotropic ($p$) and anisotropic ($q$) components of the tensor which are calculated using

$$p = \sqrt{3}MD \,, \tag{2-12}$$

$$q = \sqrt{(\lambda_1 - MD)^2 + (\lambda_2 - MD)^2 + (\lambda_3 - MD)^2} \,, \tag{2-13}$$

where *MD* is calculated from Equation (2-6).

Figure 2-12 shows DTI protocols (i.e. *p*- and *q*-map) for the same patient which was shown in Section 2.3.6 (Figure 2-8).



a b

Figure 2-12 MRI different protocols: a) *p*-map and b) *q*-map.

## 2.5 Brain Tumours

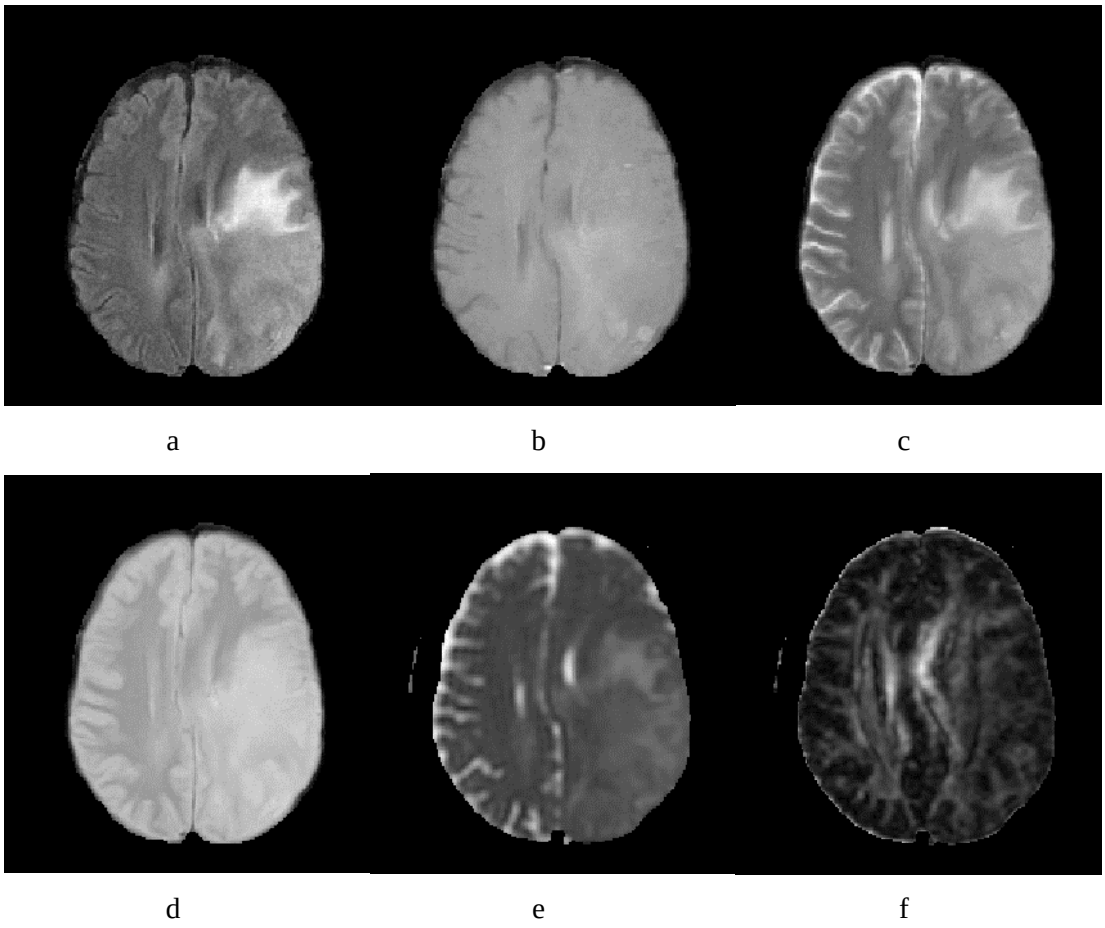
Brain tumours are caused by abnormal and uncontrolled growth of the cells inside the brain or spinal canal. The primary tumours are those started in the brain and are categorised in four main types: Gliomas, Meningioma, Pituitary adenomas and Nerve sheath tumours. The most

popular grading system for tumours is that suggested by the World Health Organisation (WHO). Regarding to the WHO grading system, the tumours are graded from I to IV, corresponding to least advanced to the most advanced diseases, respectively (Louis *et al.*, 2007). The most recent WHO classification of tumours was presented in 2016 (Louis *et al.*, 2016) which is the revised version of the 4$^{th}$ edition (Louis *et al.*, 2007). Most brain tumours cause oedema that is a swelling around the tumour and occurs when fluid enters the brain tissue (Papadopoulos *et al.*, 2004).

### 2.5.1 Low-grade Glioma

Low-grade glioma (LGG) are related to the primary central nervous system (CNS) neoplasm which is considered as the origin of the glial cell. They can be considered in WHO grade I and II categories which are less developed and tend to be more benign and have better prognosis. LGG includes many types of tumour with their own specific characteristics and treatments (Perry, 2003). The most common types of LGG tumours are astrocytomas, oligodendrogliomas and oligoastrocytomas which have more invasive and malignant properties (Cavaliere *et al.*, 2005).

LGG tumours have a wide range of imaging features which are related to their histological categories. LGG are hyper-intense on T2-weighted images and may appear with diffuse indistinguishable boundaries or focal shapes with clear borders. In this kind of tumour, a cyst may appear in the image and usually they have little oedema. Contrast imaging can enhance the appearance of 15-39% of LGG (Shaw *et al.*, 2002). Grade III or grade IV tumours may also appear as LGG with the error rate of up to 30% (Scott *et al.*, 2002). LGG are usually infiltrative which invade the normal brain tissue far from the tumour core in microscopic scale. Figure 2-13 shows examples of LGG tumours in C-MRI and DTI protocols.



a b c

d e f

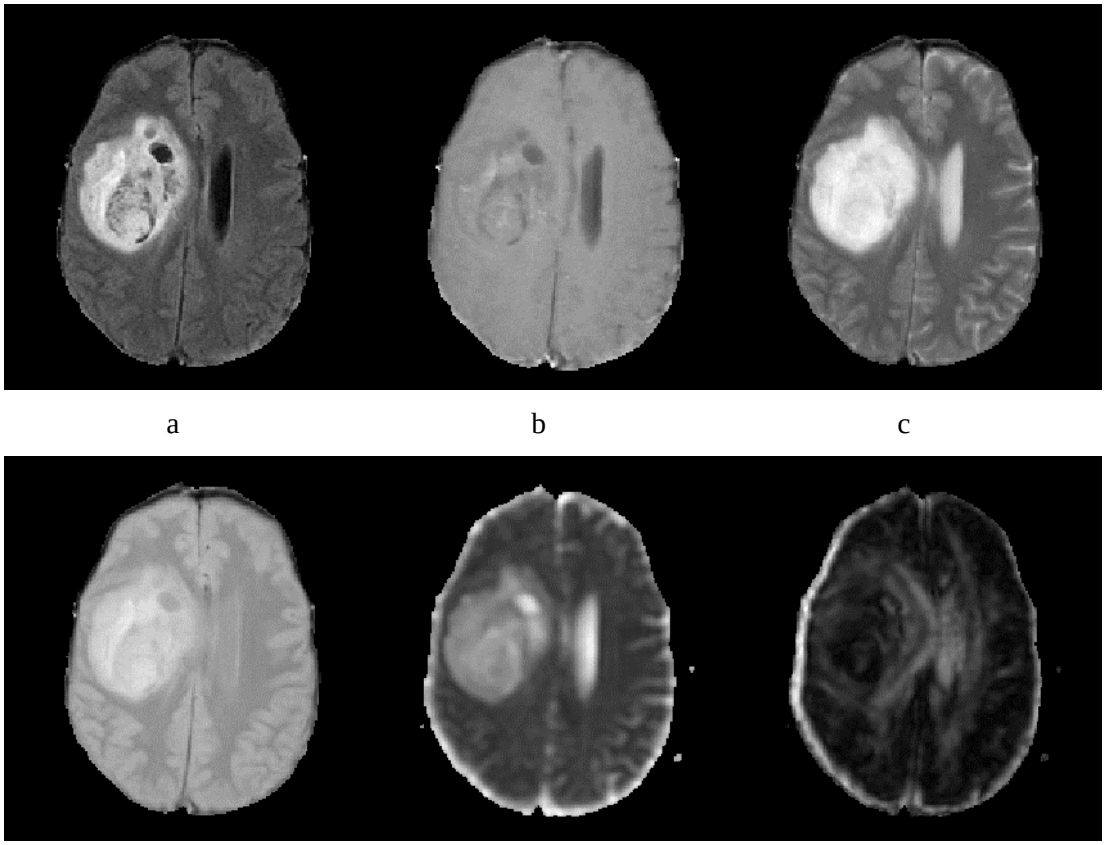
Figure 2-13 MRI images of low-grade glioma: a) FLAIR, b) T1-contrast, c) T2-weighted, d) PD, e) *p*-map, f) *q*-map.

### 2.5.2 High-grade Glioma

High-grade glioma (HGG) are the most aggressive type of glial tumours and tend to be malignant. They correspond to the WHO grade III and grade IV categories. The may be located in the cerebral hemisphere with infiltrate the sounding brain without causing substantial destruction. Therefore, the may cause enlargement of the affected brain structure. Some of them may have poorly distinguishable boundaries and their necrosis may include up to 80% of the tumour volume.

In the MRI images, the HGGs appear as a heterogeneous mass with a surrounding oedema. They often have a necrosis in the central region with extensive oedema. Necrosis which is related to cellular membrane breakdown and oedema which related to increasing in extracellular water reduce the restriction which results in increasing the diffusivity. Advanced MRI techniques such as DWI, DTI, and dynamic contrast-enhanced MRI are increasingly used for the clinical HGG related tasks (Young, 2007). Several studies (Peña *et al.*, 2006; Price *et al.*, 2007, 2004; Wang *et al.*, 2009) show that the metrics *p* and *q* can distinguish the

differences between tumour and normal brain tissues in terms of microstructural abnormalities related to tumour infiltration (Price *et al.*, 2004). Figure 2-14 shows examples of HGG tumours in C-MRI and DTI protocols.



a b c

d e f

Figure 2-14 MRI images of high-grade glioma: a) FLAIR, b) T1-contrast, c) T2-weighted, d) PD, e) *p*-map, f) *q*-map.

## 2.6 Datasets Used in the Thesis

Two datasets are used in this study to develop the algorithm and establish a comprehensive evaluation and comparison with other methods in the literature. The first dataset is the clinical data and the second set is publicly available data. The following subsections will describe the details of data acquisition, manual delineation and clinical evaluation protocols.

### 2.6.1 Clinical Dataset

The clinical dataset is obtained from St George's Hospital Trust, London. There are two reasons to use this dataset. Firstly, alongside with the conventional MRI, the clinical dataset contains DTI protocols which are not included in the common publicly available dataset. Secondly, the original image properties from the clinical acquisition are preserved, different

from the public dataset which is interpolated and resized. The following sections will describe the conventional and DTI dataset, respectively.

**C-MRI Clinical Dataset**

The clinical patient data were acquired using a GE Signa Horizon LX 1.5T MRI system (GE Healthcare, Milwaukee, WI, USA). The imaging machine was equipped with a maximum field gradient strength of 22 mT/m and using a quadrature head coil.

A cohort consisting of 19 patients were used, each with a brain tumour. The dataset consists of 6 grade II tumours, 3 grade III tumours, and 10 grade IV tumours. Patient ages at the time of scanning ranged from 22 to 73 (mean 54), and consisted of 7 females and 12 males.

The FLAIR sequence images were acquired in the axial plane with a field of view (FOV) 240 × 240 $mm^2$, matrix size 256 × 256 and 5 mm slice thickness with no slice gap. The acquisition parameters were TE= 133 ms, TR = 9000 ms, and TI = 2200 ms.

T1-weighted images were acquired in the axial plane with a field of view (FOV) 240 × 240 $mm^2$, matrix size 256 × 256 and 2.8 mm for T1 with no slice gap. The acquisition parameters were TE = 14 ms, TR = 600 ms, bandwidth = 122.1 Hz.

T1-contrast images were acquired with 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 × 220 $mm^2$ or 240 × 240 $mm^2$, a 256 × 256 acquisition matrix, and 29 slices with 5 mm thickness (Jones *et al.*, 2012).

**DTI Clinical Dataset**

DTI data were acquired using a diffusion-weighted spin-echo echo-planar imaging sequence. As explained in Section 2.4.2, $b_0$ acquisition was made without diffusion gradients (i.e. $b$=0 s/$mm^2$). Diffusion weighted images were acquired using $b$=1000 s/$mm^2$ with 12 gradient directions (Barrick and Clark, 2004). The FOV was 240 × 240 $mm^2$ with a 96 × 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 × 256 matrix. The diffusion parameters $p$ and $q$ for isotropic and anisotropic diffusion respectively were calculated as proposed by Peña *et al.* (Peña *et al.*, 2006) and explained in Section 2.4.5.

The multimodal dataset containing DTI protocols consists of 11 brain tumour patients from which 2 are grade III, and 9 are grade IV). Patient ages at the time of scanning ranged from 33 to 73 years (mean age 53 and standard deviation 7).

**Clinical Ground Truth**

The ground truths (GT) are provided by a trained human expert. The annotation protocol and the corresponding GT labels are as follows

0- Healthy brain tissues and background

1- Oedema, which boundaries are drawn using the FLAIR images. Oedema appears hyper-intense in the FLAIR images.

2- Tumour core, which boundaries are drawn using T1-contrast images. Enhancing tumour appears hyper-intense and necrosis appears with low intensity in T1-contrast images.

Figure 2-15 shows sample of tumour tissues, which are overlapped on FLAIR and *p*-map protocols.

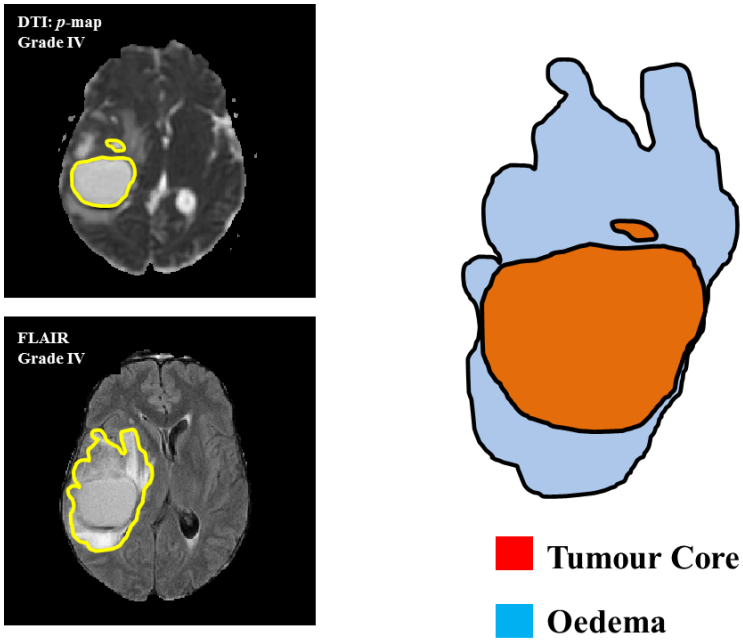



Figure 2-15 Brain tumour tissues (oedema and core) from the clinical dataset. Left) manual ground truth overlaid separately for each tissue on FLAIR (oedema) and DTI *p*-map (core) protocols. Right: the schematic illustration of the tissues.

### 2.6.2 MICCAI-BRATS Dataset

The publicly available MICCAI-BRATS 2013 dataset, which was provided by Virtual Skeleton Database (VSD), has provided a standard and unique evaluation procedure to enable the researchers and academics to fairly compare the methods and also interpret the comparison of different methods. The first BRATS dataset was released in 2012. The training dataset for both BRATS 2012 and 2013 are the same, while their challenge datasets are different. Therefore, if an experiment is performed on any of the BRATS 2012 or 2013 training dataset, it can be compared to the other one. The most recent MICCAI-BRATS dataset was released in 2017 (Bakas *et al.*, 2017a; Bakas *et al.*, 2017b, Bakas *et al.*, 2017c). The most accurate method proposed in this thesis (Chapter 6) will then be evaluated on BRATS 2017 challenge dataset. It should be noted that the BRATS 2015 was not used in this thesis for the reason that the corresponding ground truth was provided by interpolation of the selected manually annotated slices. Since the clinical dataset was segmented manually for all slices, it was deemed difficult to conduct a fair comparison.

**C-MRI Public Dataset**

The total patient dataset, which were used from BRATS 2013 challenge, consists of 40 multi-contrast MR scans of glioma patients. They are divided into training and testing datasets. For those training set, the ground-truths are provided by a trained human expert (Menze *et al.*, 2015). The annotated training dataset consists of 30 patient MRI scans of which 20 are HGG and 10 are LGG. The test dataset consists of 10 cases with LGG. For each patient data, T1, T2, FLAIR, and post-Gadolinium T1 MR images are available. Data were acquired from multi-centres and using different scanners with different field strengths (1.5 T and 3T). BRATS 2017 dataset, consists of 285 training (including 210 HGG and 75 LGG) and 46 validation patient cases.

The dataset has been already skull-removed, registered and interpolated by the BRATS challenge organisers.

**BRATS Ground Truth**

The ground truths provided by VSD system are four tumour structures which are annotated by experts which are labelled as follows: oedema, necrosis non-enhancing tumour, enhancing tumour, other tissues and background.

The BRATS challenge dataset was annotated by a team of 5 trained radiologists who drew outlines of the tumour structures in the axial slices. 3D slicer software (“3D Slicer,” 2017) was used to perform the segmentation. They segmented every third slice and the outlines were interpolated using region growing and a final visual correction was carried out by the experts. The annotation from the experts were then fused together to obtain a single segmentation and

decrease the inter-observer error. The annotation fusing was accomplished by a voting approach in which the highest class was assigned (i.e. the class that at least 50% of the observers agree on). The 30 training dataset was segmented by four observers and the 2013 test set was segmented by one observer.

The protocol for manual annotation of tumour structures in BRATS challenge is described below:

1) Oedema was segmented from T2 and FLAIR images. T2 was used for the initial segmentation and then FLAIR was used to check the extension of oedema and discriminate it from other tissues such as necrosis and ventricles.

2) The complete tumour core which includes all three tumour structures was segmented using hyper-intensity regions in T1-contrast with the heterogeneous region of hyper-intense and hypo-intense lesion in T1.

3) The enhancing core of tumour was segmented using T1-contrast using thresholding on the complete tumour core region segmented in part 2 which results in keeping the Gadolinium and subtracting the necrosis tissue. The threshold levels were determined individually for each case by visual inspection.

4) Necrosis or fluid-filled core was segmented using T1-contrast in the low-intensity structures inside the enhancing tumour which was segmented in part 3. For the rare haemorrhages the same label was considered.

5) Non-enhancing core was considered as the remaining part of the complete tumour core which were not detected in parts 3 and 4, therefore they were obtained by subtracting the corresponding regions from the complete tumour.

## 2.7 Evaluation Protocol

The clinical dataset was evaluated using binary- and multi-label classification as follows: (1) “complete tumour” which includes oedema and tumour core, and (2) Tumour core.

The VSD evaluation system categorised and grouped the tumour structures into three non-overlapped regions that are suggested to better represent the clinical application tasks (Menze *et al.*, 2015). The tumour regions in BRATS 2013 and 2017 challenges are as follows: (1) “complete tumour” which includes all the four structures, (2) “tumour core” which includes necrosis, enhancing and non-enhancing, and (3) “active tumour” which includes the enhancing

tumour only and is unique to HGG patients. It should be noted that in the BRATS 2017 dataset, the enhancing and the necrosis were merged together to form one class label.

Evaluation and comparison with the state-of-the-art methods is an important part of any newly proposed research work. The evaluation approach in the field of brain tumour segmentation is comparing to the ground-truths which are provided by the expert.

### 2.7.1 Evaluation of Classification

Regarding the true class and estimated class labels, the following categories can be considered:

- TP: Number of abnormal data classified correctly as abnormal.
- TN: Number of normal data classified correctly as normal.
- FP: Number of normal data classified incorrectly as abnormal.
- FN: Number of abnormal data classified incorrectly as normal.

Table 2-2 illustrates the above-mentioned classification categories.

Table 2-2 The classification evaluation categories.

| | | **Predicted Class** | |
|---|---|---|---|
| | | Normal | Abnormal |
| **True Class** | Normal | True Positive (TP) | False Negative (FN) |
| | Abnormal | False Positive (FP) | True Negative (TN) |

For assessment of the classification, the following standard five measures, i.e. accuracy, precision, sensitivity, specificity, and balanced error rate (BER), are calculated using

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2-14}$$

$$Precision = \frac{TP}{TP+FP}, \tag{2-15}$$

$$Sensitivity = \frac{TP}{TP+FN}, \tag{2-16}$$

$$Specificity = \frac{TN}{TN+FP}, \tag{2-17}$$

$$BER = 1 - 0.5 \times \frac{\text{Sensitivity+Specificity}}{100}, \tag{2-18}$$

### 2.7.2 Evaluation of Segmentation

Several studies have used different combinations of tumour parts for evaluation of their methods. This makes it difficult to compare different methods which used different tumour parts. The most common tumour parts are provided by the BRATS challenge organisers. The methods which used their standard can be compared fairly and more accurately. It should be noted that, in the specific case of BRATS challenge dataset, the segmented masks were uploaded to the VSD website and evaluated by the corresponding online system (blind test).

Several methods have been used for evaluation of segmentation results, including Dice similarity score (DSC), sensitivity, and positive predictive value (PPV). Figure 2-16 shows the segmented regions which are used for evaluation of the segmentation method.

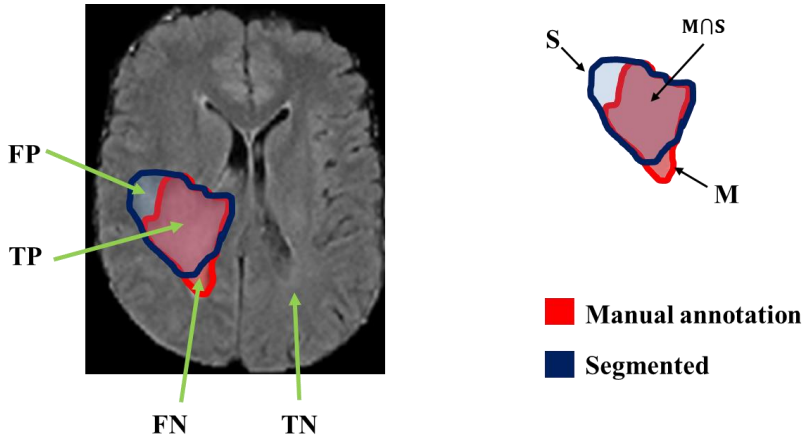



Figure 2-16 Schematic illustration of the regions which are used for evaluation of the segmentation, i.e. Dice score, PPV and sensitivity. Red boundaries represent the manual annotation (ground truth) and the blue areas represent the boundaries of the segmented region using the automated methods.

DSC (Dice, 1945) is the common overlap measure for evaluation of the segmentation, which is also the main evaluation metric in BRATS challenge (Menze *et al.*, 2015). DSC represents the overlap between the manual and segmented volumes/masks which is defined as

$$DSC = \frac{2TP}{FP+2TP+FN}. \tag{2-19}$$

DSC can also be calculated by

$$DSC = \frac{2|M \cap S|}{|M|+|S|} \tag{2-20}$$

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.

PPV is defined as

$$PPV = \frac{TP}{TP+FP}, \tag{2-21}$$

which is a measure of the number of false positives and true positives.

## 2.8 Clinical Expectations

The results of an automatic segmentation method need to be accurate and reproducible for measuring the tumour size and its changes over time. Volumetry of the tumour is very important for clinical tasks such as diagnosis, treatment planning and patient monitoring. In order to measure an accurate volumetry, the operator should be able to outline the tumour and distinguish between different tumour tissue subtypes. This task requires a significant skill and expertise in tumour diagnostic and handling the corresponding computer tools (Meier *et al.*, 2016).

Manual tumour delineation is a time consuming task. With the current tools, the average time of manual segmentation is approximately 10 minutes per tumour (Egger *et al.*, 2013).

The agreement between the result of automatic segmentation and the manual ground truth is calculated by overlap measures such as Dice similarity score. It is difficult define a specific value to qualify the agreement criterion. However, Egger *et al.* (Egger *et al.*, 2013) suggested that a Dice score of 0.88 is an acceptable agreement between a segmentation method and clinical ground truth.

An accurate segmentation can be used to calculate the volumetric size and bidimensional measure of a tumour. Bidimensional measure is used for response assessment of high-grade glioma in current clinical guidelines (Chinot *et al.*, 2013).

## 2.9 Summary

This chapter provided the clinical background for brain tumour segmentation in MRI. The conventional and DTI MRI acquisition techniques were explained. Each modality provides a specific characteristic of the tissue which may be not apparent in other protocols. The application of different MR protocols in brain tumour segmentation were explained by emphasising the importance of DTI imaging techniques which provide more information about the microstructure of the brain tissues. Clinical and publicly available datasets were described which will be used for the experiments followed by the evaluation protocols. The next chapter will explain and discuss the related work in the field of brain tumour segmentation using different MRI modalities.

# Chapter 3

# Literature Review

## 3.1 Introduction

As discussed in Chapter 2, MRI is the most common diagnosis modality for brain tumour studies (Wen *et al.*, 2010), therefore MRI imaging modality research in this field will be reviewed. Different brain tumour types were explained in Chapter 2. Gliomas are the most common primary brain tumour with the occurrence rate of 70% in adults with malignant tumours (Bauer *et al.*, 2013), which will be investigated in this thesis. There are a number of grand challenges provided for the task of brain tumour segmentation from MR images. The most popular challenge is BRATS ("MICCAI BRATS - The Multimodal Brain Tumor Segmentation Challenge," n.d.) in conjunction with MICCAI conference series (Menze *et al.*, 2015). The research work and the challenge trend will also be reviewed.

The aim of image segmentation is to partition the image into regions which in medical imaging may correspond to the tissue type, structure or function (Bauer *et al.*, 2013). Segmentation of a brain tumour in MRI is a challenging task due to the complicated tumour properties and inherent MR imaging characteristics. In terms of brain tumour structures, they usually appear with different texture and irregular boundaries. As discussed in Chapter 2, the tumour borders may be unclear and discontinuous due to infiltration. In terms of MR image characteristics, there are several limitations which make the segmentation a challenging task. For example, the contrast agent dosage and the acquisition time may affect the tumour appearance in the image and provide different intensities for the same tissue. The slice thickness ($Z$ direction) of image acquisition is usually higher than the in-plane resolutions ($X$ and $Y$ directions) which causes partial volume effects. To detect more detailed tumour structures, using different modalities is essential. In this case, the accurate registration of different modalities is very crucial and a challenging task which will affect the final segmentation results.

The MRI brain tumour segmentation techniques have been reviewed by Angelini *et al.* (Angelini *et al.*, 2007), Bauer *et al.* (Bauer *et al.*, 2013), and Gordillo *et al.* (Gordillo *et al.*, 2013). However, the studies are still growing and many improvements have been achieved recently owning to the significant advances in learning-based methods (Kamnitsas *et al.*, 2017) especially deep learning (Havaei *et al.*, 2017). Therefore, this chapter provides an overview of the state-of-the-art methods in the field especially the learning-based methods.

In the literature, different categorisation for the segmentation techniques are suggested. Bauer *et al.* divided the brain tumour segmentation algorithms into automatic and semi-automatic

approaches (Bauer *et al.*, 2013). Gordillo *et al.* (Gordillo *et al.*, 2013) categorised the tumour segmentation methods into supervised and unsupervised approaches. Menze *et al.* categorised the segmentation techniques into generative and discriminative models (Menze *et al.*, 2015). The following sections are organised mainly based on the categories suggested by Gordillo *et al.* (Gordillo *et al.*, 2013). However, the supervised techniques, i.e. the segmentation methods based on classical classifiers and deep learning will be explained and discussed in more detail.

## 3.2 Unsupervised Methods

Unsupervised segmentation techniques are used when the images are unlabelled. They can be performed based on features from anatomic or image-based objective measures (Gordillo *et al.*, 2013). The aim of unsupervised methods based on an anatomic objective measure is to segment the image into anatomical meaningful regions. The unsupervised methods based on image-based features are evaluating the regions which have similar image features, e.g. intensity, texture, etc. These methods are stronger in handling more complicated segmentation tasks, e.g. subparts of heterogenous tumours (Gordillo *et al.*, 2013). However, the lack of prior knowledge in brain tumour segmentation tasks makes them challenging for these type of methods (Popuri *et al.*, 2012). Some popular unsupervised methods, e.g. *k*-means and fuzzy *c*-means (FCM), will be explained in the following sections.

### 3.2.1 *K*-Means Clustering

The aim of *k*-means clustering is to partition a set of observations ($x_1$, $x_2$, …$x_N$) into *k* clusters $\mathbf{S} = \{S_1, S_2, \ldots, S_k\}$. *K*-means is an iterative algorithm, in which the centre of clusters ($c_1$, $c_2$, …, $c_k$) are updated to find the optimum clusters, and is initialised by randomly selecting *k* data points as initial cluster centres. The objective is choosing the centres so that to minimise the function (Hartigan and Wong, 1979)

$$\mathrm{J} = \sum_{j=1}^{k} \sum_{x_i \in S_j} \left\| x_i - c_j \right\|^2 \quad , \tag{3-1}$$

where the centre of clusters, $c_j$, is calculated from

$$c_j = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_i \quad . \tag{3-2}$$

In MRI brain images with unclear or infiltrative tumour boundaries, it is difficult to decide if a voxel belongs to which tissue class. To tackle this problem, integrating fuzzy concepts to the algorithms such as *k*-means may improve the segmentation results.

### 3.2.2 Fuzzy *C*-Means

Fuzzy *c*-means incorporates the fuzzy concept to *k*-means, in which each data point can belong to more than one cluster. The data points have sets of weights that indicate the degree of their belonging to the clusters. The weights are represented by $\boldsymbol{W} = w_{ij}$, where $w_{ij}$ is the degree of data point $x_i$ belonging to the cluster $j$ with the centroid $c_j$. The centroid is the mean of all points weighted by $w_{ij}$, and calculated using (Bezdek *et al.*, 1984)

$$c_j = \frac{\sum_{i=1}^{N} w_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} w_{ij}^{m}} , \tag{3-3}$$

where $m$ is the fuzzy partition exponent which controls the degree of fuzzy overlap ($m \geq 1$).

The objective function, which should be minimised, is calculated using

$$\mathrm{J}_{FCM} = \sum_{i=1}^{N} \sum_{j=1}^{C} w_{ij}^{m} \left\| x_i - c_j \right\| , \tag{3-4}$$

where $C$ is the number of clusters and $w_{ij}$ is calculated using

$$w_{ij} = \frac{1}{\sum_{k=1}^{c} \left( \frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}}} . \tag{3-5}$$

FCM is a popular unsupervised technique in the field of image processing, especially for brain tumour segmentation (Kong *et al.*, 2006). Firstly, a set of tissue classes of the brain and tumour should be predetermined. Then, a membership value is assigned to each voxel related to the tissue classes, based on the features such as intensity, texture, etc. The algorithm performance depends on the initialisation and selecting accurate cluster centres.

FCM clustering was firstly used by Phillips *et al.* (Phillips *et al.*, 1995) for tumour segmentation, and later was applied to multimodal MR images (Clark *et al.*, 1998). Later on, other researchers developed the FCM-based methods using priori knowledge (Fletcher-Heath *et al.*, 2001), defining spatial constraints (Hsieh *et al.*, 2011), etc. Some other methods combined FCM with other independent segmentation methods (Helen and Kamaraj, 2015; Rajendran, A. and Dhanasekaran, 2012) to improve the segmentation results. Despite its advantages, FCM is not efficient in segmenting the tumour in MR images. It is sensitive to noise and the number of clusters should be predefined by the user.

### 3.2.3 Challenges of Unsupervised Segmentation

Unsupervised methods have three main disadvantages (Gordillo *et al.*, 2013) which can be summarised as follows. Firstly, the number of clusters often should be predefined which results in a semi-automatic approach. Although some algorithms may automatically find the

number of classes during grouping the pixels, still it remains a challenging task. Secondly, if higher number of clusters are chosen, the tumour regions may be partitioned into several subparts. Thirdly, the lack of prior knowledge, e.g. shape or intensity, make the segmentation challenging. However, the unsupervised methods can be used for tumour detection rather than accurate tumour segmentation (Kamnitsas *et al.*, 2017).

## 3.3 Improving Segmentation with Probabilistic Approach

The standard clustering and classification methods consider the image pixel data as independent and identically distributed (IID). The images often have homogenous regions, in which the pixels have similar properties, e.g. intensity, texture, etc., which can be used as contextual constraints for the prediction model. Probabilistic approaches consider the label of one pixel being dependent on the probability of adjacent pixels. Although the probabilistic approaches are mentioned in this section, they also can be used to improve the supervised methods (Section 3.4).

### 3.3.1 Markov Random Fields

Markov random field (MRF) is a probabilistic approach which has been widely applied in medical image analysis (Lee *et al.*, 2005). The main point of using an MRF (Kindermann 1980) is that it takes into account the spatial information and neighbourhood dependencies between the voxels. It means that if a voxel is classified as one particular class, e.g. tumour or nontumour, then its neighbours have more probability to have the same label.

A set of observed image features $x = \{x_1, x_2, \ldots, x_N\}$ and the corresponding labels $y = \{y_1, y_2, \ldots, y_N\}$ are considered. The probability of the labelling, given the observed feature is defined as $p(y|x)$. The aim is to relabel the image and find new labels $\hat{y}$ that maximise this probability. This is called maximum a posteriori (MAP) estimate

$$\hat{y} = \arg\max_{y} p(y|x) . \tag{3-6}$$

Bayes' theorem is written as

$$p(y|x) \propto p(x|y)p(y), \tag{3-7}$$

where the first term is likelihood and the second term is a prior. Assuming that the probabilities have a factorised form, then the term $p(x|y)$ will have the following form

$$p(x|y) = \prod_{i} p(x_i|y_i) . \tag{3-8}$$

The maximal subgraphs of pixel connections that are fully connected are called cliques. The potential function of clique *c* can be written as $\psi_c(y_c|\theta_c)$. This can be any non-negative function. The joint distribution is proportional to the product of the clique potentials.

The Hammersley-Clifford theory states that a strict positive distribution (e.g. *p(y)* > 0) with the Markov properties can be factorised per cliques (Murphy, 2012)

$$p(y|\theta) = \frac{1}{Z(\theta)} \prod_{c \in C} \psi_c(y_c|\theta_c) \,, \tag{3-9}$$

where *C* is the set of all the maximal cliques, and *Z(*θ*)* is the partition function which is defined by

$$Z(\theta) = \sum_{y} \prod_{c \in C} \psi_c(y_c|\theta_c) \,. \tag{3-10}$$

A random field is an MRF when it follows the Gibbs distribution which is

$$p(y|\theta) = \frac{1}{Z(\theta)} \exp(-\sum_{c} \mathrm{E}(y_c|\theta_c)) \,, \tag{3-11}$$

where $E(y_c)$ is the energy of variables in the clique *c*. The potential function now can be written as

$$\psi_c(y_c|\theta_c) = \exp(-E(y_c|\theta_c)) \,. \tag{3-12}$$

The advantage of considering spatial information with MRF is reducing the effect of noise and the clusters overlap on the segmentation results (Gordillo *et al.*, 2013; Tran *et al.*, 2005). They also can represent complex dependencies between data, which provides higher accuracy in the field of brain tumour segmentation (Lee *et al.*, 2005). Capelle *et al.* (Capelle *et al.*, 2000) suggested that modelling the brain tissues with Gaussian mixture models followed by an MRF removes the need for postprocessing stages, e.g. morphological operations. Gering *et al.* (Gering *et al.*, 2002) developed a method to build a multilayer framework of MRF in which the layers consist of information from voxel intensity, spatial coordinates, structural coherence and user input. Doyle *et al.* (Doyle *et al.*, 2013) applied hidden MRF to the BRATS 2013 challenge dataset and obtained good results which will be mentioned in Section 3.6 (Table 3-1).

### 3.3.2 Conditional Random Fields

Conditional random field (CRF) is discriminative alternative and a variant of MRF (Lafferty *et al.*, 2001), which also takes the context of neighbouring observation into account. In CRF, the clique likelihoods are conditioned on the features of the input (Murphy, 2012)

$$p(y|x,w) = \frac{1}{Z(x,w)} \prod_c \psi_c(y_c|x,w) \,. \tag{3-13}$$

The likelihoods are often assumed having a log-linear form

$$\psi_c(y_c|x,w) = \exp(w_c^T f(x,y_c)) \,, \tag{3-14}$$

where $f(x,y_c)$ is the feature vector based on the input and the local set of the clusters $y_c$.

Therefore, it could avoid explicitly modelling the distribution over the observation. In the case of images with complex structure, such as brain tumour tissues with complex anatomic shapes, such a distribution is normally not easy to be modelled. CRF allows a flexible modelling of complex dependencies between features of voxel and its label and between the labels of the adjacent elements.

MRF and CRF have been widely used for classification, texture feature extraction and modelling the image intensity inhomogeneities (Gordillo *et al.*, 2013).

## 3.4 Supervised Methods

Supervised approaches are based on extraction of the features from the images and creating models based on the relationship between the image features and the pixel/voxel classes. A vast variety of features can be used, such as voxel intensity, histogram, and texture of the images. Using hand-designed features is common for the task of brain tumour segmentation based on classifiers.

In general, the supervised classification procedure is comprised of two main stages, i.e. training and testing. In the training stage, the model learns from the feature vector that is fed to the classifier to discriminate between different tissue classes. The model also adjusts its internal parameters or weights. In the testing stage, the unlabelled dataset, which is also called the "testing dataset", is fed to the trained model, then the MR image voxels are assigned to one of the classes.

The source of the training and testing dataset is a very important factor which affects the performance of the classifier, and can be categorised into patient-specific and inter-patient (Gordillo *et al.*, 2013). In patient-specific approaches, the model is trained based on the

labelled slices from one patient and the testing is performed on the unlabelled slices from the same patient. The inter-patient approach consists of training the model on labelled images from several patients and testing the trained classifier on a different patient from the testing dataset but for the same tumour type (Cobzas *et al.*, 2007). This approach is the common method in the brain tumour segmentation in MR images and the relevant challenges, such as BRATS challenge (Menze *et al.*, 2015).

### 3.4.1 Support Vector Machines

SVMs (Schölkopf and Smola, 2002) are efficient powerful classifiers for IID data, which have been used in many segmentation applications. The SVM finds the best hyperplane for separation of the classes, which presents the largest margin between them. The margin is defined as the maximum distance, in which in the best-case scenario, the data points of the different classes are separable in the feature space. The hyperplane for dimension *d* can be written as

$$\boldsymbol{W}\boldsymbol{X} - b = 0\,, \tag{3-15}$$

where $\boldsymbol{W}$ is a weight vector ($\boldsymbol{W} \in R^d$) and $b$ is bias (real numbers). The hyperplanes can be found for the classes, $y_i \in \{-1, +1\}$, such that

$$\begin{cases} \boldsymbol{W}^T x_i + b \geq +1 & if\ y_i = +1 \\ \boldsymbol{W}^T x_i + b \leq +1 & if\ y_i = -1 \end{cases}. \tag{3-16}$$

The terms in Equation (3-16) can be combined together and written as

$$y_i(\boldsymbol{W}^T x_i + b) \geq +1\,. \tag{3-17}$$

In the SVM problem, the aim is to find the hyperplane (Equation (3-15)) subject to the constraints in Equation (3-17) with the largest margins. The hyperplane should separate the features with the minimal error (i.e. maximum distance from the clusters' closest points). This problem can be formulated for linearly separable cases with the following quadratic programming problem (Shawe-Taylor and Cristianini, 2004)

$$\begin{gathered} \min \frac{1}{2} \|\boldsymbol{W}\|^2\,, \\ subject\ to\ \ y_i(\boldsymbol{W}^T x_i + b) \geq +1\,, \forall_i\,, \end{gathered} \tag{3-18}$$

In the case of linearly non-separable data, a set of loose variables, $\xi_i$, are introduced in the constraint. Therefore, the objective function in Equation (3-18) can be written as

.

$$\min \frac{1}{2} \|\boldsymbol{W}\|^2 + C_{SVM} \sum_{i=1}^{N} \xi_i \,, \tag{3-19}$$

where $C_{SVM}$ defines the constraint for misclassification error and subject to

$$y_i(\boldsymbol{W}^T x_i + b) = 1 - \xi_i \;,\; \xi_i \geq 0,\; i = 1, \ldots, N \,. \tag{3-20}$$

The quadratic programming problem can be solved using Lagrange multipliers, $\alpha_i \in R$. Therefore, $\boldsymbol{W}$ can be written as

$$\boldsymbol{W} = \sum_{i=1}^{N} \alpha_i y_i x_i \,, \tag{3-21}$$

where $x_i$ are non-zero values and $\alpha_i$ are the support vectors. The decision boundary is calculated using the support vectors by solving

$$\max_{\alpha} \left( \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \alpha_i \alpha_j y_i y_j x_i x_j \right) , \tag{3-22}$$

subject to

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \quad \text{and} \quad 0 \leq \alpha_i \leq C_{SVM} \;. \tag{3-23}$$

The learned parameters, i.e. $\alpha_i$ and $b$, are then used to label the data point by using the sign of the decision function

$$\hat{y}(x) = \operatorname{sgn}\left( \sum_{i=1}^{N} \alpha_i \, y_i (x . x_i) + b \right) \;. \tag{3-24}$$

Ruan *et al.* (Ruan *et al.*, 2007) applied an SVM classifier for a limited number of MRI modalities for segmentation of brain tumours. Their method could detect the whole tumour region without segmenting the tissue subtypes. They further developed their method by using feature selection to slightly improve the segmentation accuracy (Ruan *et al.*, 2011). Cai *et al.* (Cai *et al.*, 2007) and Verma *et al.* (Verma *et al.*, 2008) used intensity-based features from different MRI modalities including C-MRI and DTI and fed them to SVM for a voxel basis classification. Their methods were developed for multiclass segmentation of healthy and brain

tumour tissues. Lee *et al.* (Lee *et al.*, 2008) proposed using SVM to classify the MR image voxels and applying a CRF as spatial regularisation. Bauer *et al.* (Bauer *et al.*, 2011) proposed a SVM based classification which is regularised by hierarchical CRF which utilised prior knowledge from tissue neighbourhood system probabilities.

Wu *et al.* (Wu *et al.*, 2014) used superpixel features in a CRF framework to detect brain tumours. In their method, multimodal MR images were first segmented into superpixels, and then a multi-level Gabor wavelet was used for feature extraction from the superpixels. A SVM classifier is then used with an affinity metric model followed by a CRF to segment the tumour with maximum spatial smoothness.

### 3.4.2 Random Forests

RF is based on ensemble of decision trees, and classifies using a bagging process. A decision tree is so named because it is based on a decision model that can be presented as a tree-like graph. Decision trees can be used for both classification and regression tasks. They were first introduced by Breiman *et al.* (Breiman *et al.*, 1984) and named as Classification and Regression Trees (CART). The algorithms which are based on CART are considered among the best performing classification techniques (Wu *et al.*, 2008). The RF algorithm will be explained in Chapter 4 (Section 4.2.7).

Zikic *et al.* (Zikic *et al.*, 2012) suggested using context-aware features and decision forests classifiers for multi-structure classification of brain tumours from multimodal MR images. They utilised spatial regularisation with a smoothing constraint to eliminate the need for post-processing. Bauer *et al.* (Bauer *et al.*, 2012) proposed an energy minimisation procedure based on a CRF. The energy consists of sum of the singleton potentials and pairwise potentials which is minimised in a hierarchical way based on the output from a classifier. They used RF and used the probabilistic output of the RF to control the spatial regularisation of their CRF. The feature set included: first order features: mean, variance, skewness, kurtosis, energy, and entropy. The features were extracted from fixed sized local patches and the intensity gradient statistics in that neighbourhood, and symmetry features from across the sagittal plane. 5-fold cross-validation was used on training dataset for HGG and LGG separately. The parameters were selected for the algorithm empirically. They compared their method to the procedure which used SVM as a classifier instead of RF (Bauer *et al.*, 2011). However, their method has difficulties in segmentation of the testing dataset that was very different from the training set.

Germia *et al.* (Geremia *et al.*, 2012) also proposed the idea of creating synthetic images related to brain tumour in order to train the regression forests to improve to effectiveness of the

classifier. They developed their method by proposing spatially adaptive RF (Geremia *et al.*, 2013). Their method performed a hierarchical segmentation from coarse to fine segmentation. Tustison *et al.* (Tustison *et al.*, 2013) proposed using morphological and contextual features to better discriminate the homogeneity of the tumour. They also suggested using MRF to encourage the spatial regularisation. They later proposed a method in which the output of the first RF classifier was used to improve the second segmentation stage (Tustison *et al.*, 2014).

Festa *et al.* (Festa *et al.*, 2013) used RF with different image features which were voxel based. Their RF parameters were 50 number of trees and tree depth 25. The RF parameters were set using leave-one-out cross-validation of the training dataset. For testing on the BRATS training set, they used leave-one-out cross-validation. The training data points were downsampled and divided to half normal brain tissue and half for tumour and oedema to make the data balanced for more accurate classification. The feature sets included intensities (mean, sum, median and intensity range) from different patch sizes, context features, edge density and local binary partition. Meier *et al.* (Meier *et al.*, 2014a) proposed a semi-supervised RF method in which the classification was pixel-wise and based on patient-specific procedure. Amiri *et al.*, (Amiri *et al.*, 2016) proposed a deep RF-based hybrid method in which a SVM stage is cascaded to refine the final segmentation. Although their reported results of brain tumour segmentation were comparable to other RF-based methods, they did not outperform other machine learning based approaches.

One advantage of the tree classifiers is that they are easy to interpret. However, for larger trees the interpretations become more difficult. RF can handle unbalanced and a large number of data. Since the recent classification problems include a large dataset, this advantage of RF is more highlighted in the recent machine learning tasks. In terms of predictors, it can manage mixed types of predictor, i.e. numerical and categorical. In terms of attributes, it can somehow do feature selection. RF can handle multi-class classification tasks and also provide a probabilistic output (Criminisi *et al.*, 2012).

State-of-the-art supervised learning techniques based on RF indicated promising performance in the field of brain tumour segmentation (Rao *et al.*, 2015). They also provided promising results and best performance in challenges such as BRATS (Tustison *et al.*, 2014). However, their limitations in modelling capabilities makes their segmentation results to be outperformed by state-of-the-art deep CNN methods (Kamnitsas *et al.*, 2017).

## 3.5 Deep Convolutional Neural Networks

A CNN is a supervised technique with powerful model capability which recently has been widely used in learning tasks. Since one of the main contributions of this thesis is based on a deep CNN, it is represented with details as a separate subsection. CNN methods are based on learning the features automatically from the raw data or images. A deep CNN learns the hierarchy of the features from different complexity levels (Bengio *et al.*, 2013). The main difference with other methods is that instead of designing the features that require special knowledge, the attention is on designing the architecture of the model (LeCun *et al.*, 2015). Recently, the methods based on deep CNNs present powerful performance compared to classical classification methods. Deep CNN techniques have strong modelling capabilities and are able to learn highly discriminative features. Kamnitsas *et al.* (Kamnitsas *et al.*, 2017) suggested that the features based on deep CNN often outperform those which are predefined and hand-designed.

**Multilayer Neural Networks**

To describe the CNN, firstly multilayer neural networks are explained. The architecture of a multilayer neural network is comprised of many simple units to which most or all of them are subject to learning. These units establish a nonlinear mapping from the input toward the output. The last layer, which is a fully connected layer is called “output layer” and typically provides the class of the data. The units between the input and output layer are called the “hidden layers”.

A multi-layer neural network transforms the input data in a nonlinear way to create the class of the data in the output which are then linearly separable. Figure 3-1 presents an illustrative overview of this concept. The architecture of the network in Figure 3-1 consists of one input layer, one hidden layer and one output layer. As can be seen, the regular grid in the input space (left panel) is transformed in the hidden layer (middle panel) so that it is linearly separable. The actual networks consist of several layers with hundreds of thousands of units.

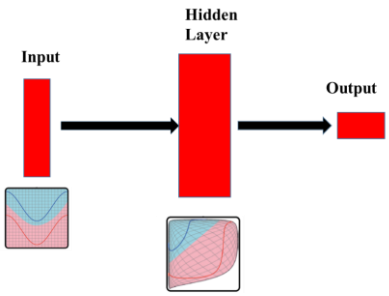



Figure 3-1 An illustrative overview of neural network concept of non-linear distortion of the input space to make them linearly separable. The network architecture consists of one input layer, one hidden layer and one output layer. The panels are reproduced from (Colah, 2017).

Setting up a deep network (i.e. with the depth of 5 to 20 layers) will lead to a system which is sensitive to very small details of the objects and at the same time they are invariant to differences such as background position and intensity contrast (LeCun *et al.*, 2015).

### 3.5.1 Introduction to CNN

The idea of convolutional neural networks is to perform an end-to-end classification, in which the inputs are the image pixels and the output is the class of that image. CNNs are designed for processing multiarray data, such as natural images (colour images), medical images (multimodal MRI), etc. Therefore, CNN consists of units which has three dimensions: width, height, and channels. An illustration of the 3D architecture of a CNN is presented in Figure 3-2. As can be seen, every layer of the CNN transforms the 3D input of the layer to a 3D output array. Furthermore, the units in each layer are arranged in 3D structure, i.e. width, height, and channels.

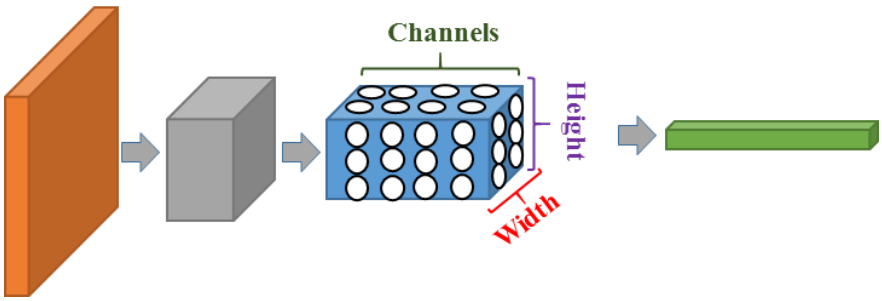



Figure 3-2 A general 3D illustration of a CNN architecture. It consists of 3D input layer (image), two 3D hidden layers, and 3D output array. The units of the second hidden layer are also illustrated that they are arranged in 3D structure. Inspired and reproduced from ("CS231n Convolutional Neural Networks for Visual Recognition," 2017)

**Layers of a CNN**

The architecture of a CNN is typically comprised of a series of stages. The first stage contains two layer types which are convolutional layers and pooling layers. Each layer consists of feature maps with units within them. Different layers of a CNN are explained in the following.

***Input Layer:*** The input layer holds the raw multi-array input data. In the case of MRI data, the width and height of the input layer are similar to the multimodal MR image spatial dimension in one slice, and the depth (third dimension) is equal to the number of MR channels, i.e. protocols.

***Convolutional Layer:*** The convolutional layer (CONV) is the main part of a CNN with most computational contribution of learning. In a convolutional layer the units are connected to the local patches of the previous layer via a set of weights which are called "filter kernels". The outputs of the previous unit are convolved with the kernel of the current convolutional layer and create the output. This means the output of a CONV layer is locally connected to the local regions in the input using dot product between the CONV weights and that local input region.

The result of convolution is a feature map. When the filter is slid over the width and height of the input, the result is a 2D filter map in which every output element is the filter response at every spatial position of the input. Every CONV consists of a set of filters, and each of them produce a 2D feature map. By concatenating these filter maps along the depth (third dimension), the output volume is created.

The sliding step of the filter is called "stride". Convolutional layers usually utilise stride 1, which means the filter is sliding one pixel in every iteration. Applying stride more than 1 generates smaller size output and is computationally more effective. However, stride more than 2 is not common in practice.

Convolution will result in decreasing the filter response dimension compared to the input array spatial dimensions. To make them similar size, it is suggested to add zeros to the spatial margins of the input array, which is called "zero-padding".

The outputs of this weighted summations are passed through a noninear function. The most popular nonlinear function is rectified linear unit (ReLU) which is represented by

$$f(z) = \max(z, 0). \tag{3-25}$$

The benefit of ReLU is that it can learn faster in networks with a large number of layers, which make it suitable for deep learning procedures.

***Pooling Layer:*** The pooling layer (POOL) merges the features with similar semantic properties into one feature value. It can be considered as a down-sampling layer along with the spatial dimensions in $X$ and $Y$ directions. A CONV or several CONVs are generally followed by a pooling layer. The pooling unit obtains the input from patches that are shifted by more than one array. The result reduces dimensionality and is less dependent on small shifts. It also reduces the number of parameters in the model.

Max-pooling is a common pooling operation in which commutes the maximum value of a local patch in one feature map. The common window for max-pooling is 2 × 2. Figure 3-3 shows an example of a max-pooling layer with filter size 2 and stride 2.

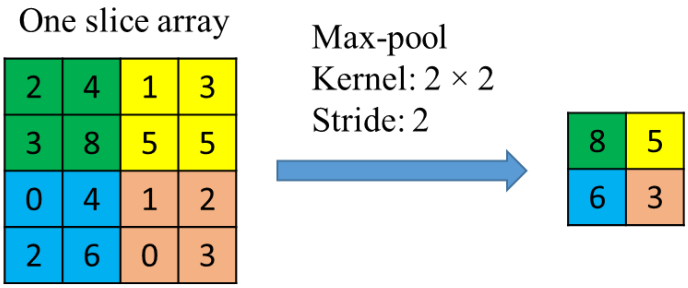



Figure 3-3 A max-pooling example with kernel size 2 × 2 and stride 2.

The pooling layer makes the network invariant with respect to small changes of each level of feature such as intensity or position in the previous layers. However, Springenberg *et al.* (Springenberg *et al.*, 2014) suggested that future CNNs will use less or no pooling layers.

***Fully Connected Layer (FC):*** The difference between FC and CONV is that in a CONV, the units are connected locally to the input, inspiring the concept of parameter sharing. Whilst in a fully connected layer, the unit is connected to whole input region, which means all the units are connected to the outputs of the previous layer. The FC layer eventually generates the classes of the data.

**Architecture of CNN**

In the previous sections the main layers of a CNN were categorised as: CONV, POOL, and FC. It should be noted that, the nonlinear function of ReLU is also mentioned when it is attached to a layer, i.e. CONV_ReLU. The common architecture of a CNN is comprised of blocks in which few CONV_ReLU layers are followed by POOL layers. These blocks are repeated until the image has a small spatial size. The last block is then connected to a FC layer which creates the class label. This pattern is presented in Figure 3-4.

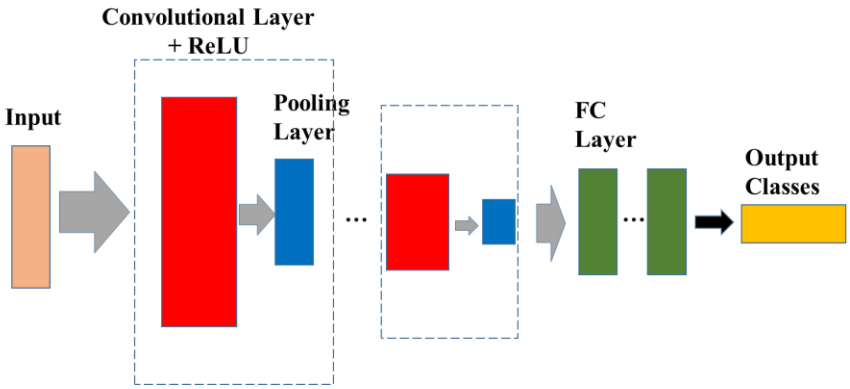



Figure 3-4 General overview of CNN architecture.

The images have properties which deep learning benefits from. The images are comprised of hierarchies, which means that higher-level features are generated by combining lower-level features which creates the hierarchies. In other words, small local edges (the lowest level feature) form the local patterns, and those local patterns create more general patterns which at the end create the objects in the images.

### 3.5.2 State-of-the-art CNN architectures

Recent CNN architectures have 10 to 20 layers of ReLUs, hundreds of millions of weights and billions of connections within the network (LeCun *et al.*, 2015). Training such a huge network model is supposed to be very time-consuming. Thanks to the developments in computer hardware and software, and using parallel processing with parallelised hardware, i.e. GPU computing, the training time has decreased dramatically in recent years.

Several architectures are proposed for designing a CNN and will be explained in the following subsections. The most popular grand challenge for research on CNN methods for object detection and image classification is ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) (Russakovsky *et al.*, 2015). The state-of-the-art CNN architectures obtain promising results in the competition.

**LeNet**

The first applicable architecture of CNN was proposed by LeCun *et al.* (LeCun *et al.*, 1998) to read the zip codes digits, etc., which is known as LeNet ("MNIST Demos on Yann LeCun's website," n.d.). In this architecture, the layers have single CONV which is followed by a POOL.

**AlexNet**

The first popular architecture of CNN, applicable in the field of computer vision, was proposed by Krizhevsky *et al.* (Krizhevsky *et al.*, 2012), which is known as AlexNet. The model was the winner of ImageNet ILSVRC challenge 2012. AlexNet is a developed version of LeNet with more layers. Also, CONVs are connected after each other and then fed to a POOL. AlexNet architecture consists of five CONV layers, some of which were followed by max-pooling, and three FCs.

**ZFNet**

ZFNet architecture was proposed by Zeiler and Fergus (Zeiler and Fergus, 2014), which was the winner of ILSVRC 2013. The ZFNet architecture is a developed version of AlexNet, in which the size of middle CONVs was expanded, smaller stride and filter size in the first layer was used. CONV has kernel size of 7 × 7 with stride 2. The max-pooling has size of 3 × 3 with stride 2. The CONV structure is repeated for five layers. The last part of the ZFNet architecture consists of two FCs.

**GoogleNet**

GoogleNet was proposed by Szegedy *et al.* (Szegedy *et al.*, 2015), from Google, which was the winner of ILSVRC 2014. They reduced the number of parameters in the network by introducing "inception module" which approximates the optimal local sparse structure from dense components. The structure of an inception module is presented in Figure 3-5.

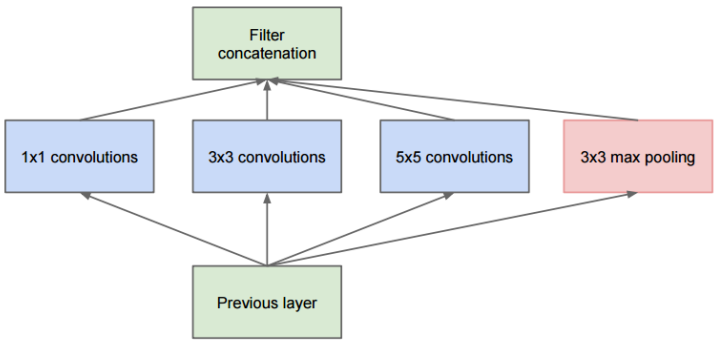



Figure 3-5 The structure of inception module proposed in (Szegedy *et al.*, 2015).

They also replace FC by average-pooling in the last layer which decreases a large number of parameters. This architecture consists of 3 CONVs, and 18 inception layers. Max-pooling is used as the POOL layer.

**VGGNet**

VGGNet was proposed by Simonyan and Zisserman (Simonyan and Zisserman, 2014), from Oxford Visual Geometry Group, in which they suggested that the depth of the network has an important effect on the performance of the model. The main characteristic of VGG was the convolution and pooling kernel size. All the CONVs consisted of 3 × 3 convolution with stride 1 and zero-padding 1. At the POOL, max-pooling size of 2 × 2 with stride 2 and zero-padding 0 were used. The main concern about VGGNet was that using a large number of parameters and memory makes it computationally expensive, which has been tackled by using recent high performance computers and parallel processing. They proposed different architectures with 11, 13, 16, and 19 layers. The 19-layers structure (VGG19) provided a slightly better performance compared to 16-layers (VGG16), however it is computationally more expensive. Therefore, VGG16 was suggested to be the better architecture which consists of 16 Convolutional/FC layers.

**ResNet**

He *et al.* (He *et al.*, 2016) proposed the architecture of residual networks (ResNet) which was the winner of ILSVRC 2015. They suggested using skip connections alongside with batch normalisation. They also eliminate the FC at the end of the network. ResNet is considered amongst the extremely deep architectures. However, the ResNet architecture involves a very large number of parameters and is computationally expensive.

**Deconvolutional NN**

Typical convolutional networks eradicate mid-level information of the edges in the images such as intersections, parallelism, etc. Deconvolution networks were introduced by Zeiler *et al.* (Zeiler *et al.*, 2010) in order to consider these level of features and were able to learn them. The main idea was using deconvolution operations and unpooling that are depicted in Figure 3-6.

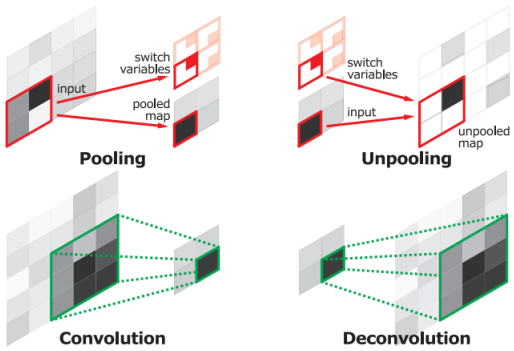



Figure 3-6 Deconvolution and unpooling compared to convolution and pooling (Noh *et al.*, 2015).

Noh *et al.* proposed a CNN architecture based on deconvolutional layers (DeCONV) for semantic segmentation of images (Noh *et al.*, 2015). Their proposed architecture was a modification of the VGG16 model and performed an end-to-end pixel-wise classification. The overall architecture of the deconvolutional CNN is presented in Figure 3-7. As can be seen, the architecture consists of two main parts, i.e. convolutional and deconvolutional networks. The left side of the architecture in Figure 3-7 is a convolutional network and derived from the VGG16 net structure which has 13 convolutional layers and 2 FC layers. The right side of the architecture in Figure 3-7 is the mirrored version of the convolutional side and consists of DeCONVs and unpooling layers.

The deconvolutional part of the architecture provides a coarse-to-fine structure of the segmented object, which was reconstructed from the convolutional part.

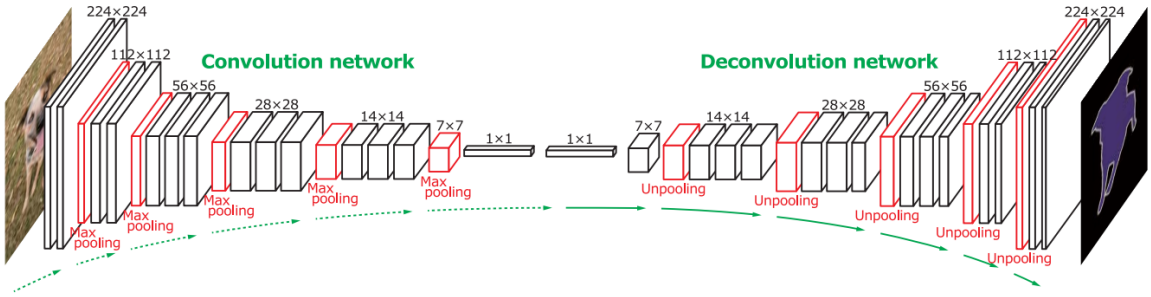



Figure 3-7 The architecture of deconvolutional neural network which was proposed by (Noh *et al.*, 2015).

**Fully Convolutional Networks**

Long *et al.* (Long *et al.*, 2015) proposed fully convolutional networks (FCN) for segmentation of images. The convolution operations are used in the last layer of a CNN to extend it for semantic segmentation. A skip layer (SKIP) architecture is introduced which combines the semantic information from the deep layers and finer appearance information from the shallow layers. It was also suggested to utilise upsampling instead of POOL. Therefore, the resolution of the output can be increased which results in finer segmentation. The image features that are learned from FCN will be used in this thesis. To provide a consistent explanation of the method, a detailed structure of FCN will be explained in Chapter 6. Long *et al.* (Long *et al.*, 2015) compared the FCN structure based on different networks, e.g. AlexNet, VGG (16- and

19-layers) and GoogleNet. They concluded that VGG16 provides the best results and performance.

**U-Net**

Ronnenberger *et al.* (Ronneberger *et al.*, 2015) proposed the U-Net architecture which was an extension of FCN (Long *et al.*, 2015) for end-to-end segmentation of the image. The main characteristic of U-Net is using data augmentation to train the network more efficiently. This enables the model to be trained efficiently with less number of training samples. U-Net also does not include any FC layers. The structure of a U-Net is depicted in Figure 3-8, which has in total 23 layers. The left side of the network architecture is a typical CNN and is called "contracting" path. This part has two 3 × 3 CONVs with 0 zero-padding followed by ReLU. Each block also includes a max-pooling of kernel size 2 × 2 with stride 2. The right side of the U-Net performs upsampling and is called "expansive" path. This part includes upsampling of the feature map, each followed by a convolution with kernel size 2 × 2 which is called up-convolution. The feature maps are concatenated with a cropped feature map from the contracting side. Then, they are followed by two CONVs with kernel size 3 × 3 and two ReLUs.

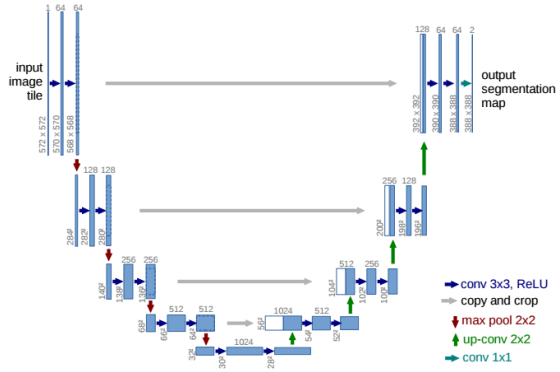



Figure 3-8 Architecture of the U-Net which is proposed by (Ronneberger *et al.*, 2015)

**Discussion of Recent CNNs**

LeNet, AlexNet, ZFNet, GoogleNet, and VGG are the state-of-the-art image classification nets. ZFNet is derived from AlexNet, both nets split an input into two main channels (colour, and grey-level). The VGG verifies that a small kernel can provide better performance than using a large kernel. Although the de-convolutional NN was proposed as a visualisation tool to understand the CNN, it was employed to implement pixel-wise prediction (semantic segmentation). FCN combines the semantic and appearance information which provides finer segmentation.

### 3.5.3 CNN for Brain Tumour Segmentation

The CNNs were firstly used for speech recognition (Waibel *et al.*, 1989) and document reading (LeCun *et al.*, 1998) by introducing one-dimensional time-delay neural networks. Then other applications of CNNs in optical character recognition and handwriting recognition were developed (Simard *et al.*, 2003). They were also developed for natural image processing including recognition of objects such as face (Lawrence *et al.*, 1997) and hand (Vaillant *et al.*, 1994). CNNs have been applied for several image processing tasks, especially image segmentation and detection, since the early 2000s (LeCun *et al.*, 2015). These applications include segmentation of biomedical images (Ning *et al.*, 2005), face detection (Taigman *et al.*, 2014), human bodies (Tompson *et al.*, 2015), etc.

CNN-based methods have been recently adopted by researchers in the field of brain tumour segmentation. Rao *et al.* (Rao *et al.*, 2015) proposed a combination of CNN and RF. In their method, the output of the last fully connected layer and the SoftMax of each CNN are fed to a RF classifier. Urban *et al.*(Urban *et al.*, 2014) *et al.*proposed using a 3D filter as kernels of the convolutional layers. The advantage of using 3D filters is that it will take into account the 3D connectivity of the original data. On the other hand, it increases the computational time and cost. Dvořák *et al.* (Dvořák and Menze, 2015) proposed structured prediction based on a CNN, in which the tumour segmentation is portioned into several binary segmentation subsets. The membership of the input to each cluster is predicted by the CNN.

Zikic *et al.* (Zikic *et al.*, 2014) used a shallow CNN architecture which comprised of two CONVs, max-pooling with stride 3, one FC, and one SoftMax layer. Lyksborg *et al.* (Lyksborg *et al.*, 2015) proposed to stack 2D binary CNNs for segmentation of 3D MRI images. The idea is setting up 2D pipelines for the three MRI orthogonal planes, i.e. sagittal, coronal, and axial. Therefore, the final volumetric segmentation mask is obtained by combining the 2D outputs of the three pipelines. Since their CNNs are binary, the procedure is performed in a multi-level approach to segment different tissues. Furthermore, a post-processing stage of cellular seed growing is used to acquire better segmentation results.

Pereira *et al.* (Pereira *et al.*, 2016) suggested that using two CONVs with kernel size $3 \times 3$ have the same effective receptive fields and fewer weights compared to a single CONV with kernel size $5 \times 5$. One advantage of the smaller kernel sizes is that since they have fewer weights, it decreases the overfitting. They proposed two different architectures for different tumour types, i.e. HGG and LGG. For LGG, they proposed a CNN with 4 CONVs with kernel size $3 \times 3$, and 3 FCs. While, due to more complex structures of HGG, they proposed a deeper network with 6 CONVs and the same kernel size. They compared their method with shallow CNN architectures and concluded that a deeper architecture provides better performance. They

applied their method on the BRATS challenge dataset and obtained promising results in terms of DSC for all the tumour structures. Their segmentation results achieved the first rank in the VSD system and remained the first rank at the time of submission of this thesis.

Davy *et al.* (Davy *et al.*, 2014) proposed a two-pathway deep network with parallel pipelines. One pipeline is "convolutional" and comprised of two CONVs and a MaxOut unit. This pipeline is connected directly to patches from the input with fixed size of 32 × 32 and considers a wider and more general region of the input. The other pipeline is "fully connected" and comprised of a FC and a MaxOut unit. This pipeline is connected with two smaller patches from the input with size 5 × 5 and considers more local detail of the image. Finally, the output of the two structures is merged together with another FC layer and a SoftMax unit. Havaei *et al.* (Havaei *et al.*, 2017) used the method which was proposed by (Davy *et al.*, 2014) and developed an architecture with two stages with different kernel sizes for global and local contextual segmentations. The global stage used larger CONV kernel size of 13 × 13, whilst the local stage used smaller CONV kernel sizes of 7 × 7 and 5 ×5.

Kamnitsas *et al.* (Kamnitsas *et al.*, 2017) proposed 3D CNN with a two-pathway architecture to consider different scales of processing the input image. The architecture has 11 layers and considers a dense training scheme to decrease the computational cost by including the neighbours of a pixel in one pass through the network. They also suggested using a fully connected CRF at the output to decrease the false positives. The architecture provides promising results which are comparable to the state-of-the-art methods. However, dense training may generate highly imbalanced classes that should be investigated in their architecture.

Dong *et al.* (Dong *et al.*, 2017) used U-Net for segmentation of brain tumours in MRI images. Their method uses a state-of-the-art deep CNN architecture that takes the information from up-convolution paths, which are applied to both deep and shallow layers. The U-Net provided more accurate segmentation compared to other CNN methods, and Chapter 6 will discuss that they are comparable to the proposed method in this thesis.

## 3.6 MICCAI-BRATS Publication Series

BRATS challenge is organised in conjunction with the international conference on Medical Image Computing and Computer Assisted Interventions (MICCAI) (Menze *et al.*, 2015). The state-of-the-art brain tumour segmentation methods have used this grand challenge (also known as MICCAI-BRATS) dataset in recent years. As explained in Chapter 2 (Section 2.6.2), a publicly available training dataset is shared by the organisers so that different research groups

optimise their methods on that dataset. Then, they applied their methods on an independent testing dataset in a structured way and evaluated using a predefined common protocol. The corresponding results have been published either in the MICCAI proceedings or other high impact journals, which are summarised in Table 3-1.

Table 3-1 Summary of MICCAI-BRATS results for the publications related to the literature review. The evaluation dataset and Dice scores for the tumour part are provided. The research papers are sorted based on the publication year.

| **Author** | **Method** | **Year** | **Testing dataset** | **Performance (Dice Scores)** | | |
|---|---|---|---|---|---|---|
| Bauer (Bauer *et al.*, 2012) | RF and CRF | 2012 | Challenge | 0.68 | 0.48 | 0.57 |
| Zikic (Zikic *et al.*, 2012) | 3D input patches are interpreted into 2D input patches to train a CNN | 2012 | Challenge | 0.75 | 0.47 | 0.56 |
| Hamamci (Hamamci *et al.*, 2012) | Generative model, uses cellular automata to obtain tumour probability map | 2012 | Challenge | 0.72 | 0.57 | 0.59 |
| Reza (Reza and Iftekharuddin, 2013) | Texture features and RF | 2013 | Challenge | 0.83 | 0.72 | 0.82 |
| Zhao (Zhao *et al.*, 2013) | Supervoxel and MRF classifier with updated unary potential | 2013 | Challenge | 0.84 | 0.70 | 0.80 |
| Cordier (Cordier *et al.*, 2013) | 3D patch atlas-based | 2013 | Challenge | 0.84 | 0.68 | 0.88 |
| Festa (Festa *et al.*, 2013) | Local context features and RF | 2013 | Challenge | 0.72 | 0.66 | 0.77 |
| Doyle (Doyle *et al.*, 2013) | Hidden MRF | 2013 | Challenge | 0.71 | 0.46 | 0.66 |
| Meier (Meier *et al.*, 2014b) | Appearance and context features with RF and CRF | 2014 | Challenge | 0.82 | 0.73 | 0.69 |
| Davy (Davy *et al.*, 2014) | Two-pathway CNN for simultaneous local and global processing | 2014 | Challenge | 0.85 | 0.74 | 0.68 |
| Kwon (Kwon *et al.*, 2014) | Generative model performs joint segmentation and registration | 2014 | Challenge | 0.88 | 0.83 | 0.72 |

| Author | Method | Year | Testing dataset | Performance (Dice Scores) | | |
|---|---|---|---|---|---|---|
| Tustison (Tustison *et al.*, 2014) | Concatenated RFs; asymmetry and first order statistical features | 2014 | Challenge | 0.87 | 0.78 | 0.74 |
| Urban (Urban *et al.*, 2014) | 3D CNN architecture using 3D convolutional filters | 2014 | Challenge | 0.86 | 0.75 | 0.73 |
| Zikci (Zikic *et al.*, 2014) | Standard CNN with 5 layer and ReLU | 2014 | Training (HG) | 0.84 | 0.74 | 0.69 |
| Rao (Rao *et al.*, 2015) | Four CNNs, one for each modality, with their outputs concatenated as an input into a RF | 2015 | - | - | - | - |
| Dvorak (Dvořák and Menze, 2015) | Local structured prediction with CNN and *k*-means | 2015 | Challenge | 0.83 | 0.77 | 0.75 |
| Njeh (Njeh *et al.*, 2015) | Graph-cut distribution matching | 2015 | Training | 0.875 | - | - |
| Thiruvenkadam (Thiruvenkadam and Perumal, 2016) | Discrete wavelet transform and FCM | 2016 | Challenge | 0.73 | 0.53 | 0.35 |
| Amiri (Amiri *et al.*, 2016) | Deep RF and SVM refinement | 2016 | Partial training | 0.72 | - | - |
| Pereira (Pereira *et al.*, 2016) | CNN with small (3x3) kernals for deeper architecture | 2016 | Challenge | 0.88 | 0.83 | 0.77 |
| Havaei (Havaei *et al.*, 2017) | Cascaded two-pathway CNNs for simultaneous local and global processing | 2017 | Challenge | 0.88 | 0.79 | 0.73 |
| Kamnitsas (Kamnitsas *et al.*, 2017) | 3D CNN with two-pathways with fully connected CRF | 2017 | Challenge* | 0.84 | 0.65 | 0.62 |
| Dong *et al.* (Dong *et al.*, 2017) | U-Net architecture | 2017 | Training* | 0.86 | 0.86 | 0.65 |

* Evaluated on BRATS 2015 dataset. The version 2013 is not reported.

## 3.7 Summary and Conclusion

Most of the brain tumour segmentation techniques used hand-designed features (such as texture, etc.) which are fed into a classifier such as SVM, RF, etc. Among the conventional classifiers, RFs presents the best segmentation results (Gotz *et al.*, 2014; Menze *et al.*, 2015; Pinto *et al.*, 2015). The challenge of supervised approaches is finding a good representation of feature, which in the case of hand-designed features requires prior knowledge for designing and parameter tuning. Furthermore, many experiments and optimisations are required to be conducted to identify the optimum parameters for feature extraction and the optimal classifier.

Several approaches based on deep CNNs have been proposed in recent years and provide promising segmentation results. However, they have the disadvantage of not considering sufficient local dependencies. Recent methods have aimed to tackle this limitation to providing finer segmentations by modifying the architecture of the CNN (Havaei *et al.*, 2017) or combining with other methods such as CRF (Kamnitsas *et al.*, 2017).

An extensive range of segmentation algorithms are proposed for brain tumour segmentation. However, there is no universal method that can handle all segmentation tasks. On the other hand, most of the methods are optimised for segmentation using a specific imaging modality. This thesis not only investigates the single modality approach, but also looks at a multimodal framework. Many of the methods are applied on different datasets so it is difficult to perform a general comparison between them. Thanks to the BRATS challenge public dataset and its straightforward evaluation protocol, now it is relatively easier to establish a fair comparison between the state-of-the-art methods. For this reason, many advanced methods in the field of brain tumour segmentation in MRI are compared and discussed in Section 3.6.

# Chapter 4

# Brain Tumour Segmentation using Superpixel in FLAIR MRI

## 4.1 Introduction

Delineation of the tumour boundary and assessment of tumour size are needed for patient management in terms of treatment planning and monitoring treatment response. Current clinical guidelines incorporate the use of both T1-contrast images and T2-weighted/FLAIR images (Niyazi *et al.*, 2016; Wen *et al.*, 2010). As explained in Chapter 2, each MRI protocol presents specific characteristics of the brain or tumour tissue. Many low-grade gliomas do not show contrast enhancement hence T2-weighted/FLAIR images are used to define the tumour extent and volume. T2-weighted/FLAIR images can also be useful to help define the target volumes for radiotherapy planning of high-grade gliomas (Aslian *et al.*, 2013; Niyazi *et al.*, 2016). From a technical point of view, using a single protocol as input to solve a binary class segmentation problem decreases the complexity of the model due to less data, smaller feature dimensionality and no need for image registration. FLAIR is considered for this task as it has been in routine clinical use as part of standard diagnosis of brain tumours. Delineation of the FLAIR hyperintensity is important to assess low-grade glioma growth (Law *et al.*, 2008), define an abnormal region from which imaging features for tumour classification can be extracted (Itakura *et al.*, 2015), aid with radiation dose planning (Stall *et al.*, 2010) and assess the treatment responses (Cho *et al.*, 2012). Detecting the complete tumour in the image can be considered as an initial stage which can be further used for tumour component segmentation tasks.

The motivation of this chapter is to use a single modality approach to detect the complete tumour structure in the clinical data. To assess the robustness of the proposed method, it is also evaluated on the FLAIR protocol of BRATS 2013 annotated training dataset.

As mentioned in Chapter 3, the methods in (Pinto *et al.*, 2015) and (Gotz *et al.*, 2014) calculated the image features based on each individual voxel, which a fixed size neighbourhood around the voxel is considered for the feature extraction. In this chapter, instead of using a fixed size neighbourhood, a patch based method using superpixel partitioning is investigated for feature extraction and the final segmentation is directly obtained from the superpixel's boundary. This will ensure that the patches are more separable for the classifier, since the pixels inside the patch have more similarity compared to the fixed sized window suggested in (Pinto *et al.*, 2015) and (Gotz *et al.*, 2014). It will also increase the

computational speed of the feature extraction and classification stages compared to the pixel-wise based approaches which require the corresponding procedures to be performed on all the pixels within the input image.

The idea of using superpixel instead of the pixel-level calculation have been used for object classification in the images. Fulkerson *et al.* (Fulkerson *et al.*, 2009) proposed using the superpixel patches of images instead of pixel level classification. They extracted bags of features for each superpixel and classify them using SVM. In this thesis, one of the contributions including the histogram of textons in the bag of features that is extracted from the superpixels.

Yi and Sun (Yi and Su, 2014) used the histogram of Gabor filter responses as the representation of the features. They suggested that using log-Gabor filters will reduce dimensionality and the computation cost. Fixed-sized non-overlapping blocks of the images was used in their method to calculate the histogram of Gabor histogram. In this thesis, flexible superpixel patches are used instead of the fixed blocks. The flexible boundaries of the superpixels create non-overlapping patches that adhere to the image edges.

Yu *et al.* (Yu *et al.*, 2012) used the bag of textons and superpixel for unsupervised image segmentation. The texton filter bank were comprised of 2D Gaussian filters. In this thesis Gabor filters are used which represent more description of spatial and frequency features.

In this chapter, a fully-automated superpixel based method will be investigated for detection and segmentation of the abnormal tissue associated with brain tumours, as defined by the hyperintensity from FLAIR MRI. In the proposed method, superpixel partitions are firstly calculated to provide accurate boundaries between different tissues. Several non-parametric and hand designed image features are then extracted from each superpixel. This will improve the accuracy of feature calculation and increase the computation speed. The superpixels are then classified using the state-of-the-art ERT which is a powerful classifier that can deal with high dimensional features and large-sized unbalanced data.

A texton is considered as a texture feature for the whole image segmentation. The idea is to use the texton histogram of a specific superpixel (which is obtained from the whole image texton map) as the main feature for that superpixel. This will be discussed in this chapter.

The contribution of this chapter can be listed as follows:

- Investigation of an automated method that provides a close match to expert delineation across all grades of glioma using the single commonly used MRI modality i.e. FLAIR.

The method could provide a faster segmentation (approximately three times faster) of brain tumours, and good agreement with the human expert delineation .

- Extraction of powerful hand designed features from superpixels instead of common pixel-wise feature extraction, with focus on the state-of-the-art texton features.

- Applying ERT directly to the superpixels instead of all the voxels (Gotz *et al.*, 2014; Pinto *et al.*, 2015), which largely reduces the data size for classification. Superpixels with the same classification label are grouped together, and are considered as the tissue ROI.

## 4.2 Methodology

This section will explain the details of the proposed method, which includes extraction of features from superpixels and superpixel classification.

### 4.2.1 General Overview of Segmentation Methods based on Classifiers

The segmentation methods which are based on classification of the pixels or voxels consist of three main essential stages which are: preprocessing, feature extraction (if necessary followed by feature selection), and classification. The majority of the methods in the literature review used this pipeline. Firstly, the MR images should be pre-processed prior to the segmentation. As the brain tissue has complex structures in MR images, different varieties of features may be needed to better distinguish the patterns. To reduce the dimensionality of the feature space, a feature selection stage may be needed. Then a classifier is utilised to label the pixels/voxels to the corresponding tissues.

The overall pipeline of the proposed method is depicted in Figure 4-1. The first stage is preprocessing and preparation of the data for the main part of the method. Then in the superpixel segmentation stage, FLAIR images are partitioned into irregular patches with approximately similar size and intensity values. Several hand-designed features are then calculated for every superpixel, which include statistical, texton, and shape features. A feature selection approach is then applied to find the most significant features. Finally, each superpixel is classified using an ERT into binary classes, i.e. tumour and non-tumour. The proposed method will be referred as superpixel ERT method (SP_ERT).

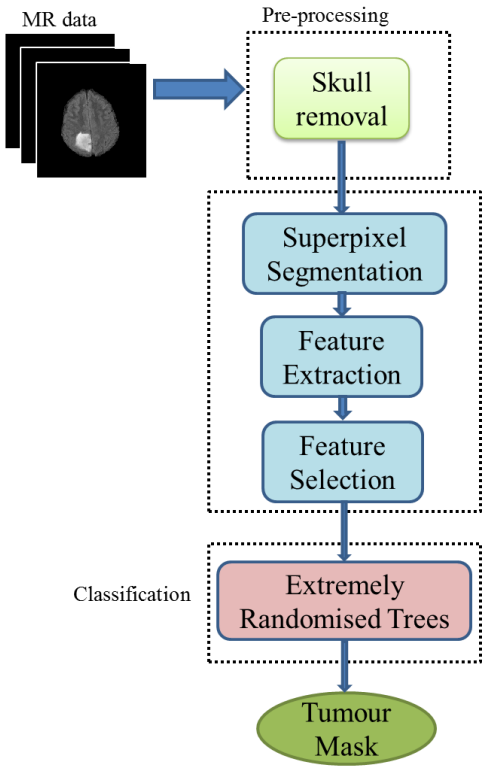



Figure 4-1 The entire workflow of the proposed SP_ERT method

### 4.2.2 Preprocessing

Preprocessing is an important part in MRI brain image segmentation which consists of skull stripping, and intensity normalisation. Skull stripping or skull removal is a common preprocessing step in most MRI brain segmentation techniques (Menze *et al.*, 2015). The aim of this step is to separate the brain tissue from the skull and non-brain tissues in MRI images. The importance of skull removal in MRI brain image analysis is discussed in (Choubey and Agrawal, 2012; Shanthi and Kumar, 2007; Zhuang *et al.*, 2006). In this work, the skull is removed from all the MRI images using fMRI of the brain (FMRIB) software library (FSL) ("FSL," n.d.) (Jenkinson *et al.*, 2012).

Intensity normalisation (Madabhushi and Udupa, 2005) is very important in parametric approaches to ensure that a unified set of parameters are obtained during parameter tuning. They are also important in classification-based segmentations to ensure the features, especially those based on intensity values, have the normalised values and share similar dynamic range. Before performing normalisation, 1% highest and lowest intensity values for each image are eliminated. The 1% highest values correspond to the hyper intensities of the remaining voxels related to the skull, while the 1% lowest values correspond to the elimination of the background noise. The intensities are initially normalised for each patient image by subtracting

the average of intensities of the image and dividing by their standard deviation. Then histogram matching algorithm (Nyúl *et al.*, 2000) is applied to ensure that all the data have a similar dynamic range. This is to ensure that the corresponding tissues in images, which are acquired from MRI machines with different magnetic field strengths (i.e. 1.5 or 3 T), have a similar intensity range. One of the patient images is selected and its histogram is considered as the reference. Then, the histogram of each patient image is matched to the reference histogram (Figure 4-2).

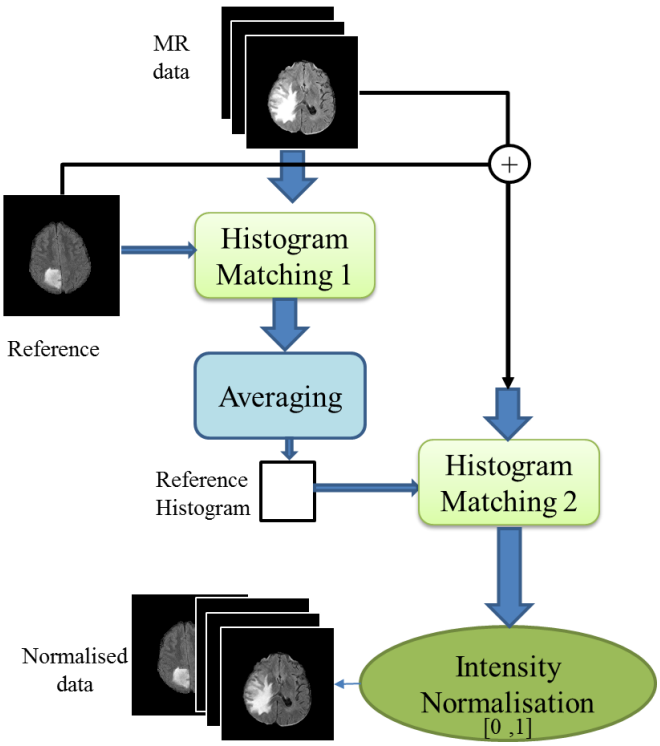


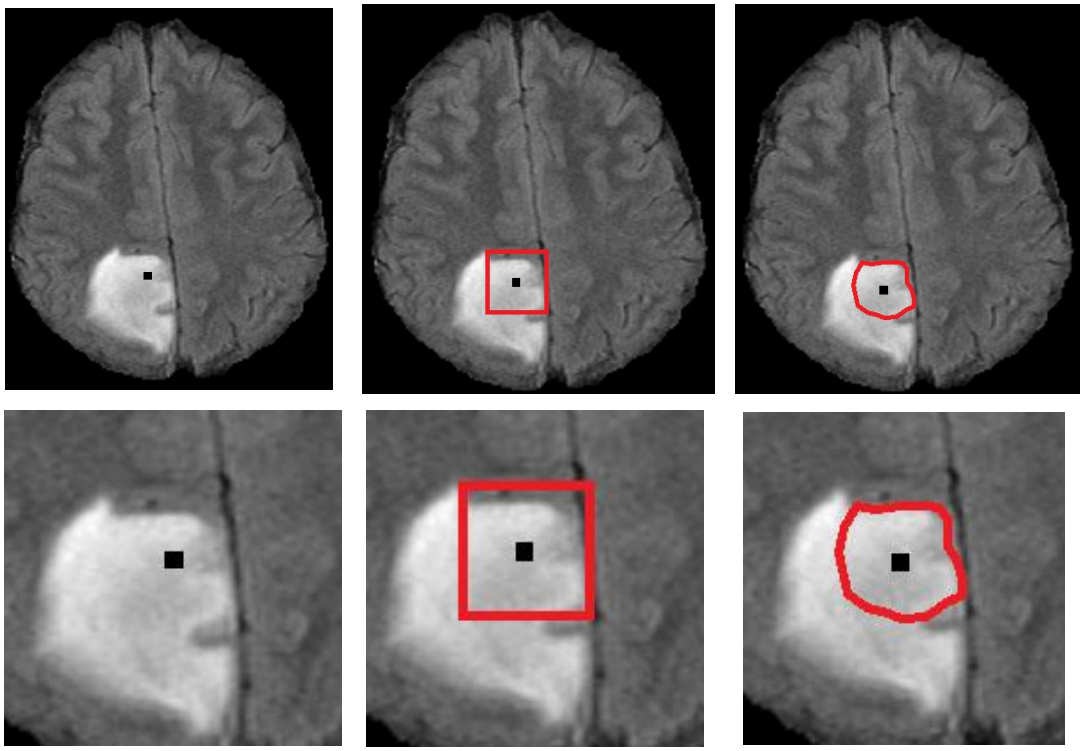

Figure 4-2 The flowchart of histogram matching and linear normalisation to the range [0, 1].

To decrease the bias towards the reference image, a second histogram matching stage is performed by calculating the average of new histograms (Nyúl *et al.*, 2000) and treat it as the new reference histogram and matching all the histograms including the initial patient image. Finally, the dynamic range of the intensities is linearly normalised to the range [0 ,1] (Figure 4-2).

### 4.2.3 Image Patch Types for Feature Extraction

Most of the image segmentation methods which are based on classification of the pixels used a fixed size area around that pixel to calculate the local features. Usually this area is a square window in which the centre is the pixel under consideration. Figure 4-3(a) shows a FLAIR

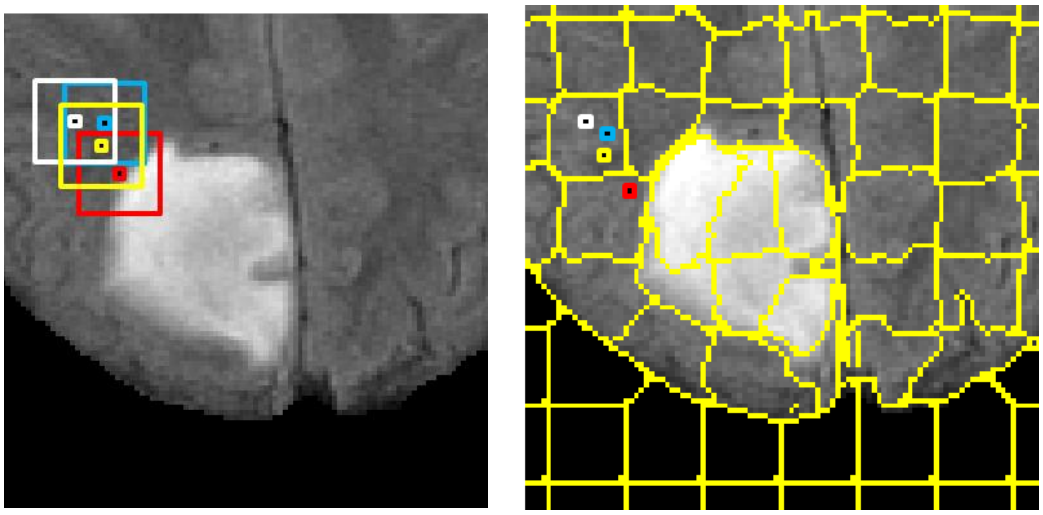
image, and the target pixel for feature extraction is shown as a black point. The fixed size square window which is illustrated in Figure 4-3(b), is used for extraction of features based on local dependencies and the neighbourhood. As can be seen, the square window contains areas with different intensities due to the complex structure of brain and tumour. Using image patches with flexible boundaries (Figure 4-3(c)), which encompass more homogenous regions, may increase the accuracy of feature extraction.



a b c

Figure 4-3 Image patch types for local image calculations. a) Fixed-size windows b) homogenous patches with flexible boundaries. The bottom row is the zoomed-in view of the upper row.

Another advantage of using a flexible homogenous image patch instead of a fixed-size window in pixel-wise approaches is that it will significantly reduce the number of calculations. The fixed-size patches use the moving window approach in which a window is assigned for each individual pixel in the image and then features are calculated for each window (Figure 4-4 (a)). Whilst for the flexible homogenous patches, the calculations corresponding to a specific patch can be assigned to all the pixels within the patch (Figure 4-4 (b)).



a b

Figure 4-4 Pixel-wise and patch-based calculations schemes. a) pixel-wise, b) patch-based.

### 4.2.4 Superpixel Segmentation

Superpixel segmentation partitions the image into flexible patches with approximately similar size and intensity values. The simple linear iterative clustering (SLIC) (Achanta *et al.*, 2012) method is used in the proposed SP_ERT method. The reason for selection of SLIC is that it has few parameters that are flexibly tuned. Hence, the trade-off between those parameters and boundary adherence can be controlled. The SLIC method is also computationally and memory efficient. The initial stage of the SLIC superpixel segmentation is gridding the natural colour image into equally sized arbitrary patches such as rectangles or squares. In the case of MRI FLAIR images, the initial grids are generated for each slice separately. In (Achanta *et al.*, 2012), it is suggested to use squares as the initial SP grid for natural images. The reason for this assumption is that the aspect ratio of the pixel dimensions in natural images is equal to 1, which means that the height and width of each pixel is the same. The MR images of the brain have identical voxel dimensions in the $X$ and $Y$ directions in each slice. Therefore, the initial grids are considered as squares with the side size of $S$. The geometrical centre of each initial segment is considered as the superpixel centre which are then updated in every further iteration. Figure 4-5 illustrates the changes in the superpixel configuration in the iterations, from the initial to the final SP.

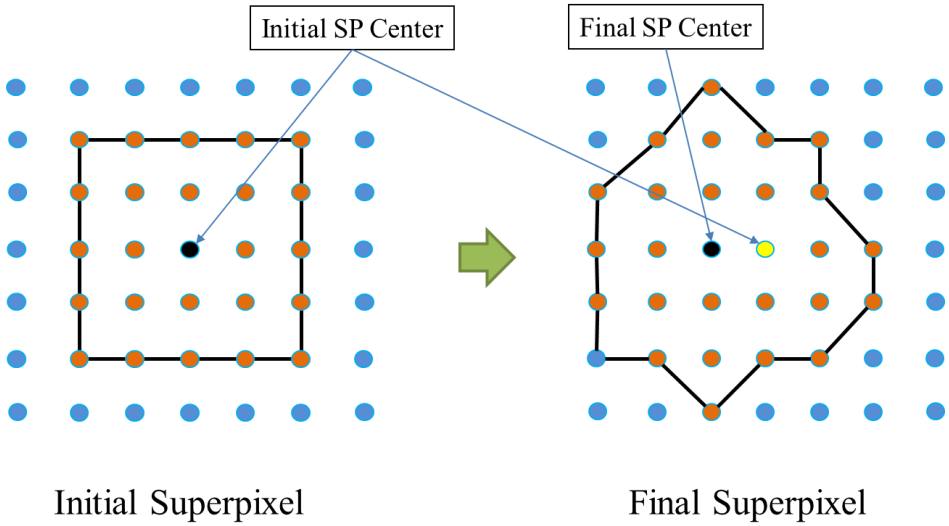



Figure 4-5 Clustering the homogenous pixels to one SP, initialling from a regular grid to the final homogenous superpixel. It should be noted that the centre of the SP may change in each iteration.

The pixels are clustered based on their spatial and intensity distance metrics. The spatial distance, $d_s$, between the $i$th and $j$th pixel is obtained using:

$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{4-1}$$

where, $x$ and $y$ are the pixel location coordinates. The intensity distance $d_c$ between the two pixels is defined as:

$$d_c = \sqrt{(I_j - I_i)^2} \tag{4-2}$$

where, $I_i$ and $I_j$ are the normalized intensity values of the $i$th and the $j$th pixel, respectively.

The overall distance measure $D$ is the combination of the spatial and intensity distances. It is calculated using:

$$D = \sqrt{d_c^2 + \left(\frac{d_s}{S}\right)^2 m^2} \tag{4-3}$$

where, $m$ is the compactness coefficient which determines the flexibility of superpixel boundaries. A higher value of $m$ increases the effect of spatial distance therefore results in more compact segments. A lower value decreases the effect of spatial distance and creates more flexible boundaries.

Figure 4-6 shows an illustration of the distance calculation for SP segmentation. The distances are calculated for a desired pixel, $P_i$, in order to assign the cluster label to that pixel. The search area is restricted around $P_i$, which is represented as the dashed square.

The way how compactness factor, $m$, affects the total distance, $D$ (Equation (4-3)), and therefore the SP boundaries, depends on the image intensity values (which determine $d_c$). The SP method proposed in (Achanta *et al.*, 2012) is optimised for natural images with the CIELAB colour space. The compactness factor in the range [1,40] is suggested for this type of images. MRI images have a various range of intensity values, which depend on the tissue and the image acquisition parameters. Therefore, it is difficult to set a generic range with the raw FLAIR voxel intensities. For this reason, the MRI image intensities used in Equation 2 are normalized to the values of [0, 1] to ensure that both the intensity and spatial distances are within the same range. This is also important in the optimisation process when a universal range or value of $m$ will be suggested which is applicable to all the new FLAIR images. SP segmentation with different compactness factors is shown in Figure 4-7 for a MR image acquired with the FLAIR protocol containing a Grade II tumour.

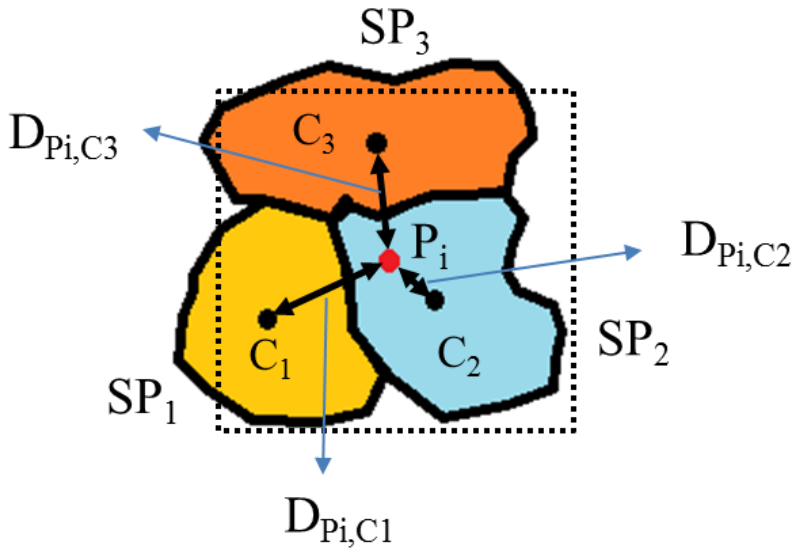



Figure 4-6 Illustration of distance in the SLIC-based superpixel algorithm. $SP_i$ presents the superpixel, $C_i$ the SP centre and $D_{pi,Cj}$ the distance between the desired pixel and the SP centres in the search area. The dashed square is the restricted search area around the desired pixel, $P_i$.

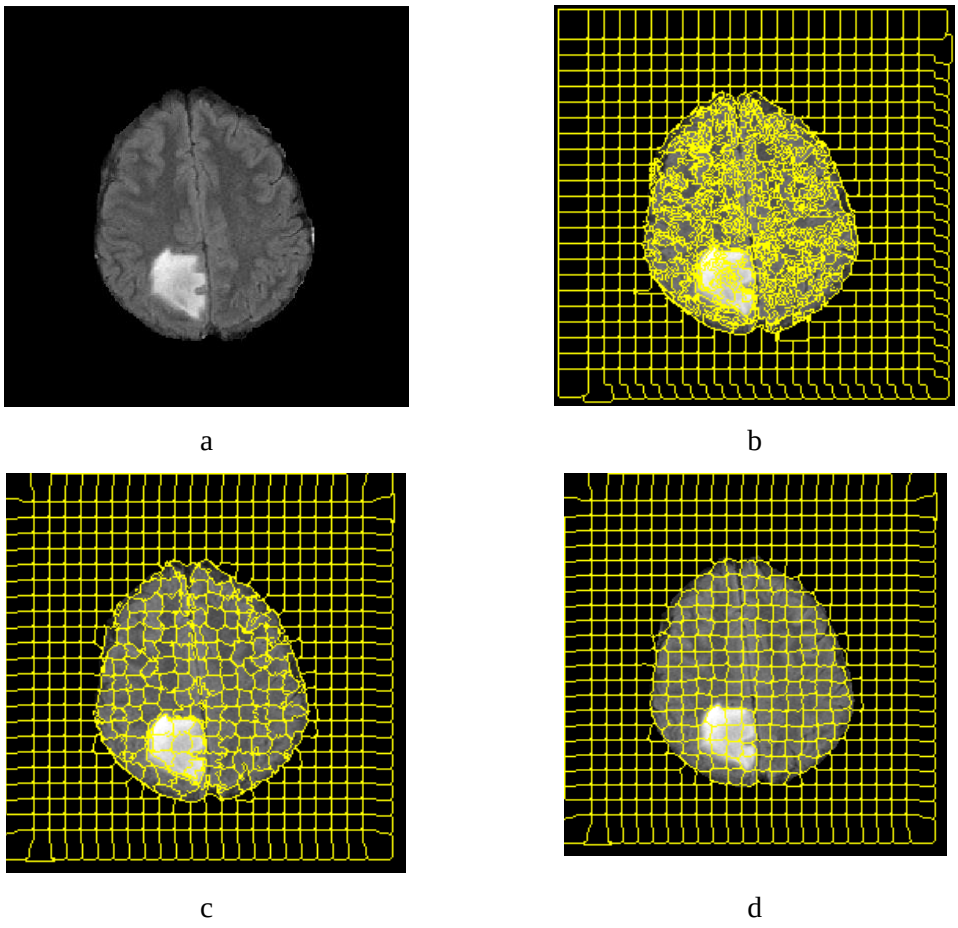



Figure 4-7 Superpixel segmentation for one slice of the MRI image different compactness factors: a) original MRI FLAIR image with a Grade II tumour, b) superpixel segmentation with $m$ = 0 and $S$ = 10, c) superpixel segmentation with $m$ = 0.2 and $S$ = 10, d) superpixel segmentation with $m$ = 0.5 and $S$ = 10.

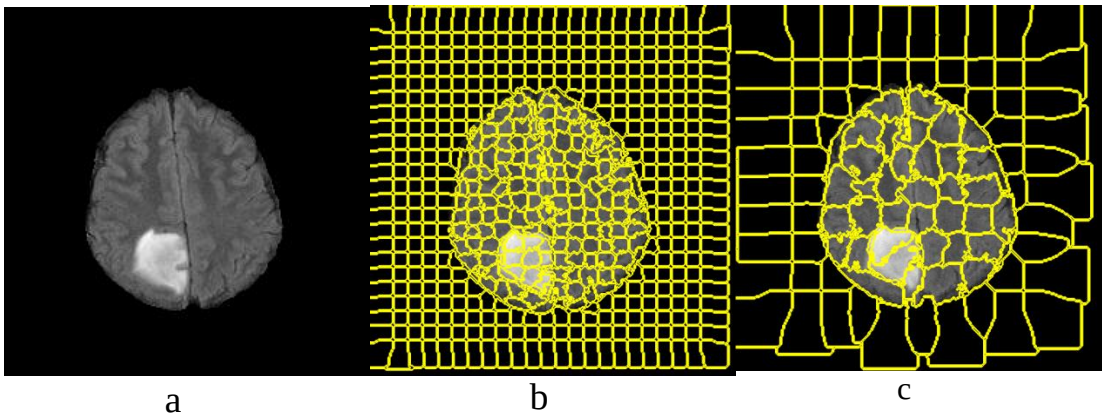



Figure 4-8 Superpixel segmentation for one slice of the MRI image with different window sizes: a) original MRI FLAIR image with a Grade II tumour, b) superpixel segmentation with $S$ = 10 (initial grids 10 x 10) and $m$ = 0.2, c) superpixel segmentation with $S$ = 20 (initial grids 20 x 20) and $m$ = 0.2.

Figure 4-8 shows the same image in Figure 4-7, which is partitioned separately to superpixels with two different side sizes, $S$. In Figure 4-8(b) and Figure 4-8(c), the superpixels are

extracted with $S = 10$ and $S = 20$, respectively. The compactness factor is fixed to $m = 0.2$ for both sizes to show only the effect of size parameter. A good partitioning occurs when the segmented regions include homogenous pixels, while the superpixel boundaries adhere to the edges in the image.

At the end of superpixel segmentation, some isolated pixels might appear. The label of these pixels is different from their surrounding pixels, which is considered as the noise of the superpixels. A post-processing procedure is designed to reduce the isolated pixels. The label of the pixels connected to the isolated pixels are counted. Then the isolated pixel is relabelled to the major connected class.

### 4.2.5 Feature Extraction

Feature extraction algorithms for medical image segmentation are categorised into intensity-based, texture, and shape features. Most of the features are considered as hand-designed, since feature extraction parameters are manually optimised for a specific task. Different types of features including intensity statistics, textons and curvature features will be considered to train a robust classifier for the detection and segmentation of brain tumour.

**Intensity statistical features**

First order intensity statistics (Jain, 1989) are referred as pixel-intensity based features. They express the distribution of grey levels within the selected region-of-interest (ROIs) which are the superpixels in the present work. For each superpixel, 16 features are calculated which will be explained in the following.

Average intensity feature of a superpixel, *SP*, is calculated using

$$Average(SP) = \frac{1}{N_P}\sum_{i=1}^{N_P} I_{SP,i} \,. \tag{4-4}$$

where $I_{SP,i}$ is the intensity value of pixel $i$ in the superpixel *SP*, and $N_P$ is the total number of pixels within the superpixel.

Standard deviation (STD) of intensities within the superpixel is calculated using

$$STD(SP) = \sqrt{\frac{1}{N_P - 1}\sum_{i=1}^{N_P} \left| I_{SP,i} - Average(SP) \right|^2} \,. \tag{4-5}$$

Variance of intensities within the superpixel is calculated using

$$Var(SP) = \frac{1}{N_P - 1}\sum_{i=1}^{N_P}\left|I_{SP,i} - Average(SP)\right|^2. \quad (4\text{-}6)$$

Coefficient of variance of the pixels is calculated using

$$CoV(SP) = \frac{STD(SP)}{Average(SP)}. \quad (4\text{-}7)$$

Skewness is a measure of asymmetry of the distribution of the intensities around the mean value of the superpixel. Skewness for a data with average $\mu$ and σ is derived from

$$Skewness = \frac{E(x-\mu)^3}{\sigma^3}, \quad (4\text{-}8)$$

where $E$ is the expectation operator. For the intensities of the pixels within a superpixels, skewness is calculated using

$$Skewness(SP) = \frac{\frac{1}{N_P}\sum_{i=1}^{N_P}\left(I_{SP,i} - Average(SP)\right)^3}{STD(SP)^3}. \quad (4\text{-}9)$$

Kurtosis is a descriptor of the shape of a distribution and a measure of tailedness of the distribution of the intensities. Kurtosis for a data generally derived from

$$Kurtosis = \frac{E(x-\mu)^4}{\sigma^4}. \quad (4\text{-}10)$$

For the intensities of the pixels within a superpixels, kurtosis is calculated using

$$Kurtosis(SP) = \frac{\frac{1}{N_P}\sum_{i=1}^{N_P}\left(I_{SP,i} - Average(SP)\right)^4}{STD(SP)^4}. \quad (4\text{-}11)$$

Maximum and minimum of the intensities within a superpixel are considered as $Max(SP)$ and $Min(SP)$, respectively which are included in the feature vector of that superpixel. The range value is calculated using

$$Range(SP) = Max(SP) - Min(SP). \quad (4\text{-}12)$$

Median and mode of the intensities of the pixels inside the superpixel are considered as $Median(SP)$ and $Mode(SP)$, respectively which are also included in the feature vector.

Mean of the absolute deviation is calculated using

$$MeanAD(SP) = \frac{1}{N_P}\sum_{i=1}^{N_P}\left|I_{SP,i} - Average(SP)\right|. \quad (4\text{-}13)$$

Median absolute deviation is calculated using

$$MedAD(SP) = Median\left(I_{SP,i}|_{i=1,\ldots,N_P} - Median(SP)\right) . \quad (4\text{-}14)$$

The third central moment is calculated using

$$Moment_3(SP) = E(x-\mu)^3 = \frac{1}{N_P}\sum_{i=1}^{N_P}\left(I_{SP,i} - Average(SP)\right)^3 . \quad (4\text{-}15)$$

Interquartile range is calculated using

$$Iqr(SP) = Q_3 - Q_1 , \quad (4\text{-}16)$$

where $Q_3$ is the upper quartile, i.e. the median of the lower half of the intensities, and $Q_1$ is the lower quartile, i.e. the median of the lower half of the data.

Entropy is calculated using

$$Entropy(SP) = -\sum_{j=1}^{N_b} p_j . log_2(p_j) \ , \quad (4\text{-}17)$$

where $p$ is the histogram count of the absolute superpixel values and $N_b$ is the number of histogram bins.

**Texton Feature**

Brain tissues have complex structures that include both normal and tumorous tissues. Therefore, intensity features are not sufficient to accurately detect and segment the tumour. To tackle this problem, texture features that work on a higher dimensionality are used to improve the accuracy of segmentation. Textons (Leung and Malik, 2001) are among the most powerful texture feature extraction (Arbelaez *et al.*, 2011) and are able to distinguish various patterns in the image. Textons are small elements of the image, generated by convolution of the image, $I$, with a specific filter bank ($F_1$, $F_2$, …, $F_{NF}$), i.e.

$$R = [F_1 * I, F_2 * I \ldots F_{NF} * I] , \quad (4\text{-}18)$$

where $NF$ is the number of filters in the filter bank and $R$ is the set of filter responses. Selecting the filter type and designing the filter bank is an important stage for texton analysis (Zhang *et al.*, 2016). Gabor filters provide strong textural descriptors by considering the local dependencies in both spatial and frequency domain (Grigorescu *et al.*, 2002). Therefore, Gabor filter (Henriksen, 2007) will be used in this work for texton feature extraction, which is defined as

$$G(x, y; \theta, \sigma, \lambda, \psi, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\,\sigma^2}\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right), \quad (4\text{-}19)$$

where, $\sigma$ is the standard deviation of Gaussian envelope, $\gamma$ is the spatial aspect ratio, $\lambda$ is the wavelength of sinusoid and $\psi$ is the phase shift. In Equation (4-19), the terms $x'$ and $y'$ are calculated from the spatial orientation of the filter, $\theta$, defined as

$$\begin{aligned} x' &= x\cos\theta + y\sin\theta, \\ y' &= -x\sin\theta + y\cos\theta. \end{aligned} \quad (4\text{-}20)$$

The values that are set for these parameters will be discussed in Section 4.3.3.

Figure 4-9 shows a set of Gabor filters with different size, directions, and wavelengths of the sinusoid. For more detailed representation, the kernels in the filter bank are categorised based on different configurations of parameters.

Assuming the number of filters in the filter bank is $N_{FB}$, the FLAIR image is convolved with all the filters, hence a response vector with length of $N_{FB}$ is generated for each pixel. Figure 4-10 shows the filter responses generated from convolution of the FLAIR image with the Gabor filters with different parameters. The parameters (i.e. size, direction, and sinusoid wavelength) are separated in order to better illustrate the effect of each parameter on the response.

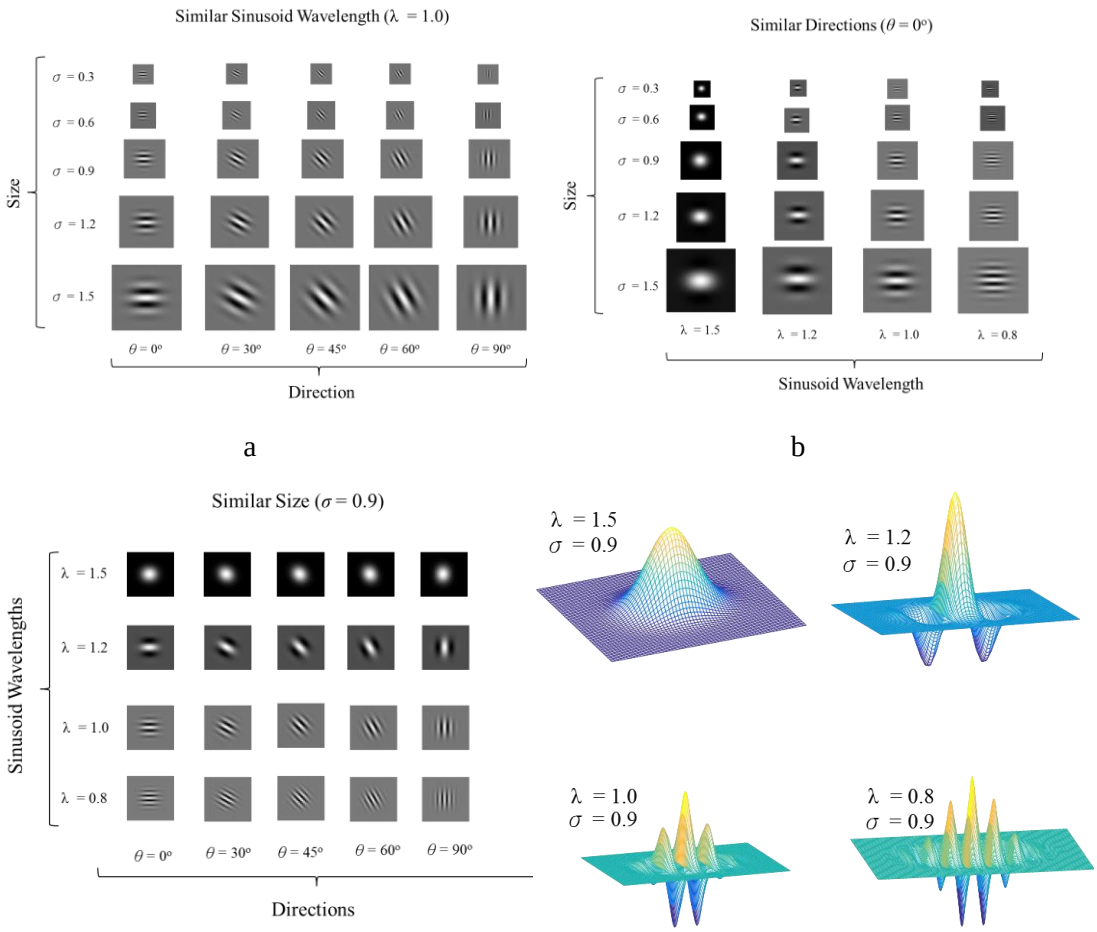



Figure 4-9 Set of Gabor filters which are used for texton feature extraction with different parameters. a) similar sinusoid wavelengths and different sizes and directions, b) similar directions and different sizes and sinusoid wavelengths a) similar sizes and different sinusoid wavelengths and directions, d) 3D representation of Gabor kernels with different sinusoid wavelengths.

The number of the filter response vectors is the same as the number of the pixels in the image. The texton maps are created from the filter bank responses by applying *k*-means clustering which is $N_{FB}$ dimensional. The number of clusters $k_{texton}$ is chosen empirically based on the number of tissues. The major tissues in a MR image of brain with tumour include WM, GM, CSF, core and oedema. Each texton is assigned a texton *ID* based on the cluster number (i.e. k = [1, 2, …,5]). The texton map is a greyscale image with values ranging in the *k* = [1, 2, …,5]. Figure 4-11 shows the process of texton map extraction. The texton feature for superpixels is defined as histogram of the texton *ID*s within that SP. The *ID*s are then sorted ascendingly based on the average FLAIR intensity of the group of pixels within each cluster. An example of the texton map and the corresponding texton histogram is illustrated in

Figure 4-11. As can be seen, the texton *ID* histogram of superpixels related to tumour are different from those of a normal brain.

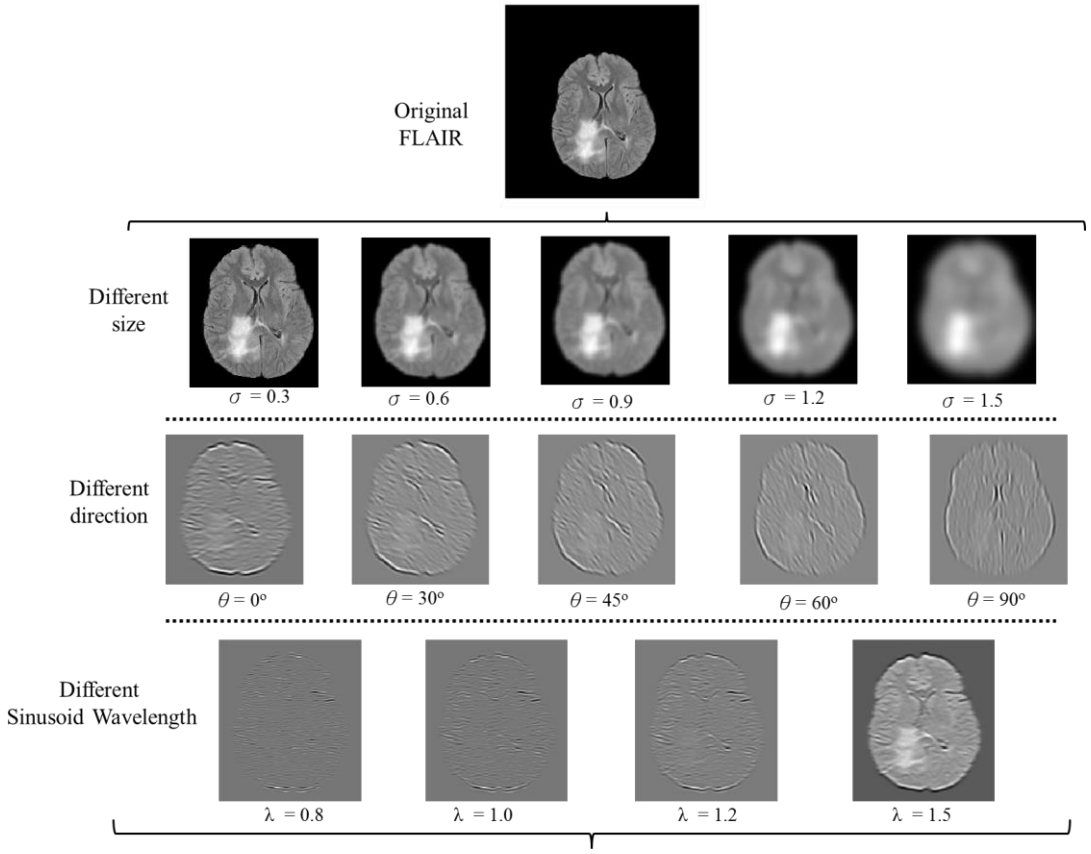



Figure 4-10 Filter responses obtained by convolving the image with the Gabor kernels in the filter bank separately for different size, direction and sinusoid wavelengths.

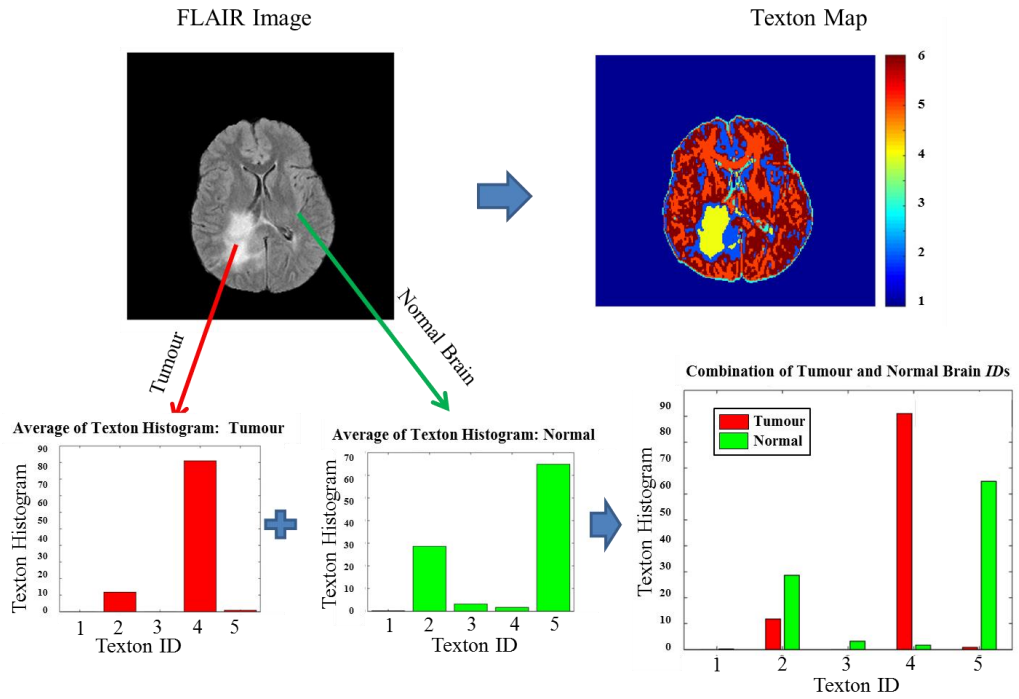



Figure 4-11 Example of calculating texton *ID*s for normal brain and tumour. The plots present the average texton histogram of the superpixels inside each region, i.e. tumour and normal brain. It should be noted that the *ID*s are sorted based on the initial *k*-means cluster points. This is an illustration example and later the clusters will be sorted ascendingly based on the average intensity value of the clusters.

**Fractal Features**

Fractal features are calculated based on a segmentation based fractal texture analysis method (SFTA) (Costa *et al.*, 2012). In this method, the image is decomposed into a set of binary images based on multi-level thresholds which are computed using the Otsu algorithm (Liao *et al.*, 2001). The number of thresholds $N_{threshold}$ is defined by the user which is the tuneable parameter of the fractal analysis. For single modality MRI data, $N_{threshold} = 3$ is selected which will be discussed in Section 4.3.3. Thereafter, all the image boundaries are extracted for each binary channel using edge detection (Canny, 1986). The fractal features are calculated from these binary edge channels which include area, intensity and fractal dimension. Area feature is the number of edge pixels in a superpixel. Intensity feature is the mean intensity of image pixels corresponding to the edge pixels in a superpixel. Fractal dimension represents the complexity of the structure of the image and is calculated from image boundary using

$$D_0 = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log \varepsilon^{-1}}, \tag{4-21}$$

where $N(\varepsilon)$ denotes the counting of hyper-cubes (rectangles in the case of 2D space) of dimension $E$ and length $\varepsilon$. An approximation of fractal distance is obtained from the binary images using box counting algorithm (Schroeder, 2009).

The flowchart of fractal analysis is depicted in Figure 4-12. Figure 4-13 shows fractal features including: area, mean intensity and fractal dimension. Figure 4-14 shows an example of fractal dimension and mean intensity features calculated from healthy and tumour superpixels from one patient data containing a Grade IV glioma. It demonstrates a good separation in feature space (mean intensity-fractal dimension) for FLAIR images.

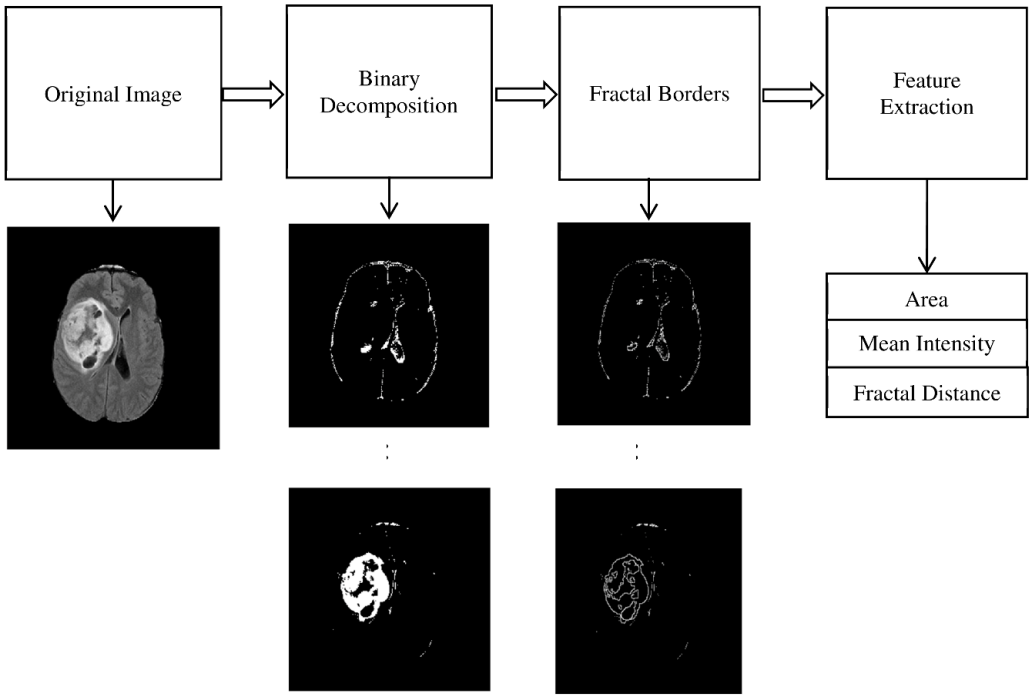



Figure 4-12 The flowchart of extracting fractal features from a grade III glioma.

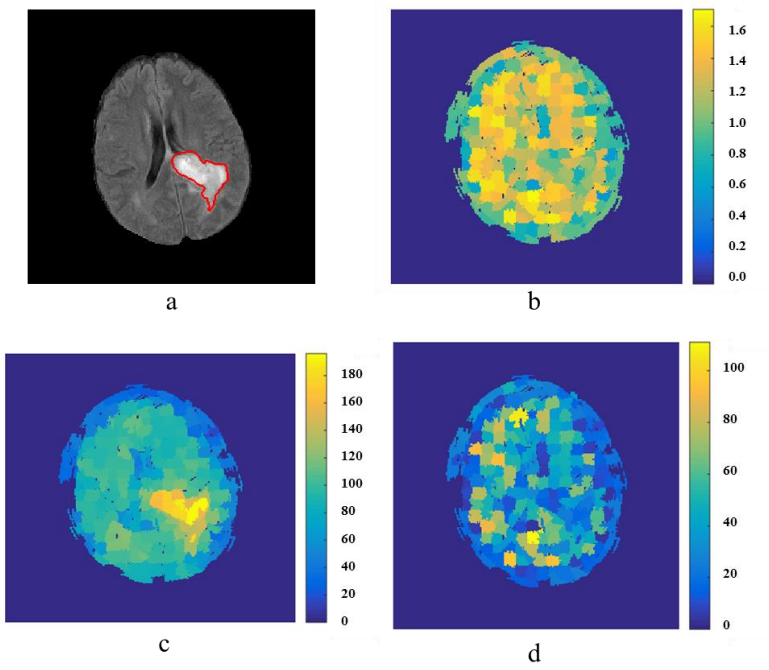



Figure 4-13 An example of fractal analysis applied to a Grade III glioma to generate superpixel based fractal feature maps: a) FLAIR image with the ground truth of oedema, b) area, c) mean intensity, d) fractal dimension.

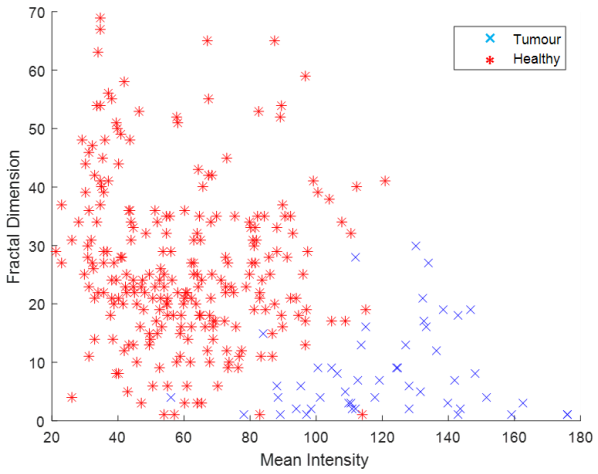



Figure 4-14 Fractal dimension vs. mean intensity for healthy and tumour superpixels calculated from one FLAIR MRI data with Grade IV glioma.

**Curvature Feature**

Image curvature is a shape-based feature which is computed by the derivatives along x and y directions of an image, $f_x$ and $f_y$. The image normal at pixel $(x, y)$ is calculated using (Arridge, n.d.):

$$\widehat{\boldsymbol{N}}(x,y) = \frac{1}{\left(f_x^2+f_y^2\right)^{1/2}}\begin{pmatrix} f_x \\ f_y \end{pmatrix}. \tag{4-22}$$

The two-dimensional curvature of the image is the divergence of the normal in Equation (4-22) and is calculated using

$$\text{Curv} = \frac{f_{xx}f_y^2+f_{yy}f_x^2-2f_{xx}f_xf_y}{\left(f_x^2+f_y^2\right)^{3/2}}, \tag{4-23}$$

where, $f_{xx}$ and $f_{yy}$ are the second derivatives of the image intensity $I(x, y)$. The curvature feature for each superpixel is the average of the curvature values for all the pixels in the superpixel. In the case of $f_x = f_y = 0$, a *null* value will be assigned to the curvature feature.

Table 4-1 Total number of features calculated from an MRI FLAIR image

| Feature name | Number of features |
|---|---|
| Statistical 1st order | 16 |
| Texton Histogram | 5 |
| Fractal | 6 |
| Curvature | 1 |
| Total | 28 |

In total, 28 features were calculated for each superpixel. The feature vector includes 5 texton histogram features from 5-clusters and 6 fractal features obtained from 3 thresholded binary images (each binary image provides 3 fractal features). All the features are normalized to the range of [0,30], except the 5 texton histogram features. The reason for selecting 30 is that the average number of pixels in the superpixels is approximately 30, which is also the maximum value for the texton histogram counts. This is to ensure that all the features have similar dynamic ranges and are close to the texton histogram values. Table 4-1 shows a list of the features. The details of parameter setting in feature calculation will be discussed in Section 4.3.3.

### 4.2.6 Feature Selection

The primary objective of feature selection is to obtain maximum accuracy with the minimum set of features. This is accomplished by effectively removing irrelevant and redundant features that may cause more classification error. Feature reduction will also decrease the computational complexity and time. At the same time, by selecting the most relevant features,

the generalisation will enhance. Feature selection methods are categorised into filter, wrapper, and hybrid approaches (Rosario and Thangadurai, 2015). In filter-based techniques, the criteria for feature selection is independent of the output class of the data (Sánchez-Maroño *et al.*, 2007). The advantages of filter based methods are their speed and independency from the classifier (Saeys *et al.*, 2007). However, since the relation between variables are not considered in the filter methods, more redundant features may be selected (Hamon *et al.*, 2013). In wrapper techniques, the subset of selected features are evaluated with a learning algorithm (Li *et al.*, 2011) and the relations between variables are taken into account. However, the wrapper methods have the problem of overfitting in the case of insufficient data samples. Furthermore, for the large dataset, the number of variables and their intervariable relations increase significantly which make the wrapper process time consuming.

Some popular feature selection methods are explained in the following. The fast correlation-based filter (FCBF) (Yu and Liu, 2003) method select features that are highly correlated with the output, while not correlated to each other. ReliefF (Kononenko, 1994) uses a statistical approach in which the features are assigned a weight regarding to their relevance to the class. A higher value of the weight represents more predictivity of the feature. Sparse logistic regression (SLogReg) (Shevade and Keerthi, 2003) identifies a sparse subset of discriminative features. Sparse logistic regression with Bayesian regularisation (SBMLR) (Cawley and Talbot, 2006) is a developed version of SLogReg, in which a Bayesian approach is integrated to makes it faster and more efficient without needing model selection for optimisation of parameter selection. Spectral feature selection (SPEC) (Zhao and Liu, 2007) is based on spectral graph theory in which the relevance of the features is related to the consistency with the structure of the graph. The minimum redundancy maximum relevance (mRMR) (Peng *et al.*, 2005) feature selection technique selects the most relevant features by removing the irrelevant ones which makes it an efficient technique for subset selection of features. This method uses the mutual information to identify the similarity between features.

In this work, the mRMR feature selection technique (Peng *et al.*, 2005) is chosen for feature subset selection due to the efficiency of the method in trading off between relevancy and redundancy amongst the filter-based techniques. A comparison of the feature selection methods for the proposed techniques in this chapter will be presented in the Section 4.3.3.

For features, $f_i$, in feature set $S$, the maximum relevance between features and class $c$ is obtained by maximising

$$\max\ D(S,c)\ ,\ D = \frac{1}{|S|}\sum_{f_i \in S} I_M(f_i; c), \tag{4-24}$$

where $I_M$ is mutual information between feature $f_i$ and the class $c$. Minimum redundancy is calculated from:

$$\min R(s),\ R = \frac{1}{|S|^2}\sum_{f_i,f_i \in S} I_M(f_i, f_j)\ . \tag{4-25}$$

The feature selection is performed on the entire feature vector which is based on leave-one-out cross validation using a voting scheme. In the cross validation, the best number of features, $N_{FEA}$, were selected for each case. When a specific feature is selected by one case, that feature will get one vote. For all the features voted by all the cases, the top $N_{FEA}$ features with highest scores will be chosen as the final features. The selected features will be used in the classification stage to classify each superpixel into tumour or non-tumour.

### 4.2.7 Extremely Randomized Trees Classification of Superpixels

The ERT classifier (Geurts *et al.*, 2006) is used to categorise each superpixel into tumour or normal brain tissue to tackle the problem of extremely imbalanced data in the clinical dataset and to improve the accuracy of the minority class (i.e. tumour). Like random forests (RF) (Liaw and Wiener, 2002), ERT is an ensemble technique which uses multiple decision trees. Each node of the tree in both methods includes a set of training examples and the predictor. Splitting starts from the root node and will continue at every node. The procedure is performed based on the feature representation and allocating the partitions to the sub-nodes. The trees are growing until a specified tree depth, $D_{tree}$, is reached. A random subset of features is selected at each attribute split during the bagging process. In the RF classifier, the most popular class is voted (Breiman, 2001) after generating large number of trees, $N_{tree}$.

ERT is an extension of RF in which a further randomisation stage is added for selecting the cut-points alongside with randomised selection of attributes like in RF. The splits of attributes and cut-points are selected randomly in ERT. Each tree is determined by $t \in \{1 \ldots T\}$ in which $T$ is the number of randomised trees. For a given data point $x$ and dataset $D_{train}$ a feature vector is represented by $f(x, D_{train})$. To classify the class $c$ of the data, for an n-dimensional feature representation, each tree learns a weak predictor of $p_t(c|f(x, D_{train}))$.

In the testing process, for an unseen data point, $x'$, the probability of belonging a class $c$ is calculated by the average of probabilities on all the trees using

$$p(c|f(x', D)) = \frac{1}{T}\sum_{t=1}^{T} p_t(c|f(x', D)). \tag{4-26}$$

The structures of randomised trees are independent of training sample outputs. The main parameters of designing ERT are the number of trees, tree depth, and the number of attributes ($k_{attribute}$) which is selected to perform the random split. RF parameters were tuned by examining different tree depths and number of trees on clinical training datasets and evaluating the classification accuracy using leave-one-out validation. In the current study, there are 20 extra trees with depth $D_{tree}$ = 15 in the ensemble and five attributes which are equal to the number of selected features. The minimum number of samples for splitting a node is 2 as this is a classification task. Setting these parameters will be discussed in Section 4.3.3.

Each superpixel is now classified into a tumour or non-tumour candidate. Based on the classes assigned for each voxel in the testing case in each leave-one-out iteration, the final segmentation mask is created by mapping back the pixel class to the segmentation mask volume. The small superpixel regions with the total number of pixels less than a pre-defined threshold (i.e. 100) are considered as a false positive (FP) region and removed from the tumour candidates. This is a post-processing stage to remove the hyper-intense regions of skulls near normal brain. The remaining tumour superpixel regions are the segmented tumour.

## 4.3 Experiments and Results

Two experiments were conducted in this section based on the datasets. In the first experiment, the clinical dataset was used for training and validation of the algorithm. In the second experiment, to assess the robustness of the proposed method, a further validation on the publicly available BRATS 2013 training dataset was conducted. In this section the dataset, implementation, parameter setting of the models, and evaluation protocols are described. Then comprehensive experimental results are presented and discussed in terms of quantitative and qualitative evaluations.

### 4.3.1 Dataset and Implementation

Data acquisition and parameters for the clinical FLAIR sequences were explained in Chapter 2 (Section 2.6.1). The clinical dataset consists of 19 patients which are entered retrospectively into the experiments, each with a brain tumour with grades: II, III, and IV, and each of them has a histological gold standard. Figure 4-15 shows some examples of the manual segmentations of the complete tumour for different tumour grades in FLAIR images from the clinical dataset.

The BRATS 2013 annotated clinical training dataset were described in Chapter 2 (Section 2.6.2). The dataset consists of multi-contrast MR scans of 30 glioma patients which were acquired from multi-centres using different scanners with different field strengths (1.5 T and 3T). In the current study, only FLAIR images were used to evaluate the proposed single modality method. As explained in Chapter 2 (Section 2.6.2), VSD system provided the tumour type classes: oedema, necrosis, enhancing and non-enhancing tumour.

The experiments were carried out using the following combinations of the tumour structures for image labelling and classification:

1. Complete tumour.
2. Healthy brain tissues and background.

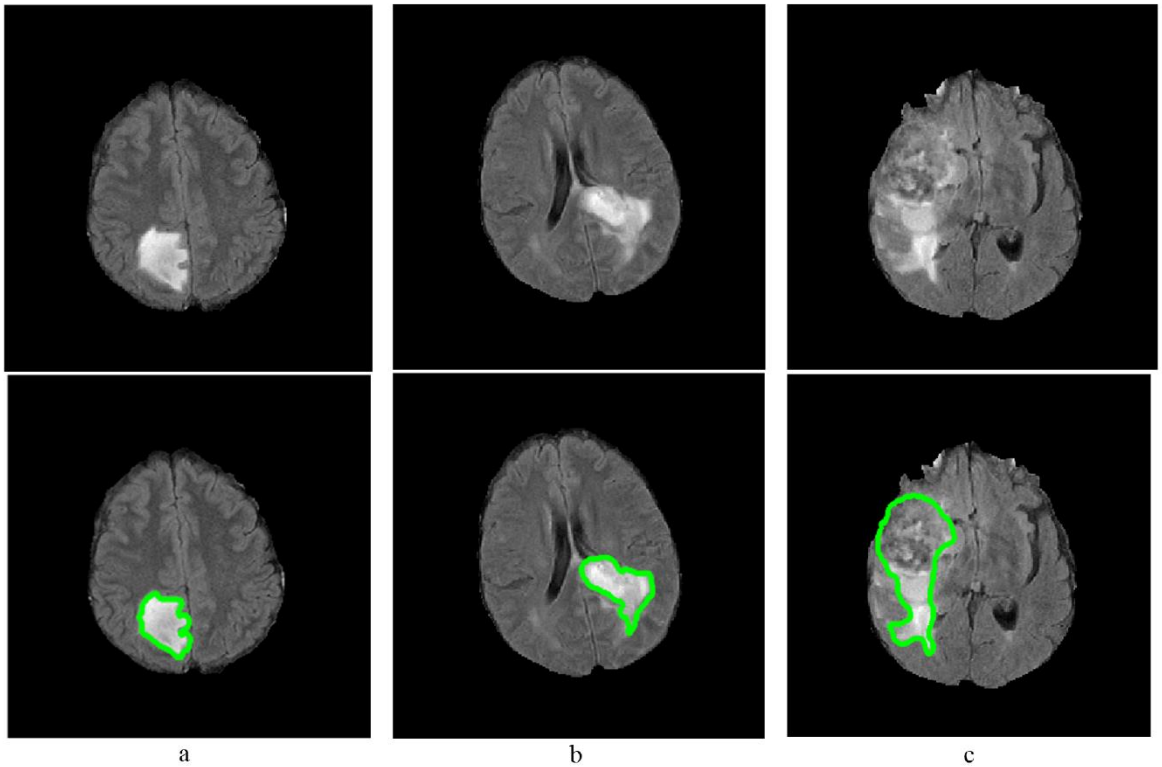



Figure 4-15 FLAIR images with different tumour grades in upper row and their ground-truth manual segmentation of the FLAIR hyperintensity in the lower row. Tumour grades are: a) Grade II b) Grade III and c) Grade IV

The reason that the BRATS testing dataset were not used is that the VSD evaluation blind test protocols were designed for analysing the segmentations with the full label classification.

Currently the aim of this work is to determine the complete tumour region which is a binary classification problem. Therefore, using VSD which is designed for a specific multi-class evaluation (Chapter 2, Section 2.7.1), is not applicable for the current binary classification experiments. The evaluation of testing dataset using VSD blind test system will be carried out later in Chapter 6 for segmentation of all tumour tissue subtypes.

The methods were performed on MATLAB 2015b on a PC with CPU Intel Core i7 and 16 GB RAM with the operating system Windows 8.1. The ERT classifier was implemented using the open source code provided in (Taormina, n.d.) which was created based on the ERT method proposed by Geurts *et al.* (Geurts *et al.*, 2006).

To evaluate the performance of ERT classifier, it was compared with support vector machine (SVM) (Furey *et al.*, 2000) for the classification of superpixels. The reason for selecting SVM for this comparison is that it is a powerful binary classification which makes it an appropriate candidate to assess the performance of ERT for the two-class single modality segmentation. The SVM-based method will be referred as SP_SVM in this section.

The application of the automated single modality method on the clinical data was evaluated by comparing the segmentations with the manual annotation provided by an expert. The BRATS 2013 dataset experiments were compared with the manual annotation provided by the VSD system.

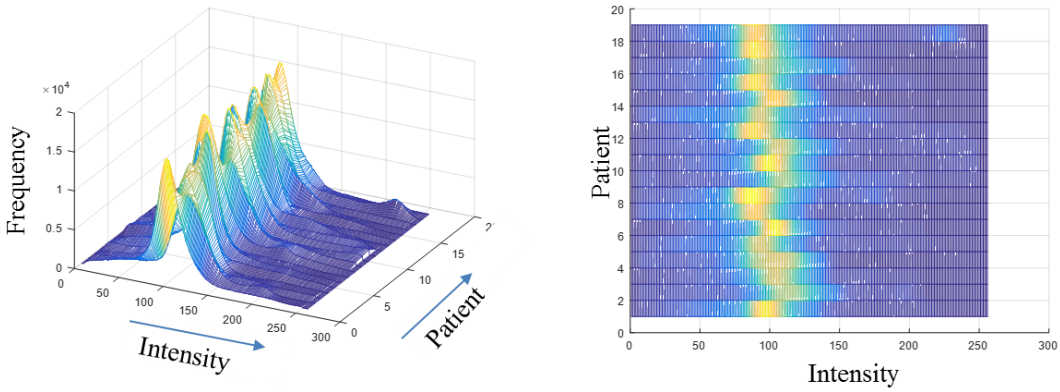
The SVM and ERT classifiers were evaluated using standard classification evaluation measure, i.e. sensitivity (Equation (2-16)), specificity (Equation (2-17)) and BER (Equation (2-18)). The segmentations were quantitatively evaluated using DSC (Equation (2-19)).

### 4.3.2 Preprocessing

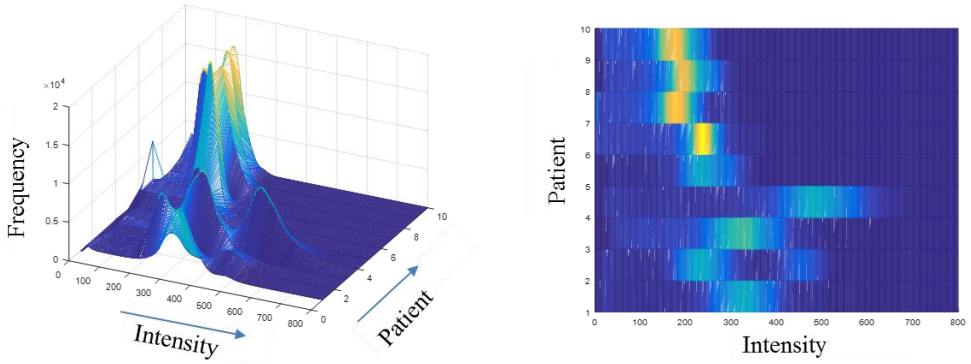
In the clinical dataset experiments, all data were acquired from the same scanner. Figure 4-16(a) shows the original histograms for all the 19 patient data. As can be seen, the data are quite consistent and there are similar histogram distributions across the different subjects.

Therefore, in the initial study, instead of applying histogram matching, during the feature calculation stage, all the features (e.g. 1st order intensity, fractal and curvatures), except the 5 texton histogram features, are normalised to the range of [0,30], which is referred as partial normalisation. This is to ensure that all the features have similar dynamic ranges and are close to the textons histogram values. However, for BRATS dataset, there are a large inconsistency between histograms across different patients as shown in Figure 4-16(b) (low-grade) and

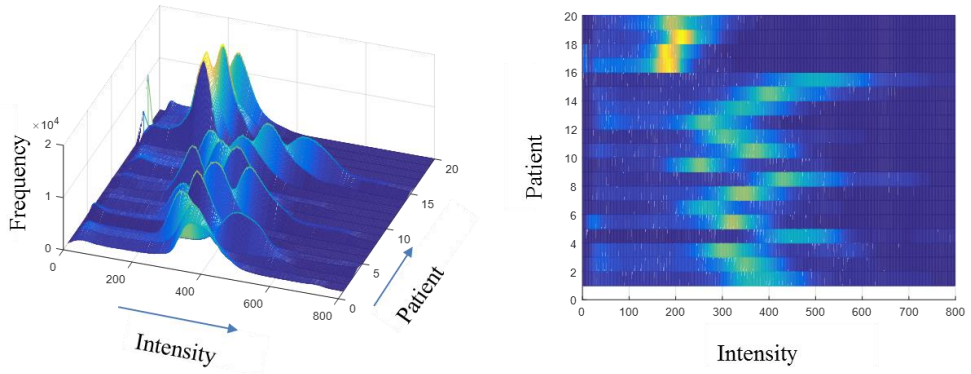
Figure 4-16(c) (high-grade). Therefore, the MRI image histogram normalisation stage is essential.




a




b




c

Figure 4-16 FLAIR MRI image histograms for: a) the clinical data original histograms (19 patients), b) BRATS 2013 LGG original histograms (10 patients), c) BRATS 2013 HGG original histogram (20 patients). HGG and LGG are separated for better illustration of the histogram plots.

Table 4-2 Dice overlap comparison results for the BRATS data with histogram normalisation and without histogram normalisation (but normalising the feature ranges, so it is called "partial normalisation" in this table).

| Case No | Grade/ID | SP_SVM | | SP_ERT | |
|---|---|---|---|---|---|
| | | Partial normalisation | Histogram matching | Partial normalisation | Histogram matching |
| 1 | LG-01 | 0.84 | 0.85 | 0.89 | 0.89 |
| 2 | LG-02 | 0.93 | 0.93 | 0.95 | 0.95 |
| 3 | LG-04 | 0.71 | 0.78 | 0.8 | 0.87 |
| 4 | LG-06 | 0.79 | 0.84 | 0.87 | 0.91 |
| 5 | LG-08 | 0.86 | 0.88 | 0.91 | 0.92 |
| 6 | LG-11 | 0.85 | 0.86 | 0.88 | 0.89 |
| 7 | LG-12 | 0.88 | 0.88 | 0.92 | 0.92 |
| 8 | LG-13 | 0.71 | 0.75 | 0.79 | 0.81 |
| 9 | LG-14 | 0.81 | 0.80 | 0.86 | 0.84 |
| 10 | LG-15 | 0.45 | 0.78 | 0.54 | 0.88 |
| 11 | HG-01 | 0.88 | 0.89 | 0.91 | 0.92 |
| 12 | HG-02 | 0.83 | 0.83 | 0.88 | 0.88 |
| 13 | HG-03 | 0.8 | 0.82 | 0.89 | 0.91 |
| 14 | HG-04 | 0.91 | 0.90 | 0.92 | 0.92 |
| 15 | HG-05 | 0.5 | 0.74 | 0.53 | 0.78 |
| 16 | HG-06 | 0.72 | 0.79 | 0.84 | 0.91 |
| 17 | HG-07 | 0.7 | 0.78 | 0.78 | 0.85 |
| 18 | HG-08 | 0.89 | 0.89 | 0.91 | 0.91 |
| 19 | HG-09 | 0.85 | 0.86 | 0.87 | 0.89 |
| 20 | HG-10 | 0.5 | 0.65 | 0.59 | 0.71 |
| 21 | HG-11 | 0.83 | 0.87 | 0.87 | 0.92 |
| 22 | HG-12 | 0.85 | 0.88 | 0.88 | 0.91 |
| 23 | HG-13 | 0.78 | 0.81 | 0.86 | 0.89 |
| 24 | HG-14 | 0.78 | 0.86 | 0.82 | 0.90 |
| 25 | HG-15 | 0.71 | 0.78 | 0.83 | 0.91 |
| 26 | HG-22 | 0.79 | 0.84 | 0.82 | 0.88 |
| 27 | HG-24 | 0.76 | 0.85 | 0.8 | 0.89 |
| 28 | HG-25 | 0.75 | 0.84 | 0.89 | 0.90 |
| 29 | HG-26 | 0.55 | 0.75 | 0.61 | 0.79 |
| 30 | HG-27 | 0.62 | 0.81 | 0.69 | 0.91 |
| **Mean** | **All** | **0.76** | **0.83** | **0.82** | **0.88** |
| **STD** | **All** | **0.13** | **0.06** | **0.11** | **0.05** |

The comparison experimental results for the BRATS data with histogram normalisation and without histogram normalisation (but normalising the feature ranges used in the initial study) are shown in Table 4-2. This shows that the histogram normalisation improves the segmentation performance by using both SP_SVM and SP_ERT.

### 4.3.3 Selection of Parameters

The superpixel segmentation parameters should be optimised to ensure obtaining accurate boundary patches in a reasonable time with sufficient number of pixels within each superpixel for feature calculation. Statistical features are non-parametric and calculated directly from the

intensity values of the pixels within the superpixels. Parameter setting is required to calculate other features, i.e. fractals and textons. Optimum parameters for the ERT classifier should be selected for an accurate and fast classification. In this study, the parameters are determined through the training stage; in which a total number of 6 patient data is randomly selected including 2 Grade II, 1 Grade III and 3 Grade IV which are referred as the training parameter tuning dataset. The testing procedure was performed on the whole dataset with a leave-one-out approach. In the following sections, the parameters selection procedure will be explained in detail.

**Superpixel Parameters**

The first stage is to investigate the effect of compactness factor, *m*, which was defined in Equation (4-3). As discussed in Section 4.2.4, the intensity values of the FLAIR voxels within the brain are normalised to the range of [0, 1] to ensure a normalised value for *m* in different patient images. Different values from 0 to 1 with step 0.1 were applied and the superpixels boundaries were inspected visually. A compactness factor $m = 1$ increases the weight of location distance in Equation (4-3), hence results in more rigid boundaries. On the other hand, $m = 0$ decreases the effect of location distance, therefore a higher weight of intensity distance produces very flexible boundaries but increases the variation and irregularity of the superpixels' shapes. Examples of varying the compactness factor on the superpixel patches are shown in Figure 4-18. Different values for *m* were examined by visually inspecting the superpixel boundaries and area for some slices from the selected patients. The value of $m = 0.2$ presents more coherent boundaries. For more accurate optimisation of the parameter *m*, a quantitative approach such as a distance measure (e.g Hausdorff distance) between superpixel boundaries and the ground truth can be used.

The next parameter that must be optimised is the superpixel size and is determined by the initial superpixel grid size. Different initial window side sizes are considered in the optimisation stage which are mentioned in Table 4-3. The compactness factor is fixed to $m = 0.2$ for all the experiments. The superpixels that include more than 90% pixels from the tumour in the manual segmentation mask are selected. The Dice measure is used for assessing the performance of superpixel segmentation. The experiment ran on the training parameter tuning images, which are from different tumour grades, and the average results are presented in Table 4-3. The results show that increasing the superpixel size decreases the segmentation accuracy. A superpixel size of $S = 6$ is chosen which has a good performance and meanwhile contains sufficient information within the superpixel for calculating the texture based features.

It should be noted that the above-mentioned parameters were optimised independently. However, using joint optimisation may provide more optimised parameters. For the superpixel size $S$ = 6, the overlap measure test was applied to the superpixels with different $m$ values that include more than 90% pixels from the tumour in the manual segmentation. The noise level was considered as the average of the total number of isolated pixels. Figure 4-17 shows both plots of the overlap measure and noise. As it can be seen, the compactness $m$ = 0.2 results in a low noise value while having high accuracy.

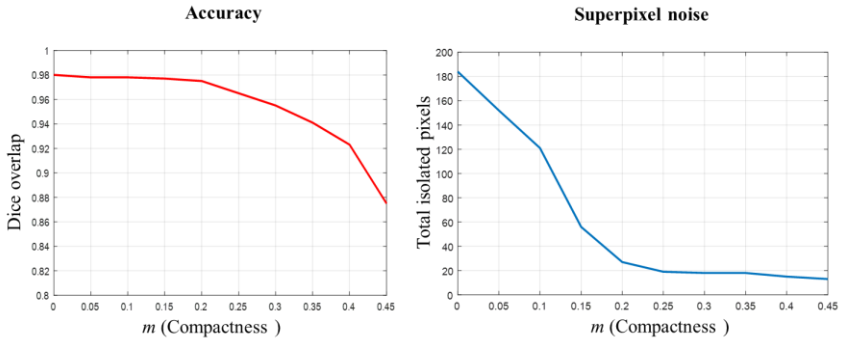



Figure 4-17 Comparison of accuracy and superpixel noise for superpixels with $S$ = 6 and different $m$ values in the range [0, 0.45] with the step 0.05.

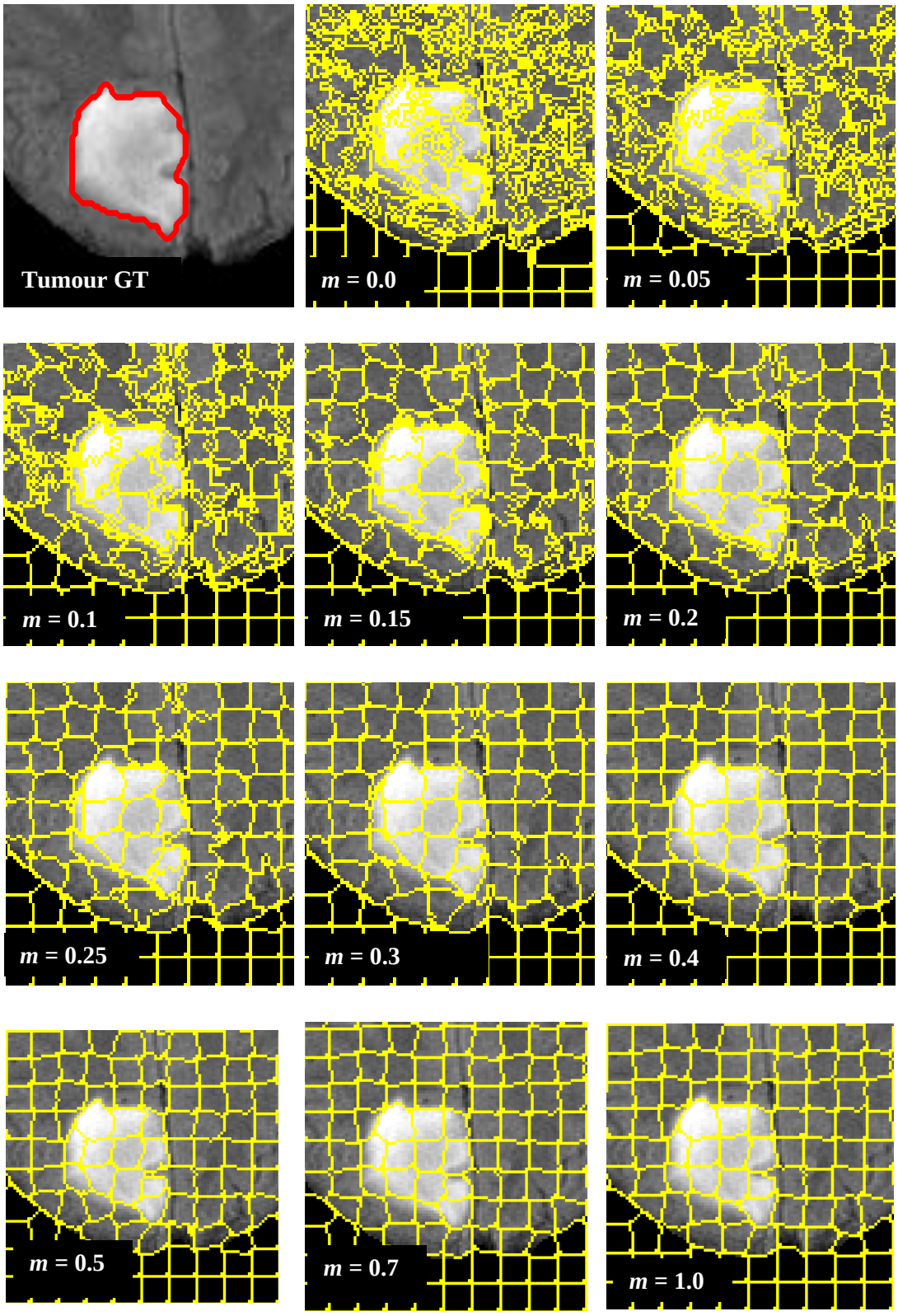



Figure 4-18 Superpixel segmentation with $S$ = 10 and different compactness factors; $m$ = [0.0, 0.5, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.7, and 1.0]. The upper left image is the original FLAIR overlaid with the tumour ground truth, which is the close-up of Figure 4-7.

Table 4-3 Examples of the impact of different initial superpixel side sizes, *S*, on the segmentation accuracy of the tumour in FLAIR images with compactness factor $m = 0.2$

| **Superpixel Side Size** | 4 | 6 | 8 | 10 | 15 | 20 |
|---|---|---|---|---|---|---|
| **Dice Overlap** | 0.98 | 0.96 | 0.92 | 0.85 | 0.73 | 0.56 |

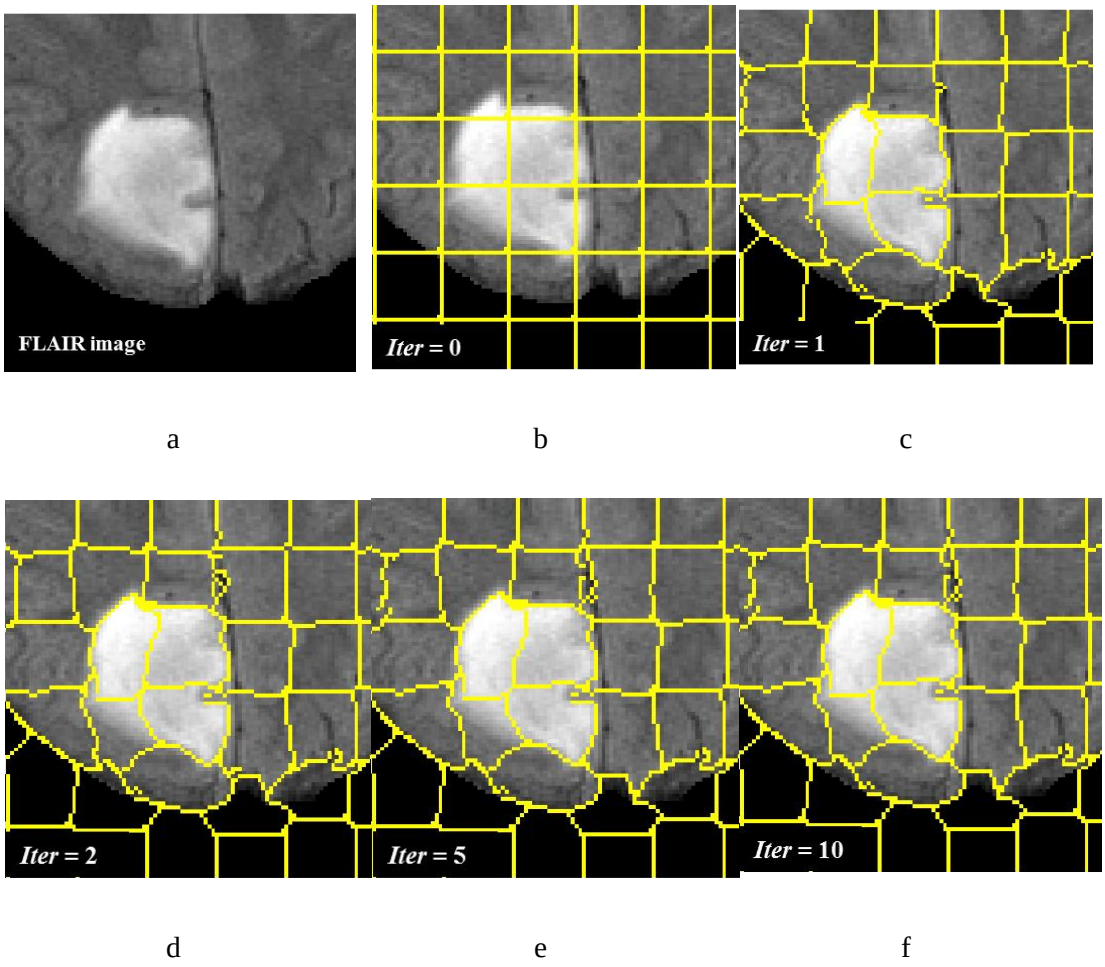



Figure 4-19 SP segmentation with different iterations; i.e. *Itr* = [0, 1, 2, 5, 10].

Another parameter which is considered in iterative superpixel segmentation is the number of iterations for updating of the clusters. Figure 4-19 shows the superpixel segmentation boundaries after different iterations. As it can be seen, after the second iteration, no considerable changes are visible in the boundaries. However, the number of iterations, *Itr* = 10 is considered to ensure the convergence of the SPs for all the patient images which is also suggested by (Achanta *et al.*, 2012). The experiments show that the SLIC method converges to a stable superpixel boundary in a few number of iterations. However, it might depend on the complexity and homogeneity of the images. To evaluate the effect of the number of iterations on the superpixels, the difference between superpixel labels from consequent

iterations is calculated. The superpixel map from an iteration is compared to the one from the previous iteration by counting the number of pixels with different superpixel labels. Figure 4-20 shows the superpixel difference for 10 iterations for three sample cases and the average of all the cases. It can be seen that from *Itr* = 8, there is no difference between iterations. However, the *Itr* = 10 is considered as a safe margin.

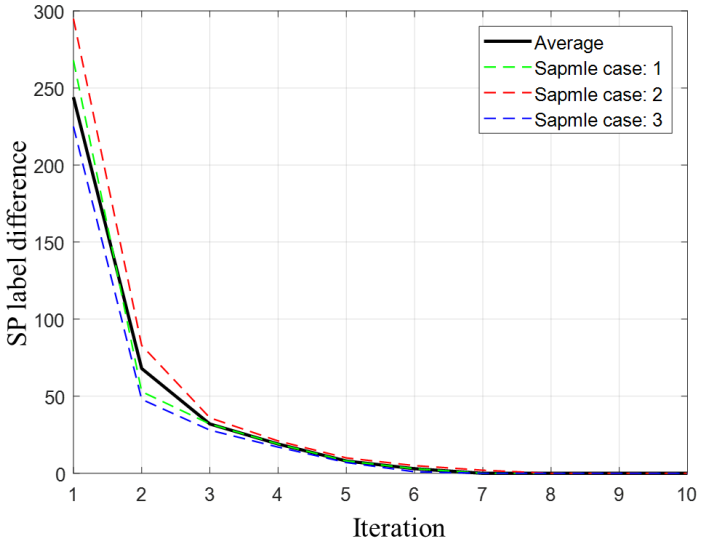



Figure 4-20 Superpixel label difference between iterations for three sample cases and the average of all the cases with $S = 6$ and $m = 0.2$.

**Texton Features**

For the orientation of the Gabor filters, six degrees in the range: [0º, 30º, 45º, 60º, 90º, 120º] are chosen to ensure covering the whole space of the region with a reasonable step. Adding more orientations seems to include more detail to the features, but on the other hand it will also increase the computation time and may add redundant information which may affect the classification accuracy.

The maximum and minimum values for the Gabor filter size were selected empirically by visually inspecting the filter response to the input image. For the small filter size, i.e. values under 0.3, filtered images are very close to the original image. Whilst for the large filter size i.e. the values above the 1.5, the images are intensively blurred. Therefore, the kernel sizes are selected within this range with the increment of 0.3, i.e. [0.3, 0.6, 0.9, 1.2, 1.5]. The wavelengths of sinusoid were selected empirically by visual inspection of the filter responses in the range of [0.8, 1.0, 1.2, 1.5].

As discussed in Section 4.2.5, the texton map is created by applying *k*-means clustering to different filter responses. A key question in using *k*-means clustering is to determine the number of clusters, which in the current method is equivalent to the number of textons, $k_{texton}$. However, it is not straightforward to provide an accurate number of the structures presented in the image. Theoretically, increasing number of clusters for texton generation provides more specific texton differences between clusters. Selecting a large $k_{texton}$ may result in over-classifying and increasing the computational cost. In the current experiments, the number of clusters (textons) $k_{texton}$ =5 is chosen empirically based on the number of tissues which may present in the FLAIR images, i.e. grey matter, white matter, tumour, oedema and other tissue types.

To evaluate the effectiveness of texton features in the classification task, two separate experiments were conducted. The classification pipeline was similar to SP_ERT method but with different sets of features. Firstly, all the SP features except textons are fed into ERT classifier. In the second experiment, only textons of the SPs are considered for classification. The accuracy of the leave-one-out experiments (STD and average) are presented in Figure 4-21. As can be seen, texton alone has better performance compared to the other features, while combining all the features provides the best classification accuracy.

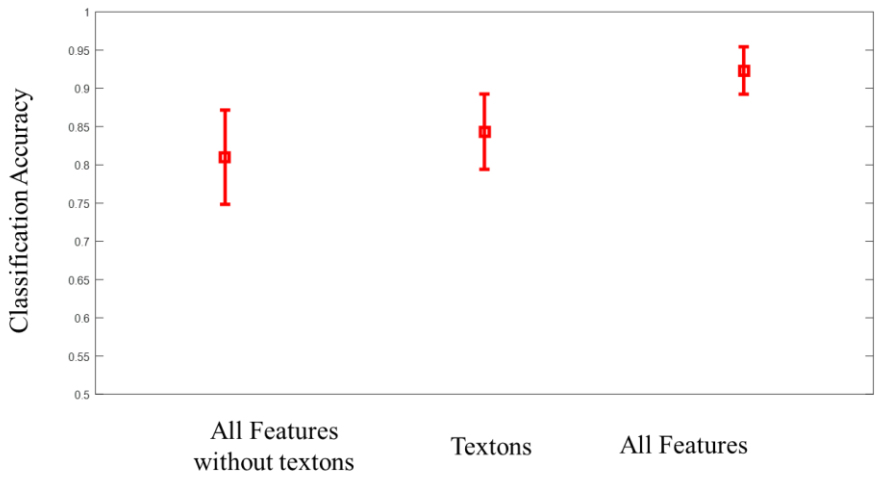



Figure 4-21 Comparison of classification accuracy of superpixels on the testing dataset by using different combination of the features (e.g. all features without textons, textons only and all features including textons).

**Fractal Features**

To evaluate the effect of threshold levels on fractal feature extraction, a different number of thresholds, $n_{threshold}$ = 1,…,8 have been examined. The measure to assess the effect of number of threshold level is the accuracy of superpixel classification using fractal features only. Figure 4-22 shows the average accuracy of the selected testing dataset using the different thresholds for Grades II, III, and IV. As can be seen, after increasing $n_{threshold}$ = 3 levels of threshold (which creates 6 binary channels) the overlap measure does not increase significantly. On the other hand, increasing each level of threshold will add 6 more features to the feature vector, since each binary channel has 3 fractal features. Augmenting the feature vector makes the classification more complicated and increases the computation time for both fractal feature calculation and classification. Therefore, the optimum level of threshold $n_{threshold}$ = 3 is selected for the segmentation of oedema and tumour core in the proposed method.

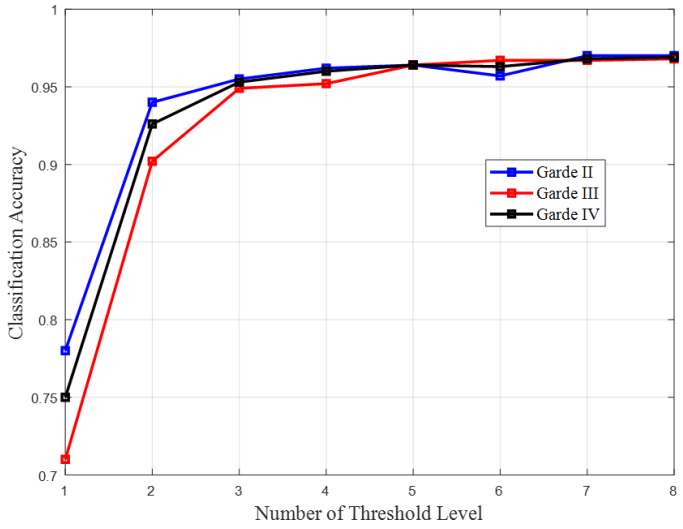



Figure 4-22 Effect of number of threshold levels on the classification accuracy for Grades II, III, and IV. Adding more threshold levels than $n_t$ = 3 does not affect the accuracy.

**Extremely Randomised Trees Parameters**

To assess the impact of ERT parameters on the classification performance, the experiment ran on the selected training images with different sizes of trees. The maximum depth of the trees, $D_{tree}$, for the ERT was set to 15. Minimum sample size, $n_{min}$, for splitting a node is selected to be 2, which is an optimal value for most classification tasks according to (Geurts *et al.*, 2006). The number of attributes for random split is considered as 5 which is equal to the number of selected features after applying the mRMR feature reduction. In general, the optimum value for $k_{attribute}$ for the classification tasks is $k_{attribute} = \sqrt{N_{feature}}$ where $N_{feature}$ is the total number of

features. However, the selected features in the current study are few, so the $k_{attribute} = 5$ is chosen as the optimal parameter. ERT classifier models with a different number of trees were assessed and the average classification accuracy on the testing dataset were calculated which are presented in Table 4-4. The average time of training the ERT with a different number of trees is also presented in Table 4-4. As can be seen, there is no significant improvement for the classifier accuracy after adding more than 20 trees to the ERT. On the other hand, increasing the number of trees will increase the computation time. As it can be seen in Table 4-4, the accuracy measures of classification for $N_{tree} = 20$ and $N_{tree} = 50$ are 98.22 and 98.28, respectively, whilst the training time for $N_{tree} = 50$ is more than two times higher than the ERT with $N_{tree} = 20$. Therefore, the size of trees $N_{tree} = 20$ is used for the ERT classifier in the current experiments.

Table 4-4 Impact of the number of trees on ERT classifier accuracy and training time.

| **Number of Trees** | 5 | 10 | 20 | 50 | 100 |
|---|---|---|---|---|---|
| **Classification Accuracy (%)** | 92.35 | 97.86 | 98.22 | 98.28 | 98.28 |
| **Training Time (mins)** | 134 | 253 | 498 | 1218 | 2437 |

**Feature Selection Methods**

Some popular feature selection methods were explained in Section 4.2.6, which are applied to the proposed SP_ERT method. *k*-fold cross validation with $k = 4$ is used for evaluation. Table 4-5 provides the average accuracy of the classification on the evaluation dataset in each fold. As can be seen, the mRMR method provides the best performance for the feature subset selection. The classification accuracy after selecting the features given by mRMR method was 98.11%, which was the highest amongst other feature reduction methods.

Table 4-5 Comparison of feature selection techniques. The accuracy of ERT classifier after feature selection is considered as evaluating the efficiency of the selected feature subsets.

| **Feature Selection Method** | **Accuracy (%)** |
|---|---|
| Fast Correlation-base Filter (FCBF) (Yu and Liu, 2003) | 95.23 |
| ReliefF (Kononenko, 1994) | 94.68 |
| Sparse logistic regression with Bayesian regularisation (SBMLR) (Cawley and Talbot, 2006) | 97.03 |
| Spectral feature selection (SPEC) (Zhao and Liu, 2007) | 96.54 |
| Minimum Redundancy Maximum Relevance (mRMR) (Peng *et al.*, 2005) | 98.11 |

### 4.3.4 Comparative Experimental Results

Leave-one-out cross validation (LOO-CV) was performed on single-channel MR FLAIR data for the classification stage. The brain MR images were partitioned into superpixels based on Equation (4-3) using the initial window side size of $S = 6$ pixels and the compactness factor $m = 0.2$. All the superpixels inside the brain area were used for classification. Superpixels are split in two classes based on the manual annotation: normal tissue and brain tumour including tumour core and oedema. Any superpixel with at least 50% of tumour pixels in manual annotation were considered as a tumour superpixel. The remaining superpixels were labelled as normal. The classifier was trained based on these two labels. For the testing stage, the trained classifier was then applied and labels were assigned to all the superpixels inside the brain. The tumour area was obtained by grouping the superpixels related to tumour class together.

Five features were used after mRMR feature selection. The final features were the normalised mean intensity, fractal dimension, two texton channels (cluster numbers 3 and 5) and mean curvature within the superpixel. The texton channels were selected in the mRMR process. However, the texton channel 3 mostly covers the white matter, while texton channel 5 represents the most overlap with the tumour. The reason for selecting 5 features was the comparison with SVM and the fact that the optimum attribute selection by ERT was $k_{attribute} = 5$. ERT can be used directly as feature selection and classification. However, in the current experiments, to ensure a fair comparison between the ERT and SVM classifiers the same feature set is considered.

Evaluation of the classifiers has been carried out quantitatively using three classification measures for the detection. From the standard four classification measures (accuracy, precision, sensitivity, specificity), accuracy and specificity will give very high values due to the imbalanced nature of the clinical data (approximately 10:1 ratio of healthy to tumour). Therefore, only precision and sensitivity were considered to properly evaluate the classification performance.

Table 4-6 presents the evaluation measures for classification of the superpixels using SVM and ERT. The results show that ERT produces a better classification performance compared to that of SVM, with an overall classification precision of 87.86%, sensitivity of 89.48% and BER of 6% for ERT, and of 83.59%, 87.82% and 7% for SVM, respectively.

Table 4-6 Comparison evaluation on superpixel classification in SP_SVM and SP_ERT, respectively, on the 5 features selected using mRMR. The classification is performed for tumour including oedema and active tumour core versus normal brain tissue. (BER is balanced error rate).

| Case No | Grade | SP_SVM | | | SP_ERT | | |
|---|---|---|---|---|---|---|---|
| | | Precision (%) | Sensitivity (%) | BER | Precision (%) | Sensitivity (%) | BER |
| 1 | II | 62.71 | 97.33 | 0.02 | 69.85 | 97.45 | 0.02 |
| 2 | II | 58.65 | 98.14 | 0.02 | 90.24 | 98.65 | 0.01 |
| 3 | II | 72.55 | 98.41 | 0.02 | 74.21 | 99.12 | 0.01 |
| 4 | II | 68.53 | 94.88 | 0.03 | 70.24 | 96.05 | 0.02 |
| 5 | II | 76.33 | 55.64 | 0.23 | 78.43 | 56.32 | 0.22 |
| 6 | II | 75.83 | 73.45 | 0.14 | 85.63 | 71.32 | 0.15 |
| 7 | III | 84.75 | 98.75 | 0.01 | 86.07 | 99.35 | 0.01 |
| 8 | III | 88.54 | 83.32 | 0.09 | 90.78 | 85.64 | 0.08 |
| 9 | III | 88.92 | 98.11 | 0.01 | 91.44 | 98.67 | 0.01 |
| 10 | IV | 95.22 | 83.25 | 0.09 | 97.44 | 89.03 | 0.06 |
| 11 | IV | 93.45 | 88.53 | 0.07 | 96.57 | 91.65 | 0.05 |
| 12 | IV | 81.55 | 73.98 | 0.14 | 84.33 | 75.92 | 0.13 |
| 13 | IV | 80.35 | 92.68 | 0.04 | 82.53 | 95.73 | 0.03 |
| 14 | IV | 90.12 | 92.51 | 0.04 | 91.32 | 95.87 | 0.03 |
| 15 | IV | 93.42 | 93.76 | 0.04 | 96.78 | 94.02 | 0.03 |
| 16 | IV | 87.45 | 83.06 | 0.09 | 90.21 | 84.15 | 0.08 |
| 17 | IV | 95.34 | 87.75 | 0.06 | 96.81 | 91.87 | 0.04 |
| 18 | IV | 98.33 | 82.56 | 0.09 | 98.43 | 85.33 | 0.08 |
| 19 | IV | 96.21 | 92.51 | 0.05 | 98.12 | 94.03 | 0.04 |
| **Mean** | **All** | **83.59** | **87.82** | **0.07** | **87.86** | **89.48** | **0.06** |
| **STD** | **All** | **11.76** | **11.09** | **0.06** | **9.27** | **11.23** | **0.06** |

The segmentation results were evaluated qualitatively by visual inspection and quantitatively using the Dice overlap measure. Figure 4-23 shows the Dice score overlap measure of the segmented tumour masks obtained by both SP_SVM and SP_ERT methods against the ground truth for each individual patient. The Dice overlap measure using the SP_ERT method is much better than that of SP_SVM for all the three tumour grades, with mean and standard deviation Dice score of $0.91 \pm 0.04$ for SP_ERT; and $0.87 \pm 0.05$ for SP_SVM. The segmentation results of SP_ERT method for Grade IV patients show a consistent increase for all cases compared to SP_SVM in Figure 4-23. Grade IV tumours appear with clearer boundaries in FLAIR images. It is difficult to compare the results for Grade III tumours since there are three cases with different comparison results. One reason for this discrepancy can be the low numbers of Grade III patient cases in the training phase. The difference between segmentation results for the Grade II patient cases varies for different cases. This is because the tumours in Garde II cases appear with more unclear boundaries due to infiltration, compared to the other grades. However, the SP_ERT method provides a better accuracy and robustness for the segmentation of Grade II tumours.

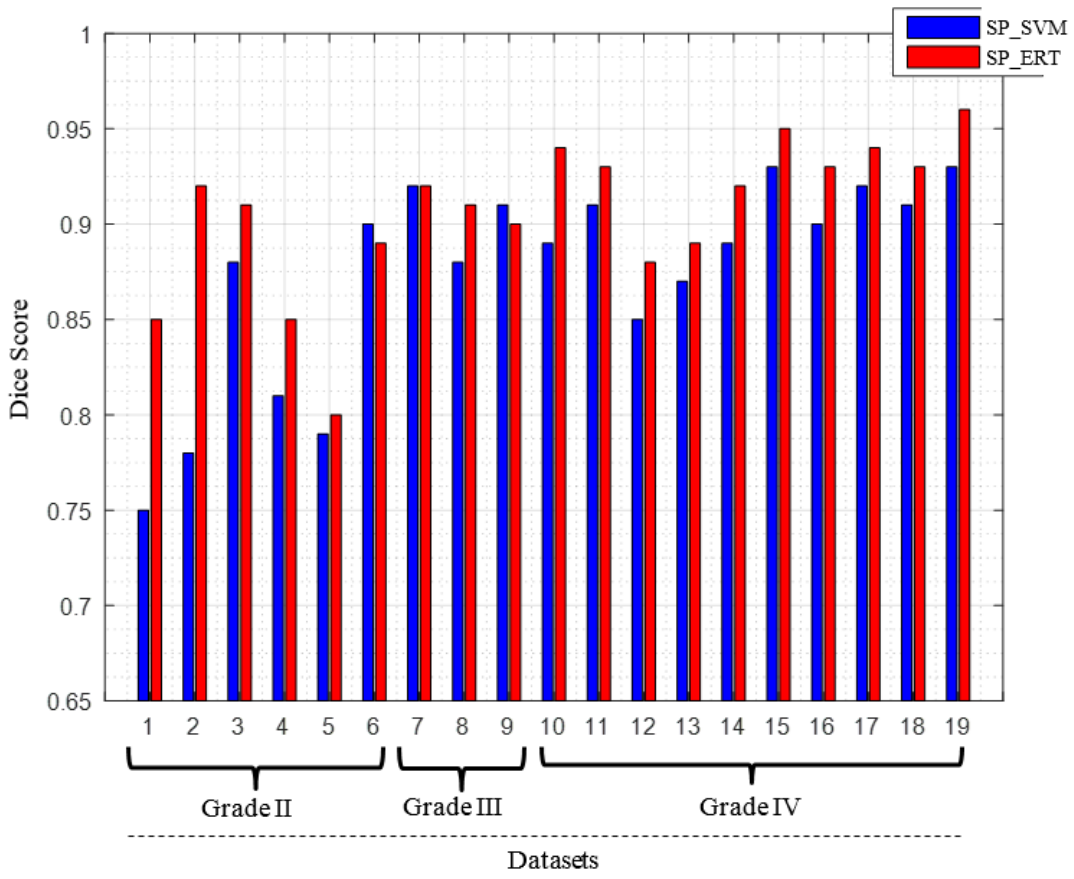



Figure 4-23 Comparison of Dice Score overlap measure of SP_SVM vs. SP_ERT for all the clinical patient data (19 scans). Dice score in the vertical axis starts from 0.65 for better illustration.

The Dice score overlap measure (mean and standard deviation) for SP_SVM vs. SP_ERT are compared in Figure 4-24 for different tumour grade types from II to IV. The results show that the method based on ERT classifier provides more accurate segmentations for all grades of tumour type. For segmentation results of different tumour grades using the SVM classifier, an evident difference can be seen between Dice overlap measures. The results are not good for grade II with mean overlap of 0.81, compared to the other two grades with mean overlap of 0.90. It should be noted that the minimum DSC in recent related publications is 0.85. Whilst the segmentation results based on ERT classifiers are consistent for all tumour grade types, with mean overlap of 0.91. The reason for the small difference between SP_ERT and SP_SVM results of Grade III was the insufficient number of patient cases.

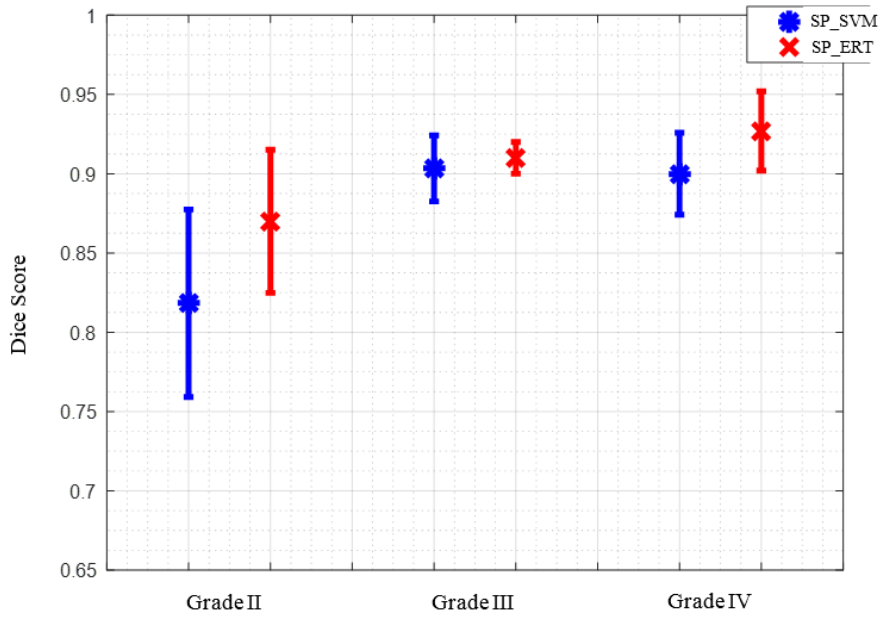



Figure 4-24 Comparison between average and standard deviation of Dice score overlap measure for SP_SVM vs. SP_ERT for different tumour grade types II to IV.

To determine the differences between both the segmentation measure of the Dice overlap and classification measures of precision and sensitivity, obtained using the two different classifiers (i.e. SVM and ERT), the Wilcoxon signed-rank test is used at a 99% confidence level, with 19 subjects. Table 4-7 shows the statistical parameters of the analysis which is based on the p and z values of the statistical test. The analysis demonstrates that using the ERT classifier instead of the SVM provides a statistically significant improvement in the segmentation measures of Dice overlap, and in the classification measures of precision and sensitivity.

Table 4-7 Statistical parameters of the Wilcoxon signed-rank test

| | **p** | **Z value** |
|---|---|---|
| **Dice** | < 0.001 | -3.826 |
| **Precision** | < 0.001 | -3.823 |
| **Sensitivity** | 0.001 | -3.340 |

Figure 4-25 shows examples of segmentation results for SP_ERT and SP_SVM methods overlaid on the FLAIR image. The manual annotation for the corresponding images are also overlaid to compare the segmentation boundaries. Both SP_SVM and SP_ERT methods

obtained accuracy above 0.80 for the segmentation of different tumour types, while SP_ERT method provided slightly better segmentations compared to the SP_SVM method. Figure 4-26 shows examples of much better detection and segmentation results obtained from SP_ERT methods, compared to that from SP_SVM. Most of the false positive superpixels from SP_SVM (e.g. Figure 4-25 (c4) and Figure 4-26 (c1)) were effectively eliminated using SP_ERT. Also, some tumour superpixels which were wrongly classified to the normal brain tissues by using SP_SVM (e.g. Figure 4-26 (c2) and (c3)) were classified correctly as tumour by using the SP_ERT, which demonstrates the higher sensitivity of the SP_ERT method. The SP_SVM method did not correctly segment the tumour core in Figure 4-26 (c2), while the SP_ERT method successfully segmented it. Comparison examples of segmentation for Grade II tumour in the first row of both Figure 4-25 and Figure 4-26 shows that the segmented tumour boundaries from SP_ERT (d1) were closer to the manual annotation, compared to that of SP_SVM (c1).

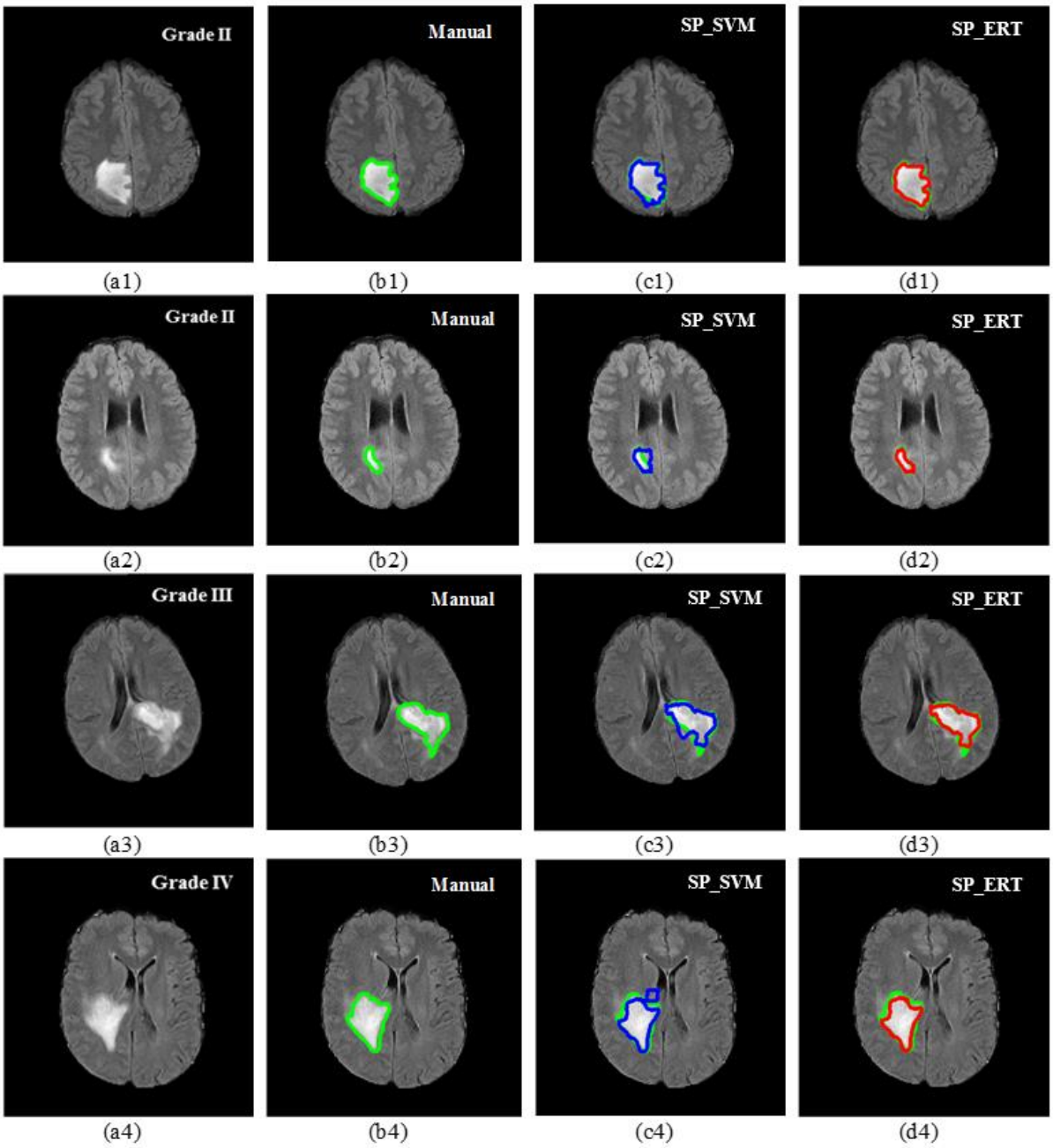



Figure 4-25 Examples of segmentation results overlaid on manual segmentation (green). FLAIR image with tumour Grade II (a1), Grade II (a2), Grade III (a3) and Grade IV (a4); (b1)-(b4) manual segmentation; (c1)-(c4) results using SP_SVM; (d1)-(d4) results using SP_ERT.

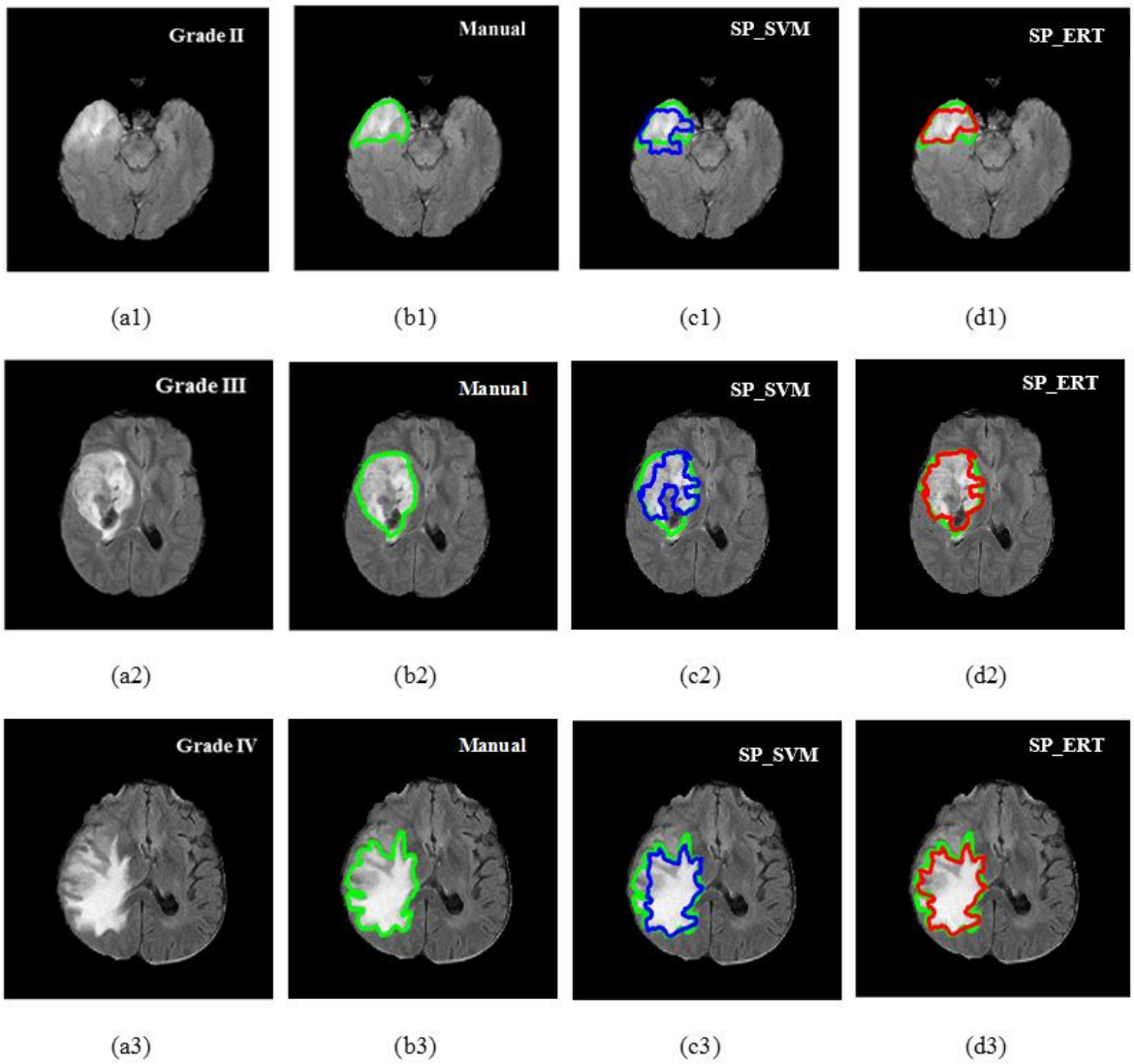



Figure 4-26 Examples of good segmentation results obtained from SP_ERT methods. FLAIR image with tumour Grade II (a1), Grade III (a2), Grade IV (a3); (b1)-(b3) manual segmentation; (c1)-(c3) results using SP_SVM; (d1)-(d3) results using SP_ERT. Most of the false positive superpixels from SP_SVM (e.g. (c1) and (c3)) can be effectively eliminated using SP_ERT; while some tumour superpixels which are wrongly classified to the normal brain tissues by using SP_SVM (e.g.(c2)) can be correctly classified as tumour by using the SP_ERT.

Figure 4-27 and Figure 4-28 show the 3D graphical view of the corresponding slices in Figure 4-25 and Figure 4-26, respectively. The 3D surfaces of the segmented tumours using the SP_SVM (blue) and SR_ERT (red) are overlaid on the ground truth surface (green). This provides a better representation of the whole segmentation in all slices. As it can be seen, the SP_SVM results in several false positive isolated small volumes. It should be noted that the post-processing stage was performed for both SP_SVM and SP_ERT methods. However, both

methods provide similar segmentation volumes for the tumour region. SP_ERT method has very few false positive isolated regions close to the tumour area. In Figure 4-28, the segmentation volume for Grade III has a large under-segmented surface. However, the corresponding slices in Figure 4-26 (middle row) show that the segmentation from the SP_ERT method (Figure 4-26 (d2)) has a closer surface to the ground truth, compared to the SP_SVM segmentation (Figure 4-26 (d2)).

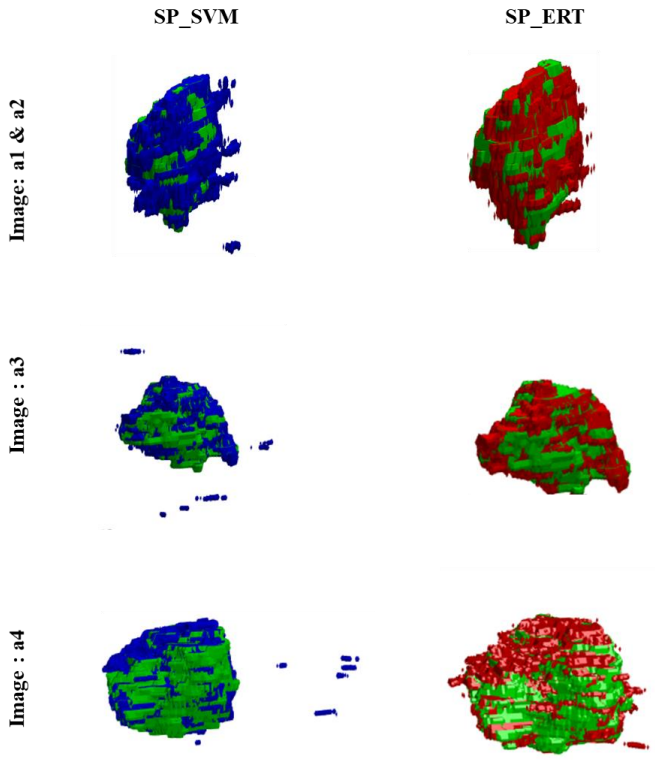



Figure 4-27 The 3D graphical representationof the segmented tumours in Figure 4-25 using SP_SVM (blue) and SP_ERT (red) overlaid on the ground truth (green).

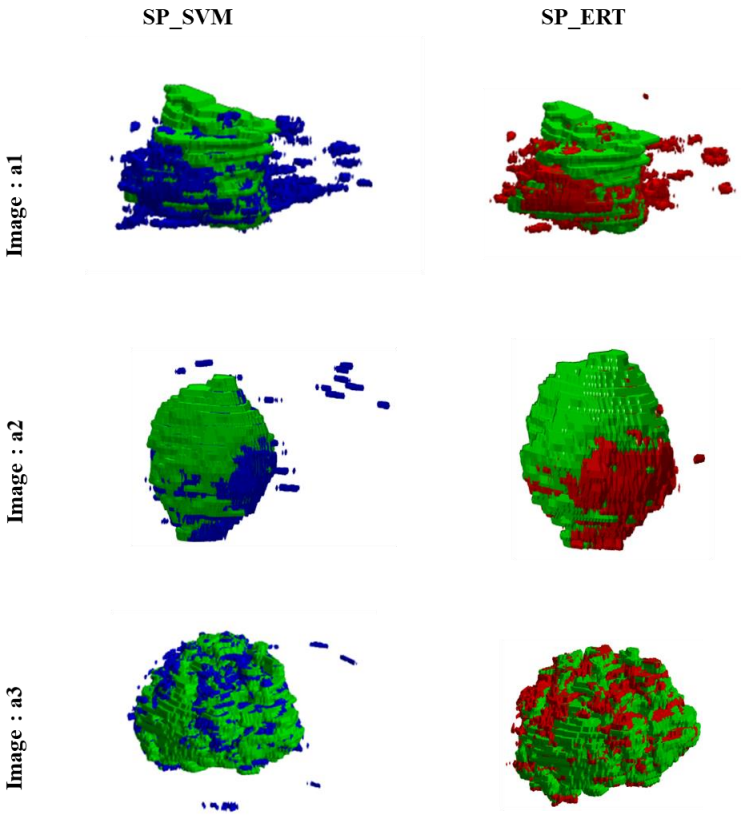



Figure 4-28 The 3D graphical representationof the segmented tumours in Figure 4-26 using SP_SVM (blue) and SP_ERT (red) overlaid on the ground truth (green).

### 4.3.5 BRATS 2013

Most of the parameters which are tuned for the clinical dataset are directly used in the experiments on BRATS dataset. These parameters include the number of clusters for texton generation, compactness coefficient for superpixel segmentation and the number of threshold levels for fractal features. All the parameters of superpixels and feature extraction for both SP_ERT and SP_SVM experiments are identical. However, only the superpixel size and filter size used for Gabor filter (Equation (4-19)) are slightly adjusted differently for the clinical and BRATS dataset regarding to their different image sizes and resolutions. The image dimension in *X*-*Y* plane for the clinical dataset is 256 × 256, while for the BRATS 2013 dataset has an average of 170 × 220. The superpixel side size 5 is optimal for segmentation of BRATS 2013 dataset. However, for the clinical dataset, using superpixel side size 5 produces more superpixels, which increases the computation time. Therefore, superpixel side size 6 was selected. The superpixel size of 5 is used in the BRATS dataset while it was 6 for the clinical dataset. A slightly smaller range of Gabor filter size (e.g. [0.3 0.5 0.8 1.1 1.4]) is used for texton feature extraction for the BRATS dataset. All the five features selected using mRMR

in the clinical dataset experiment are also used in BRATS dataset for the classification of each superpixel.

Table 4-8 presents the evaluation measures for SP_SVM and SP_ERT classification of superpixels into tumour and non-tumor. ERT produces better classification performance, compared to that of SVM, with an overall classification precision of 89.09%, sensitivity of 88.09% and BER of 6% for ERT, and of 83.79%, 82.72% and 9% for SVM, respectively.

Table 4-8 Comparison evaluation on superpixel classification using SP_SVM and SP_ERT classifier, respectively, on the BRATS 2013 dataset using 5 features selected by mRMR. The classification is performed for tumour including oedema and active tumour core versus normal brain tissue (BER is balanced error rate).

| **Case No** | **Grade/ID** | **SP_SVM** | | | **SP_ERT** | | |
|---|---|---|---|---|---|---|---|
| | | **Precision (%)** | **Sensitivity (%)** | **BER** | **Precision (%)** | **Sensitivity (%)** | **BER** |
| 1 | LG-01 | 87.68 | 89.43 | 0.06 | 91.84 | 88.18 | 0.06 |
| 2 | LG-02 | 96.98 | 88.60 | 0.06 | 99.02 | 92.63 | 0.04 |
| 3 | LG-04 | 75.59 | 81.95 | 0.10 | 78.40 | 90.67 | 0.05 |
| 4 | LG-06 | 84.57 | 87.42 | 0.07 | 92.15 | 90.05 | 0.05 |
| 5 | LG-08 | 90.95 | 83.54 | 0.09 | 93.11 | 91.05 | 0.05 |
| 6 | LG-11 | 89.91 | 82.67 | 0.09 | 91.41 | 86.78 | 0.07 |
| 7 | LG-12 | 91.42 | 83.19 | 0.09 | 92.18 | 84.19 | 0.08 |
| 8 | LG-13 | 74.48 | 79.19 | 0.11 | 79.28 | 85.86 | 0.08 |
| 9 | LG-14 | 83.17 | 80.37 | 0.10 | 88.03 | 82.58 | 0.09 |
| 10 | LG-15 | 76.15 | 80.60 | 0.10 | 82.64 | 89.29 | 0.06 |
| 11 | HG-01 | 92.77 | 92.55 | 0.04 | 98.47 | 95.91 | 0.03 |
| 12 | HG-02 | 83.51 | 82.15 | 0.09 | 90.45 | 88.62 | 0.06 |
| 13 | HG-03 | 85.46 | 79.59 | 0.11 | 91.31 | 88.68 | 0.06 |
| 14 | HG-04 | 94.08 | 89.30 | 0.06 | 98.69 | 90.96 | 0.05 |
| 15 | HG-05 | 78.96 | 72.06 | 0.14 | 83.16 | 77.70 | 0.12 |
| 16 | HG-06 | 81.54 | 74.77 | 0.13 | 93.13 | 90.32 | 0.05 |
| 17 | HG-07 | 75.48 | 79.60 | 0.11 | 83.16 | 87.81 | 0.07 |
| 18 | HG-08 | 87.87 | 90.58 | 0.05 | 89.21 | 93.88 | 0.04 |
| 19 | HG-09 | 84.78 | 87.04 | 0.07 | 87.56 | 90.35 | 0.05 |
| 20 | HG-10 | 67.77 | 65.63 | 0.18 | 73.17 | 71.84 | 0.15 |
| 21 | HG-11 | 90.53 | 85.68 | 0.08 | 92.39 | 90.21 | 0.05 |
| 22 | HG-12 | 88.58 | 86.82 | 0.07 | 92.08 | 89.36 | 0.06 |
| 23 | HG-13 | 80.10 | 84.35 | 0.08 | 88.64 | 89.23 | 0.06 |
| 24 | HG-14 | 84.74 | 87.99 | 0.07 | 88.80 | 91.76 | 0.05 |
| 25 | HG-22 | 78.21 | 80.75 | 0.10 | 88.79 | 92.83 | 0.04 |
| 26 | HG-24 | 82.50 | 85.14 | 0.08 | 88.87 | 87.98 | 0.07 |
| 27 | HG-25 | 82.23 | 86.08 | 0.07 | 90.95 | 88.16 | 0.06 |
| 28 | HG-26 | 84.41 | 82.60 | 0.09 | 91.71 | 89.84 | 0.06 |
| 29 | HG-27 | 77.16 | 72.67 | 0.14 | 80.93 | 75.54 | 0.13 |
| 30 | HG-22 | 82.10 | 79.19 | 0.11 | 93.09 | 90.42 | 0.05 |
| **Mean** | **All** | **83.79** | **82.72** | **0.09** | **89.09** | **88.09** | **0.06** |
| **STD** | **All** | **6.63** | **5.95** | **0.03** | **6.00** | **5.22** | **0.03** |

The Dice overlap ratio between the ground-truth from manual annotation and the segmented tumour using SP_ERT and SP_SVM experiments for the BRATS dataset is presented in Table 4-9. The overlap ratio using the SP_ERT method is much better than that of SP_SVM for all the three tumour grades, with mean Dice score of 0.88 for SP_ERT; and 0.83 for SP_SVM.

Table 4-9 Comparison results for Dice overlap ratio between manual annotation and the automated segmentation using SP_SVM and SP_ERT for BRATS 2013 dataset (30 scans).

| **Case No** | **Grade/ID** | **Dice** | |
|---|---|---|---|
| | | **SP_SVM** | **SP_ERT** |
| 1 | LG-01 | 0.85 | 0.89 |
| 2 | LG-02 | 0.93 | 0.95 |
| 3 | LG-04 | 0.78 | 0.87 |
| 4 | LG-06 | 0.84 | 0.91 |
| 5 | LG-08 | 0.88 | 0.92 |
| 6 | LG-11 | 0.86 | 0.89 |
| 7 | LG-12 | 0.88 | 0.92 |
| 8 | LG-13 | 0.75 | 0.81 |
| 9 | LG-14 | 0.80 | 0.84 |
| 10 | LG-15 | 0.78 | 0.88 |
| 11 | HG-01 | 0.89 | 0.92 |
| 12 | HG-02 | 0.83 | 0.88 |
| 13 | HG-03 | 0.82 | 0.91 |
| 14 | HG-04 | 0.90 | 0.92 |
| 15 | HG-05 | 0.74 | 0.78 |
| 16 | HG-06 | 0.79 | 0.91 |
| 17 | HG-07 | 0.78 | 0.85 |
| 18 | HG-08 | 0.89 | 0.91 |
| 19 | HG-09 | 0.86 | 0.89 |
| 20 | HG-10 | 0.65 | 0.71 |
| 21 | HG-11 | 0.87 | 0.92 |
| 22 | HG-12 | 0.88 | 0.91 |
| 23 | HG-13 | 0.81 | 0.89 |
| 24 | HG-14 | 0.86 | 0.90 |
| 25 | HG-15 | 0.78 | 0.91 |
| 26 | HG-22 | 0.84 | 0.88 |
| 27 | HG-24 | 0.85 | 0.89 |
| 28 | HG-25 | 0.84 | 0.90 |
| 29 | HG-26 | 0.75 | 0.79 |
| 30 | HG-27 | 0.81 | 0.91 |
| **Mean** | **All** | **0.83** | **0.88** |
| **STD** | **All** | **0.06** | **0.05** |

Figure 4-29 and Figure 4-31 show examples of segmentation results for SP_ERT and SP_SVM methods compared to the manual annotations. Figure 4-29 shows the segmentation results for high-grade tumour and Figure 4-31 for low-grade tumour. Both SP_SVM and SP_ERT methods obtained accuracy above 0.80 for the segmentation of different tumour types. However, SP_ERT method provided slightly better segmentations than SP_SVM method.

Most of the false positive superpixels from SP_SVM (e.g. Figure 4-29 (c2) and Figure 4-31 (c3)) can be effectively eliminated using SP_ERT. Furthermore, some tumour superpixels which are wrongly classified as normal brain tissue by using SP_SVM (e.g. Figure 4-31 (c2)) can be correctly classified as tumour by using the SP_ERT, which demonstrates the higher sensitivity of the SP_ERT. Comparison examples of segmentation for both HG and LG tumours in Figure 4-29 and Figure 4-31 illustrate that the segmented tumour boundary from SP_ERT is closer to the manual annotation, compared to that of SP_SVM.

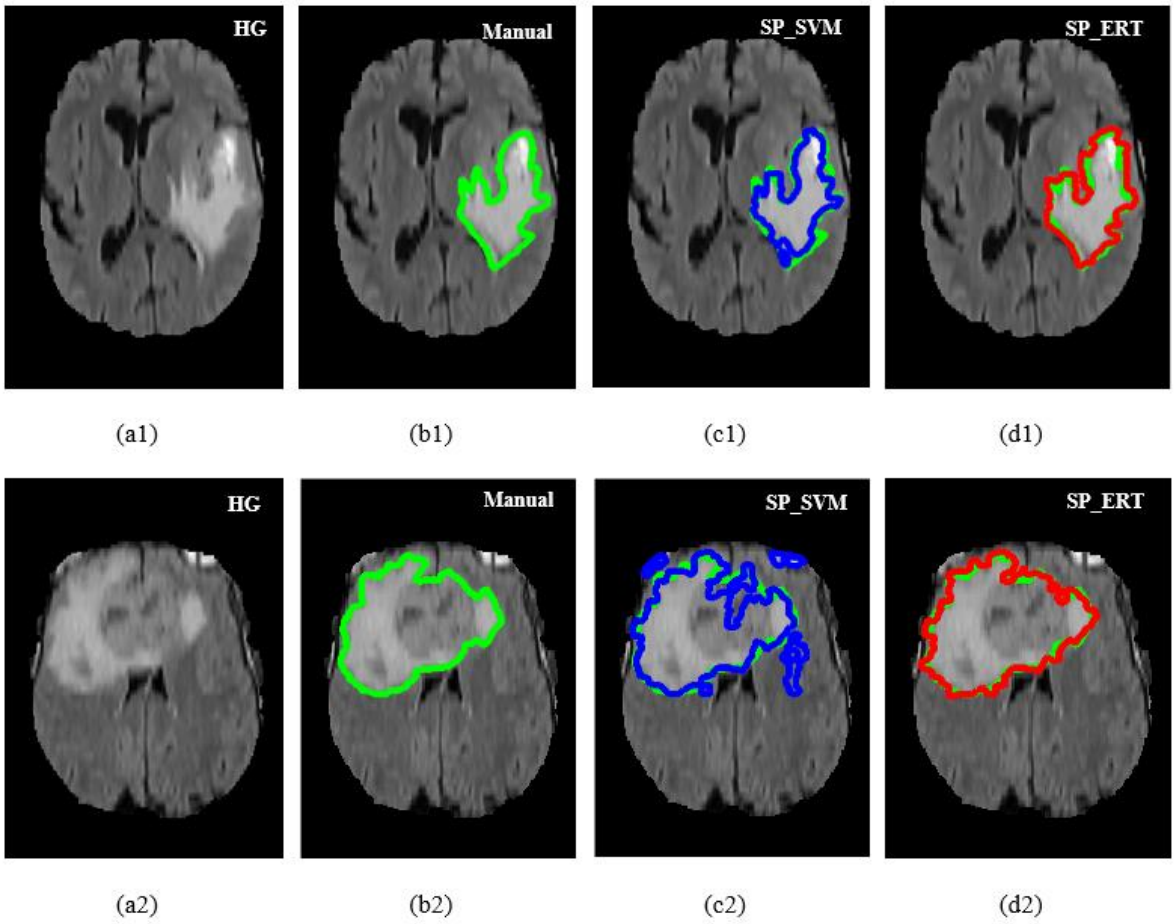



Figure 4-29 Examples of segmentation results obtained from SP_ERT methods on BRATS 2013 data. FLAIR image with high grade tumour Case HG-01 (a1), HG-15 (a2); (b1)-(b2) manual segmentation; (c1)-(c2) results using SP_SVM; (d1)-(d2) results using SP_ERT.

Figure 4-30 shows the 3D graphical view of the corresponding slices in Figure 4-29. The 3D surfaces of the segmented tumours using the SP_SVM (blue) and SR_ERT (red) are overlaid on the ground truth surface (green). As it can be seen, similar to the segmentation of clinical datasets, the SP_SVM method produces several false positive isolated volumes. The SP_ERT

method has very few false positive small volumes. However, both methods provide similar segmentation volumes for the tumour region.

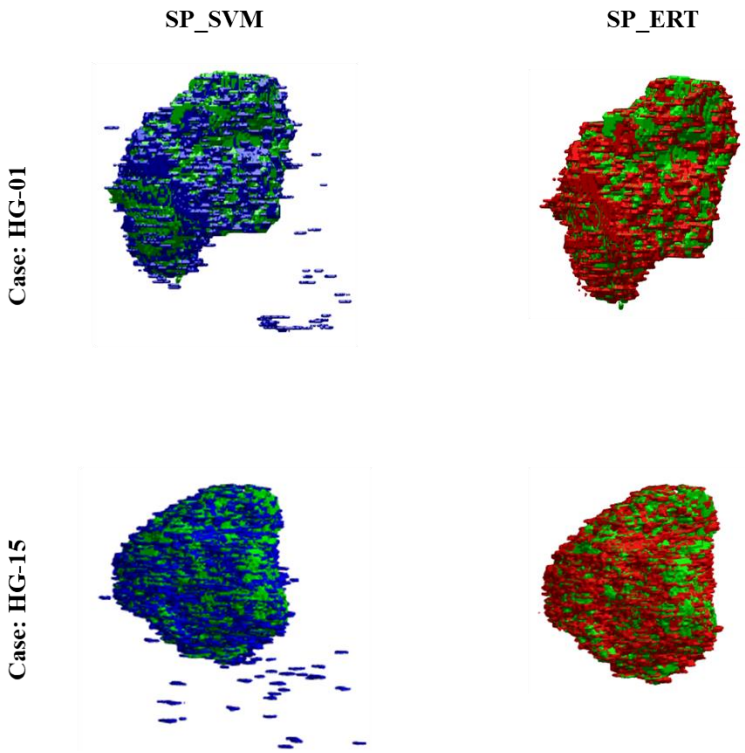



Figure 4-30 3D graphical representationof the segmented tumours in Figure 4-29 (HG cases) using SP_SVM (blue) and SP_ERT (red) overlaid on the ground truth (green).

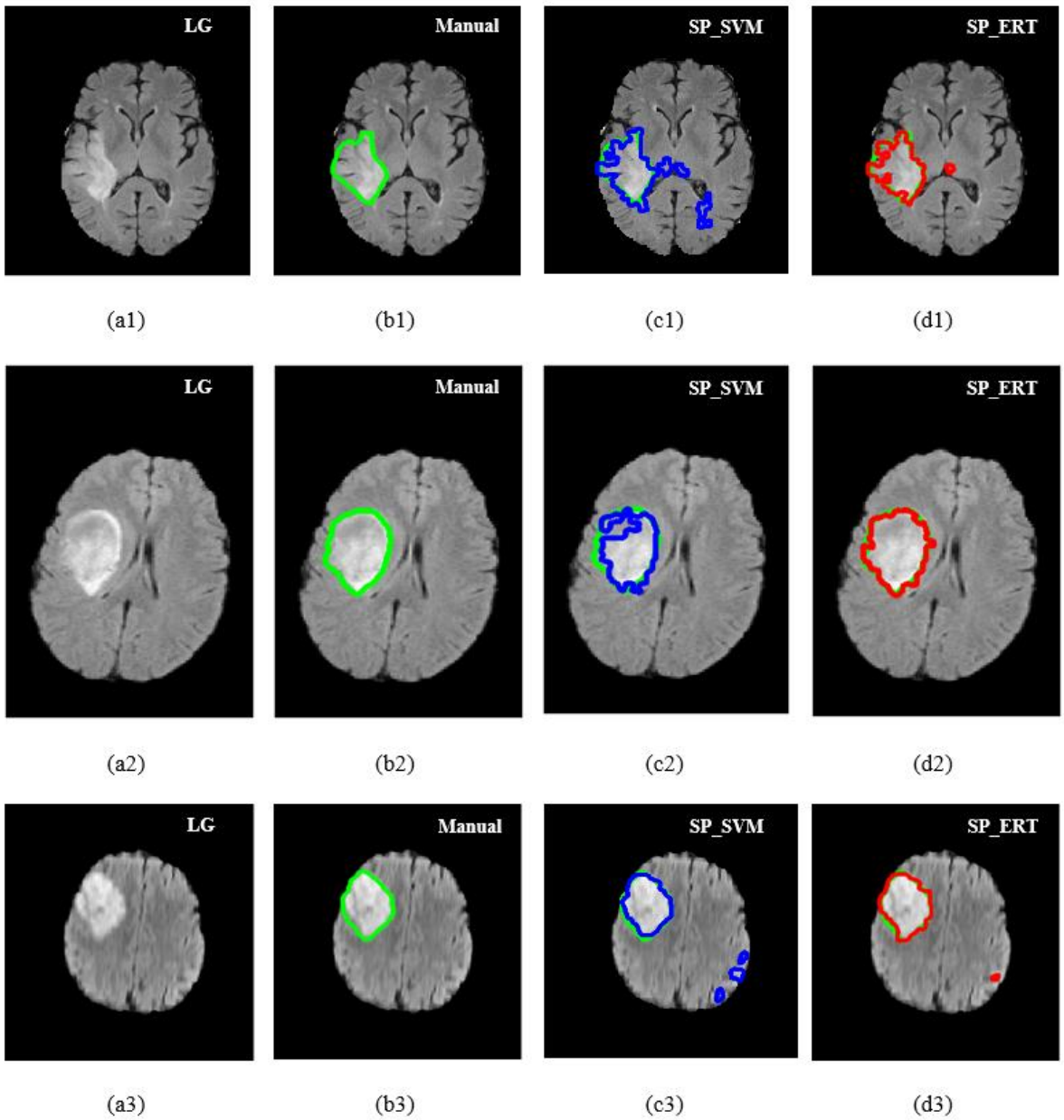



Figure 4-31 Examples of segmentation results obtained from SP_ERT methods on BRATS 2013 data. FLAIR image with low grade tumour Case LG-04 (a1), LG-11 (a2) and LG-12 (a3); (b1)-(b3) manual segmentation; (c1)-(c3) results using SP_SVM; (d1)-(d3) results using SP_ERT.

Figure 4-32 shows the 3D graphical view of the corresponding slices in Figure 4-31. The 3D surfaces of the segmented tumours using the SP_SVM (blue) and SR_ERT (red) are overlaid on the ground truth surface (green). As it can be seen, similar to the segmentation of LG tumours, the SP_SVM method produces several false positive isolated volumes. The SP_ERT method has very few false positive small volumes.

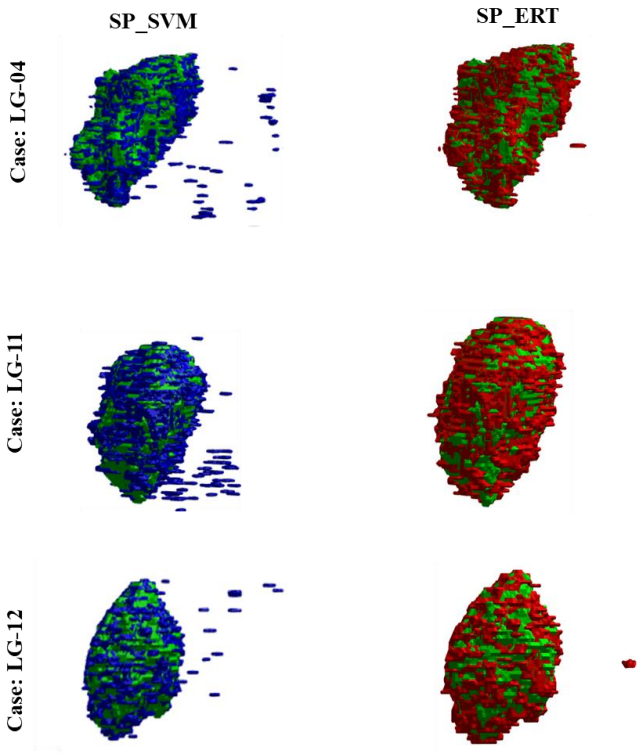



Figure 4-32 3D graphical representationof the segmented tumours in Figure 4-31 (LG cases) using SP_SVM (blue) and SP_ERT (red) overlaid on the ground truth (green).

## 4.4 Discussion

### 4.4.1 SP_ERT Single Modality Method

FLAIR images are commonly acquired in clinical practice as part of standard diagnostic clinical MRI of brain tumours. The experimental results in this chapter and Section 4.3.4, which are shown in Table 4-6, Table 4-7, Figure 4-25 and Figure 4-26, demonstrate high performance of automated detection and segmentation of the brain tumour oedema and core regions in FLAIR MRI. The method was also further validated on BRATS 2013 training dataset (FLAIR) with the similar model parameters and features tuned for the clinical dataset. The experimental results in Section 4.3.5, which are shown in Table 4-8 and Table 4-9, suggests the robustness of the SP_ERT single modality method.

Selecting an appropriate superpixel size is essential for increasing the overall segmentation accuracy within an optimum calculation speed. Selecting large superpixel size requires fewer number of total superpixels, hence it can ensure fast computation and meanwhile may provide sufficient information for feature extraction such as stable texture features. On the other hand, a large superpixel size may contain more than one class of pixel which may cause inaccurate feature calculation (such as small areas of calcification or haemorrhage), and it is also not

appropriate for small sized lesions. Whereas, a small superpixel size has higher probability to purely contain one class of pixel, hence it is preferred for segmentation of a small lesion. However, they may not be enough pixels for calculating stable features. Also, the computation time for generating the small size partitions is very high due to the large number of total superpixels. The aim of optimisation is to find a superpixel size which provides a good trade-off between computation time and segmentation accuracy. In Section 4.3.3, the size of superpixel is obtained through exhaustive parametric searching during the training stage on the selected data.

The compactness factor is another important parameter for superpixel segmentation which should be tuned. As explained in Section 4.3.3, higher value of compactness factor provides more rigid partitions which are more stable and usually less noisy. In the case of superpixel segmentation noise is considered as holes or sparse separated pixels. However, the rigid superpixels may not follow the tissue boundaries very well, especially in the cases where there are no sharp or clear boundaries. On the other hand, lower values for the compactness factor provides more flexible and accurate boundaries, but produces more isolated and disconnected pixels. They also may generate very narrow superpixels which are not appropriate for texture analysis. In the Section 4.3.3, the compactness factor is determined using visual inspection of matching the superpixels with the boundaries of the ground truth.

The application of SP_ERT method on the BRATS data is compared to the methods published in (Menze *et al.*, 2015) which used the same data in MICCAI challenge. However, some of the corresponding methods are assessed on the training dataset, whilst others are on the separate testing dataset. Since the current study is based on binary classification (i.e. tumour including oedema and active tumour core versus normal brain tissue) using a single FLAIR protocol, it is difficult to have a direct comparison with the current published methods on BRATS data. However, the experimental results of the SP_ERT method are in the same range of other methods and are close to the best segmentation of the complete tumour which demonstrates the strength of the method. The image patches from the superpixels are following the edges in the images, therefore including more homogeneous pixels. This increases the robustness of the final segmentation after classification of superpixels. However, in the case of small tissues that encompass few superpixels or those smaller than an average superpixel, the algorithm might fail. Misclassification of these superpixels will result in assigning the small tumorous region as healthy brain tissues.

This study also emphasises the importance of MRI histogram normalisation in the preprocessing stage. This is of importance especially when the method is applied to the data that are from multi-centres and different scanners such as BRATS dataset. When the histogram

normalisation was applied on the clinical data, there were only slight differences for the Dice scores, for the individual patient data, whereas the mean Dice score for all the 19 data is the same as before (i.e. 0.91). This is mainly because the clinical data are quite consistent. Also, the feature normalisation step was found out to be quite important, especially when it was applied to the BRATS data, the segmentation results improved significantly using both SP_SVM and SP_ERT. This may be the reason why even without histogram normalisation, the mean Dice score of 0.82 for SP_ERT was still obtained (partial normalisation) as shown in Table 4-2.

### 4.4.2 Applying the SP_ ERT on BRATS dataset

The BRATS training dataset was used to evaluate the robustness of the method. As discussed in the Section 4.3.5, most of the parameters are the same as those optimised for the clinical data. Figure 4-33 shows the overall average and standard deviation of Dice score overlap measures for all 19 clinical patient data and 30 BRATS 2013 dataset using both SP_ERT and SP_SVM methods. The results show that using the state-of-the art ERT for classification of superpixels provides more accurate and robust segmentation compared to that of an SVM classifier. For the clinical dataset, the Dice score overlap measure for SP_ERT segmentation is 0.91 ± 0.04, while for SP_SVM method it is 0.87 ± 0.05. For BRATS 2013 dataset, the Dice score overlap measure for SP_ERT segmentation is 0.88 ± 0.05, while for SP_SVM method it is 0.83 ± 0.06. The mean Dice scores obtained from BRATS training dataset is closer to that from the clinical dataset, this suggests robustness of the method.

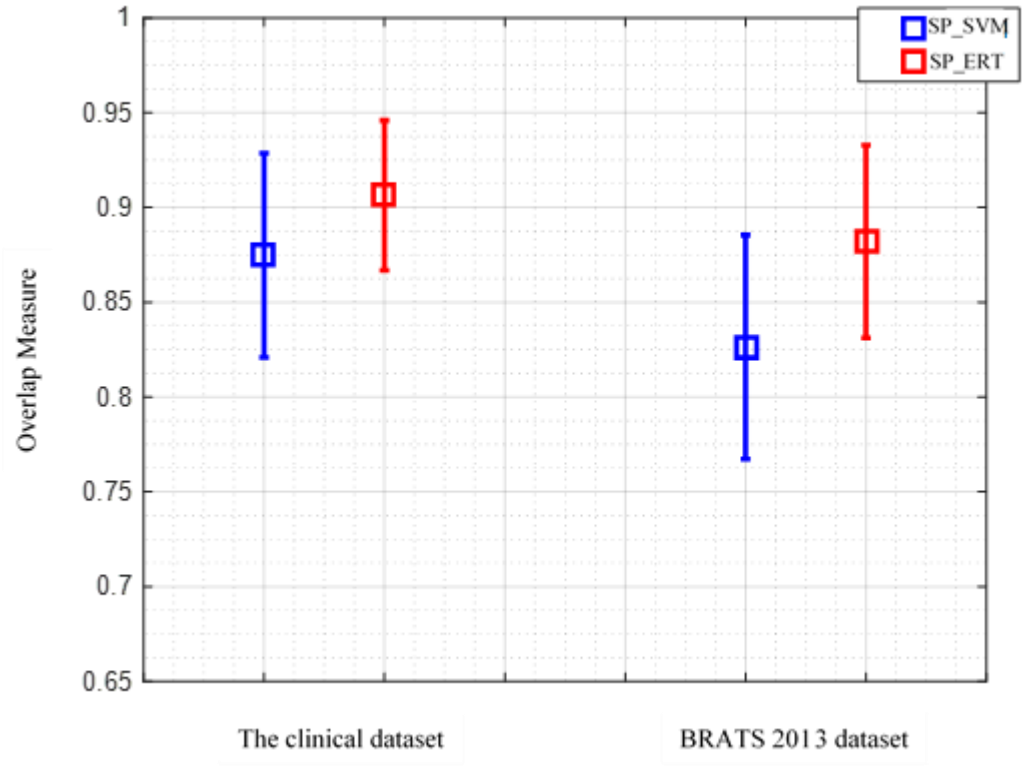



Figure 4-33 Comparison of the average and standard deviation of Dice score overlap measures for SP_SVM vs. SP_ERT for all 19 data scans in the clinical dataset and 30 clinical scans in BRATS 2013 dataset.

Table 4-10 presents the comparison of applying the SP_ERT single modality method on BRATS 2013 dataset with the best scores in the MICCAI challenges (Menze *et al.*, 2015). The method proposed by Tustison *et al.* (Tustison *et al.*, 2013) was the winner of on-site BRATS 2013 challenge and performed on the challenge data. The best on-site score could provide a comparable reference using BRATS dataset despite the difference between datasets. The SP_ERT single modality method was also compared to the method proposed by Reza *et al.* (Reza and Iftekharuddin, 2013) which has the best result for the training set of the BRATS multi-protocol dataset. This is the same dataset which was used in the experiments of the SP_ERT method, however only the FLAIR protocol was used. This work has achieved the average Dice overlap of 0.88 which is closer to that of 0.92 by Reza's method. As explained in Section 4.3.5 to emphasise the robustness of the SP_ERT method, the similar optimum parameters and the same five features selected for the clinical dataset are directly applied to the BRATS dataset. The algorithm is trained particularly on 1.5T clinical data from a single centre, whereas the BRATS data contains multicentre data from 1.5T and 3T MRI scanners. This may be the reason for the slightly difference of the results between the two datasets.

Table 4-10 Comparison with other related methods using BRATS dataset (MICCAI 2013). Note: the proposed SP_ERT method and Reza *et al.* (Reza and Iftekharuddin, 2013) are performed on BRATS clinical training data and the other work (Tustison *et al.* (Tustison *et al.*, 2013)) is performed on BRATS challenge data.

| Method | Description | Comment | Complete tumour (Dice) |
|---|---|---|---|
| **Tustison** (Tustison *et al.*, 2013) | Random forests (ANTs/ANTsR package) | Best MICCAI 2013 on-site | 0.87 |
| **Reza** (Reza and Iftekharuddin, 2013) | Random forests + texture features | Best on training MICCAI 2013 | 0.92 |
| **Proposed SP_ERT** | ERT + superpixels | Training MICCAI 2013 | 0.88 |

## 4.5 Limitations

Although the current segmentation algorithm method was evaluated on FLAIR images, it should be straightforward to apply the same superpixel methodology to other protocols such as contrast enhanced T1-weighted images to determine the signal intensity and higher order features that best segment the contrast enhancing region of high grade gliomas. In the results Section 4.3.4, it was noted that in Figure 4-26 (a2), the small hypointense spots in the FLAIR (and corresponding T1-weighted) maybe calcifications, and the hypointense FLAIR region, which is excluded by the SP_SVM method (Figure 4-26 (c2)) but included in the SP_ERT analysis (Figure 4-26 (d2)) is haemorrhagic since there is hyperintensity in the T1-weighted MRI. This is a limitation of the current single modality analysis if these regions need to be separately specified. This will be further investigated in the next chapter by developing and extending the superpixel-classification based method to multimodal data. This will also include the segmentation of different tumour tissue structures (e.g. necrosis, active tumour, non-enhancing tumour, and oedema) by considering information from multimodal clinical MRI, including perfusion and diffusion imaging.

The ground truth for the clinical dataset, were provided based on one expert's manual annotation. Some errors may occur in the manual annotations, which may include intratumoural bleeding or calcification in the tumour (e.g. in Figure 4-26 (b2)). Using those disputable annotations to train the model may result in some errors in the final segmentation. As it can be seen in Figure 4-31 (b1), a small part of the normal brain (hypo-intensity), was included in manual annotation. Both SP_SVM and SP_ERT excluded these dark regions, which increase the segmentation overlap error.

The methods proposed in this chapter were evaluated on three patient cases with Grade III tumour. This is not sufficient to conduct a good comparison with other Grades (i.e. II and IV). Although the provided number of Grade III in the clinical dataset was three, the reason for not

excluding them from the experiment was to assess the generality of applying the proposed methods on different types of brain tumours. However, it is difficult to conclude with certainty that the SP_ERT method is more accurate for Grade III tumour compared to SP_SVM.

A major limitation of superpixel based segmentation methods is the labelling procedure for heterogeneous superpixels. They may contain pixels with different labels that make it difficult to allocate a certain label for this kind of superpixels. This is because the superpixel segmentation is based on a single resolution (initial size) which is a limitation for more heterogeneous and complex structures.

The parameter selection procedure to find an optimal value for compactness factor, *m*, was conducted by visually inspecting some selected slices from different cases. To find a more accurate value, quantitative measures can be used, such as the spatial distance between the ground truth edges and the superpixel boundaries.

## 4.6 Conclusion

In this chapter a fully automated method for the detection and segmentation of brain tumour from FLAIR MRI images was proposed. This work has been published in the International Journal of Computer Assisted Radiology and Surgery (IJCARS) (Soltaninejad *et al.*, 2016). The method is based on the formation of superpixels by grouping voxels with similar properties and extraction of the features from superpixels. The feature space includes nonparametric features i.e. curvature and statistical intensity features, and hand designed features i.e. Gabor texton feature and fractals. ERT is then used for classification of superpixels into tumour or healthy brain tissue. The experimental results demonstrate the high detection and segmentation performance of the SP_ERT, with average sensitivity of 89.48%, BER of 6%, and Dice overlap measure of 0.91 for the complete tumour. The method was further evaluated on BRATS 2013 dataset to assess the robustness of the method. The experiments provided similar good performances of 88.09%, 6% and 0.88 for sensitivity, BER, and Dice overlap measure, respectively. This provides a close match to expert delineation across all grades of glioma, leading to a faster and more reproducible method of brain tumour delineation to aid patient management.

Although the current SP_ERT method provides promising results for the segmentation of the complete tumour using only single modality i.e. FLAIR, the information from a single modality is not sufficient for segmentation of all the tumour tissue subtypes. Whilst adding more protocols may provide further separation between different tumour parts (e.g. oedema, tumour core, necrosis, and enhancing core) within the complete tumour region. To incorporate

more MR protocols into the current framework for multi-class analysis of the tumour, some of the algorithms, e.g. superpixel segmentation, feature extraction, and classifiers should be extended to multidimensional input data and cope with multiple classes. Future automated methods are likely to incorporate information from multimodal clinical MRI as in the BRATS database studies, and that also include perfusion and diffusion imaging to detect tumour tissue subtypes (e.g. necrosis, active tumour, infiltrative tumour, oedema) (Sauwen *et al.*, 2015). The use of multimodal MRI for the segmentation of brain tumour tissue subtypes will be further investigated in the next chapter (Chapter 5).

# Chapter 5

# Multimodal MRI Supervoxel-based Brain Tumour Tissue Classification

## 5.1 Introduction

In the previous chapter an automated method was proposed that utilised a single protocol commonly used in clinical applications, i.e. FLAIR, to segment the tumour in brain MRI images. The results presented a promising performance compared to the state-of-the-art methods for segmentation of the complete tumour from the normal brain tissue. However, a single protocol does not include all information for separation between different tumour parts (oedema, tumour core). As explained in Chapter 2, to determine the tumour grade type and severity of the disease and other clinical planning, the segmentation and measurement of each tumour part is essential.

Most of the existing brain tumour segmentation studies were performed on conventional MRI protocols (i.e. FLAIR, T1-weighted (with contrast) and T2-weighted), which are based on qualitative image intensities. In this chapter, the isotropic ($p$) and anisotropic ($q$) diffusion components derived from DTI (Peña *et al.*, 2006) will also be considered, in addition to the conventional MRI sequences. As explained in Chapter 2, DTI protocols provide parameters that are related to the average microscopic movement of water within tissue structure ($p$) and whether this movement has an anisotropic element of diffusion ($q$). The hypothesis is that combining DTI and C-MRI may provide quantitative features that increase the classification accuracy and improve tumour segmentation results.

Most previous supervised learning methods are voxel-wise, in which a window or subarea around a voxel is normally used to extract features for classifying each individual voxel. The multimodal MRI data is comprised of millions of voxels (i.e. the sum of all voxels across each image modality), therefore voxel based methods usually require significant computational time. However, few studies have used superpixels or supervoxels in conjunction with other methods such as CRF (Wei Wu, 2013), or MRF (Zhao *et al.*, 2013) to segment the tumour. In this chapter, a supervoxel based method is considered, which partitions an image into several small 3D patch volumes. The supervoxel based method can reduce the required computation for classification in the new feature space regarding to the average number of voxels within the supervoxels. For example, in the case of supervoxels with an average of 50 voxels, the computation time will be 50 times faster. In general, the feature vector size of supervoxel-based methods is less than those that are based on image voxels (i.e. moving window).

Su *et al.* proposed multimodal superpixel approach for brain tumour segmentation in MRI, which was based on SLIC (Achanta *et al.*, 2012). They adopted the SLIC method formation so that the intensities from different protocols was used for pixel clustering. A specific weight can be assigned for each protocol to control its impact on final segmentation. However, they considered the voxels in a raw slice as a pixel, without considering the voxel characteristics (e.g. voxel dimensions). In this thesis, a supervoxel segmentation method is introduced to portion the volumetric MRI into 3D volumetric patches based on their homogeneity. Incorporating the voxel dimensions and slice thickness into the supervoxel calculation is the main contribution of this chapter.

The previous texton (Yu *et al.*, 2012) and (Yi and Su, 2014) Gabor-based methods have applied 2D filter bank. In this chapter, 3D Gabor filters are used for texton feature extraction from MRI volumetric datasets.

As explained in Chapter 2, each MRI sequence provides specific information about normal brain and tumour tissues. Therefore, the use of different MRI modalities can enhance the supervoxel segmentation by identifying image boundaries simultaneously across all available images. Unlike the existing methods (Wu *et al.*, 2014) in which the supervoxel are calculated using one single MRI protocol, providing a framework in which the information from several protocols is considered in calculating the distances for superpixel segmentation in Equation (4-3) may improve boundary detection of multiple tissue segmentation.

Textons have demonstrated their advantages of providing significant information to distinguish various patterns. This was also demonstrated in Chapter 4, where superpixels, taking into account the connectivity within the slice (*X*-*Y* plane), were used for texton calculation. To consider the connectivity in the *Z* direction, the approach is extended to 3D. For example, the 3D texton maps are generated in a 3D volume. In order to calculate the texton histogram as a feature vector, the texton *ID*s are required to be counted in 3D patches, i.e. supervoxels.

The main contributions of this chapter can be summarised as follows:

- A unified framework is built to classify each supervoxel using features calculated from multimodal MRI, including FLAIR, T1-weighted (with contrast), T2-weighted, *p* and *q* diffusion maps for segmentation of brain tumours.
- The supervoxel boundaries across multiple images are formed using a combination of the information from multimodal MRI, unlike the existing methods (Wu *et al.*, 2014) in which the supervoxel is calculated using a single MRI protocol.
- A novel histogram of texton descriptor for each supervoxel, calculated using a set of Gabor filters with different sizes and orientations is considered to provide better

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.

## 5.2 Methodology

The general framework of the multimodal learning based method is similar to the single modality method that was discussed in Chapter 4. The overall workflow of the proposed method is shown in Figure 5-1.

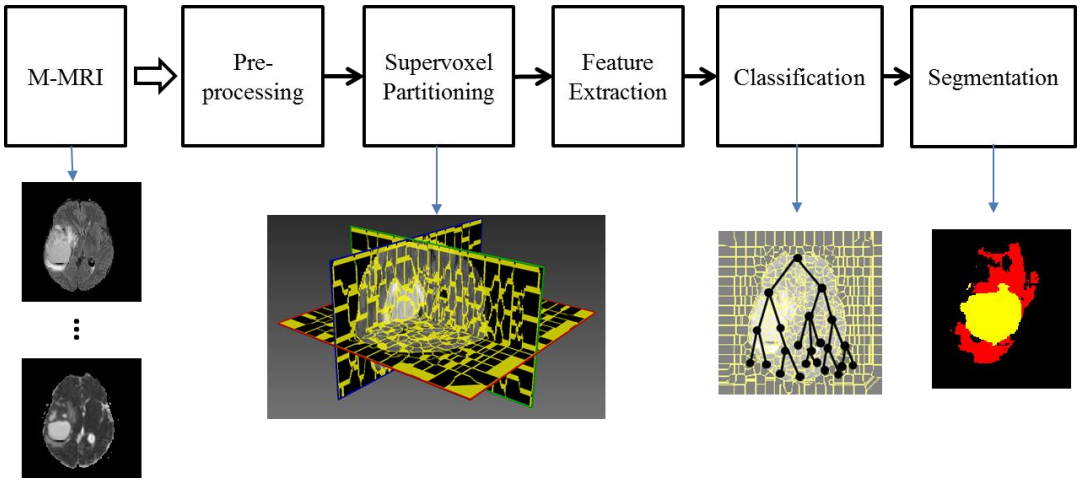



Figure 5-1 The workflow of the proposed automated multimodal MRI segmentation method for segmentation of brain tumour tissue subtypes.

The important main step of the proposed automated multimodality method is the supervoxel partition which comes after preprocessing and before feature extraction. Supervoxels are equally sized patches with similar intensity ranges calculated based on a distance matrix which is formed using a combination of multimodal images. The features are extracted for each supervoxel to further be classified into oedema, tumour core and nontumour. The following sections will describe these stages in more details with focus on the multimodality aspect of the proposed method.

### 5.2.1 Preprocessing

As explained in Chapter 4, the preprocessing stage for most MRI brain tumour segmentation comprised of skull removal, noise reduction and registration (for multimodal data). DTI data were realigned to remove eddy current distortions using eddy correction (FSL Software Library by FMRIB ("FSL," n.d.)) prior to generating *p*- and *q*- maps. Images were skull stripped using the 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. Statistical Parametric Mapping (SPM12 ("SPM - Statistical Parametric Mapping," n.d.)) was used to avoid interpolation of quantitative diffusion characteristics.

The intensities of the multimodal images are normalised with a two-step procedure: histogram matching and dynamic range normalisation. The core procedure is similar to single modal normalisation which was introduced in Chapter 2. The normalisation procedure is modified and developed for multimodal MR images, and is illustrated in Figure 5-2. Firstly, 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 (left and right pipelines in Figure 5-2). To eliminate the bias of the matched histogram to the reference case, another block ("Histogram Matching 2" in Figure 5-2) is added to the process according to (Nyúl *et al.*, 2000). In this procedure, 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 normalised to the dynamic range of the corresponding FLAIR related to that case. The FLAIR was chosen since it is the most commonly acquired protocol in the clinical tasks. Therefore, it is more available than other acquisition protocols. This is to ensure that, in the feature extraction stage, for each case, images from different protocols have similar intensity dynamic ranges.

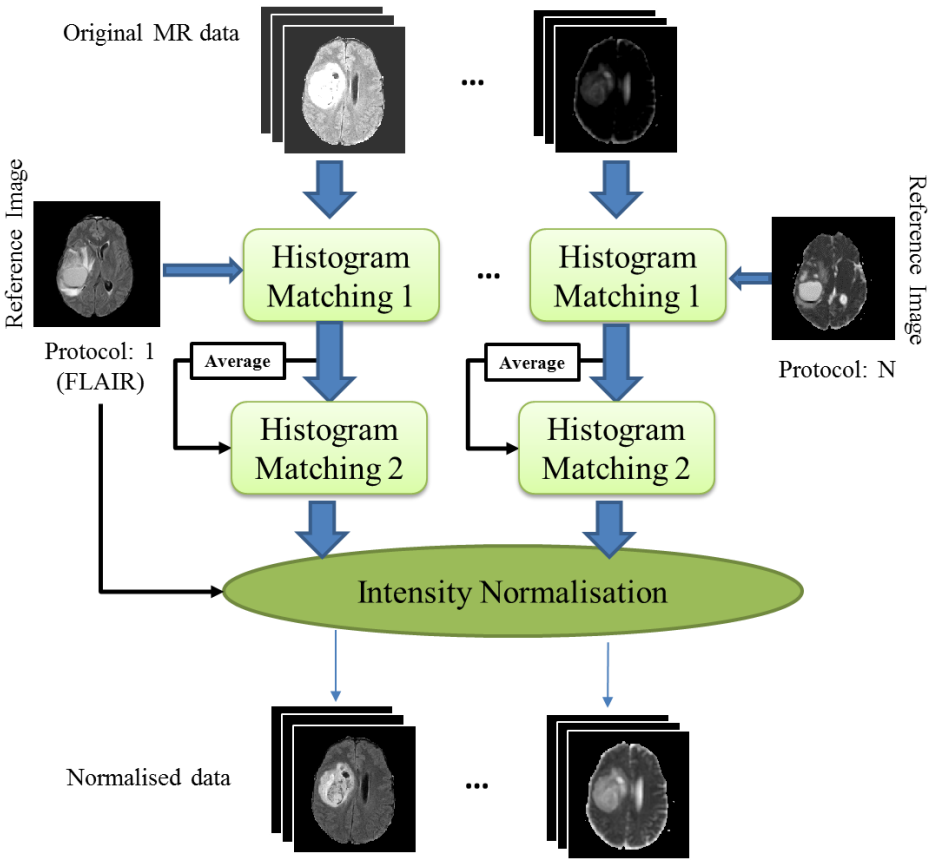
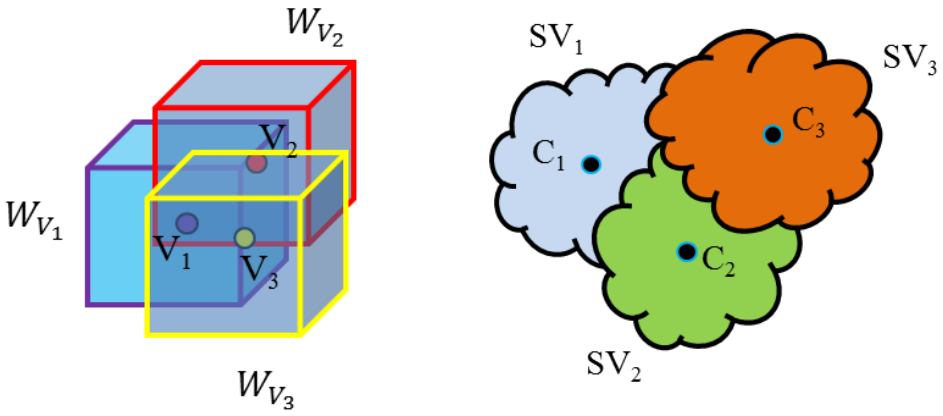



Figure 5-2 Flowchart of the multimodal normalisation and histogram matching of the MR dataset.

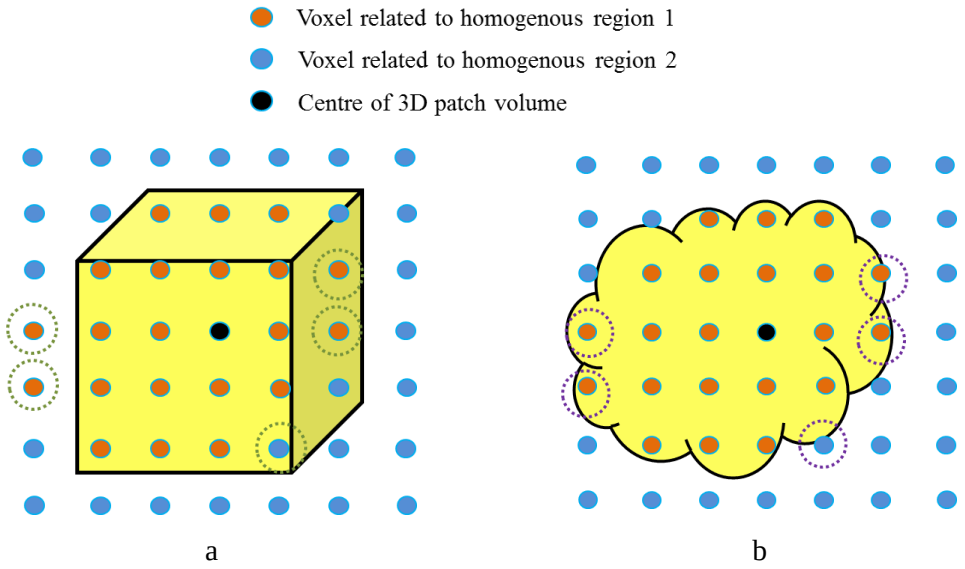
### 5.2.2 Three-dimensional Patch for Feature Extraction

Most of the voxel-wise classification algorithms used fixed 3D patches (Figure 5-4-(a)). For example, Festa *et al.* (Festa *et al.*, 2013) used a 3D cube (Figure 5-5-(a)) which is centred in each voxel for feature extraction and then assigned the features to that voxel. Instead of a fixed 3D cube, supervoxels are used as the patch for feature extraction (Figure 5-4-(b) and Figure 5-5-(b)).




a b

Figure 5-3 Fixed and flexible 3D volumes for feature extraction. a) fixed size cubic patches; $W_{V_i}$ represents the window of neighbour voxels around voxel $V_i$ . b) supervoxel patches; $SV_i$ represents the supervoxel $i$ with the centre $C_i$ .




a b

Figure 5-4 Fixed and flexible 3D volumes for feature extraction. a) fixed size cubic patch. B) flexible homogenous patch volume. The dotted circles are the homogeneous voxels related to regions 1 and 2 that are assigned differently in both patch systems.

### 5.2.3 Multimodal Supervoxel Segmentation Algorithm

The supervoxel method clusters an image into a predefined number of three dimensional portions, which have similar intensity range (Figure 5-5). Here, the simple linear iterative clustering (SLIC) superpixel method (Achanta *et al.*, 2012) is extended to extract 3D supervoxels for the segmentation of a brain tumour. A brief description of SLIC and the modifications for application in three dimensional multimodal MRI is given below.

In the proposed method, the height of the initial supervoxel grids 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 (i.e. $R_x$ and $R_y$). Therefore, $R_s$ is obtained using

$$R_S = \frac{R_x}{R_y}. \tag{5-1}$$

In the case of the clinical MRI dataset, the resolutions in $X$ and $Y$ directions are the same ($R_x$= $R_y$), so $R_s$=1. All the data were co-registered in the preprocessing stage, therefore the slice thickness, $R_t$, for each dataset is consistent through all the slices. The registration of the data is an important prerequisite of performing this multimodal supervoxel segmentation. Assuming the initial width of a supervoxel to be $W_S$ (voxels), its initial height, $H_S$, is calculated from the ratio of slice spatial resolution to slice thickness

$$H_S = round\left(W_S \times \frac{R_S}{R_t}\right). \tag{5-2}$$

The minimum value for supervoxel height, $H_S$, is considered 3, whilst, $H_S$=1 results in 2D segments which are considered as superpixels. The spatial resolutions of the images are 1 mm × 1 mm × 1 mm for the BRATS, and 0.9375 mm × 0.9375 mm × 2.8 mm for the clinical datasets. In the case of BRATS dataset, $R_S$ = $R_t$ = 1. Supervoxel side size, $W_S$, is integer. Therefore, according to Equation (5-2), superpixel height is equal to its side size ($H_S$ = $W_S$). For the clinical dataset, $R_S$ = 1 and $R_t$ = 2.8. Therefore, for a predefined $W_S$ = 8, the supervoxel height is calculated as

$$H_S = round\left(6 \times \frac{1}{2.8}\right) = round\,(2.857) = 3\,. \tag{5-3}$$

Figure 5-5 presents a schematic illustration of calculating the initial supervoxel parameters from the MR input data considering the voxel resolutions.

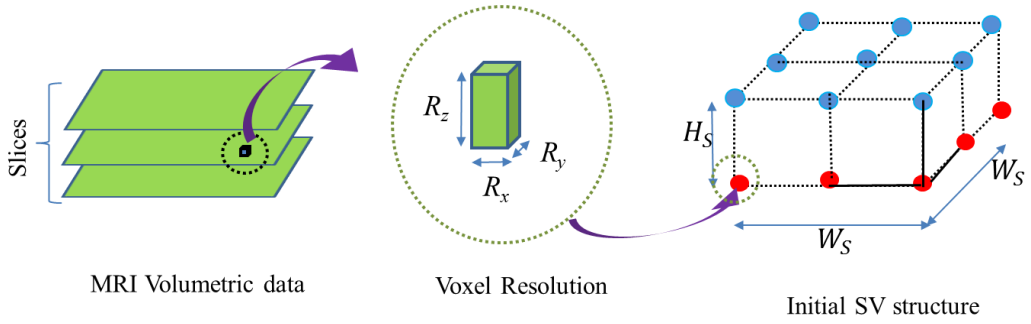



Figure 5-5 Initial supervoxel structure calculation based on MR voxel resolution parameters. $W_s$ and $H_s$ represent initial supervoxel width and height. $R_x$ and $R_y$ relate to spatial resolution of the voxel in $XY$ plane, and $R_z$ relates to slice thickness.

Supervoxel computation based on SLIC is an iterative procedure in which the initial iteration starts from the initial SV grids. The geometrical centres of the initial grids are considered as supervoxel region centres (Figure 5-4-(a)). 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 (Figure 5-4-(b)). 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 voxels using

$$d_c = \sqrt{(I_j - I_i)^2}, \tag{5-4}$$

where, $I_i$ and $I_j$ are the normalised intensity values of the corresponding voxels. This is the intensity distance for a single modality. The extension of the intensity distance to multimodal will be explained in Equation (5-7).The location distance, $d_s$, between the two voxels *i* and *j* is calculated using

$$d_s = \sqrt{(R_x(x_j - x_i))^2 + (R_y(y_j - y_i))^2 + (R_z(z_j - z_i))^2}, \tag{5-5}$$

where $(x_i, y_i, z_i)$ are the coordinates of voxel *i* and $R_x$, $R_y$ and $R_z$ are the voxel resolutions.

The total distance measure (Achanta *et al.*, 2012) is then calculated using

$$D = \sqrt{d_c^2 + \left(\frac{d_s}{W_S}\right)^2 m^2}, \tag{5-6}$$

where *m* is the compactness coefficient. As discussed in Chapter 4 (Section 4.2.4), the intensities are normalised into the range [0, 1] to ensure a normalised value for compactness factor, *m*. A higher value of *m* results in more compact segments and a lower value creates more flexible boundaries.

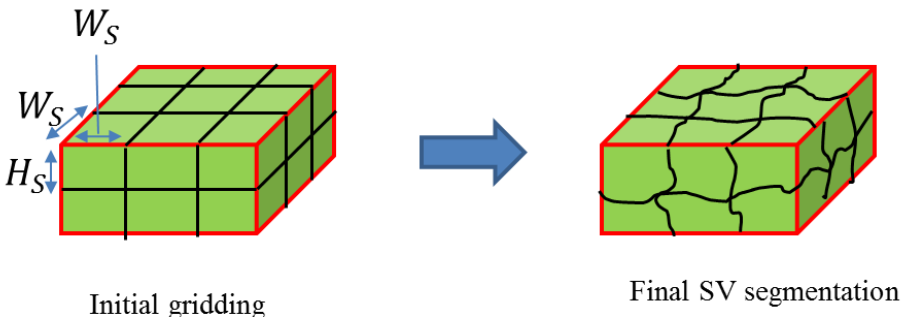
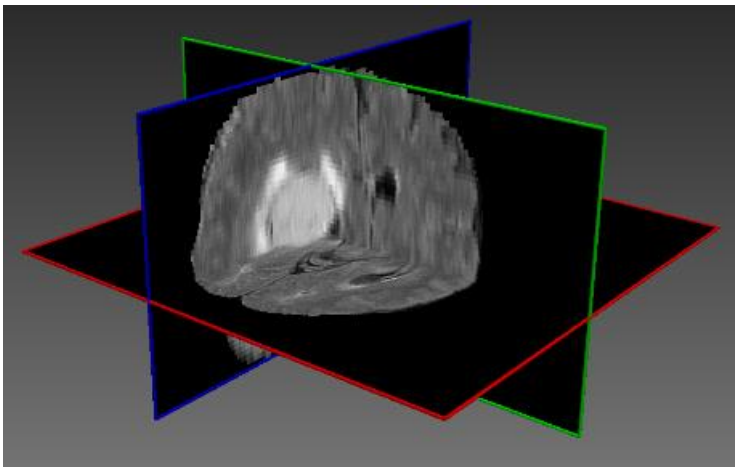
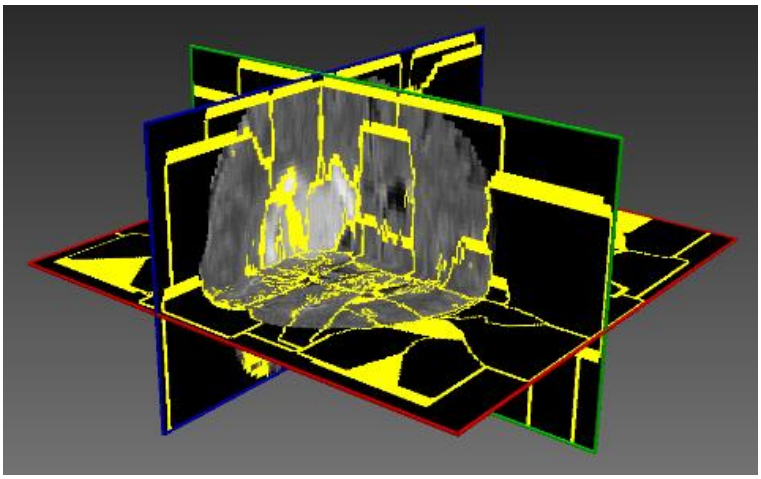



Figure 5-6 Iterative SV portioned the volume into 3D homogenous segments.

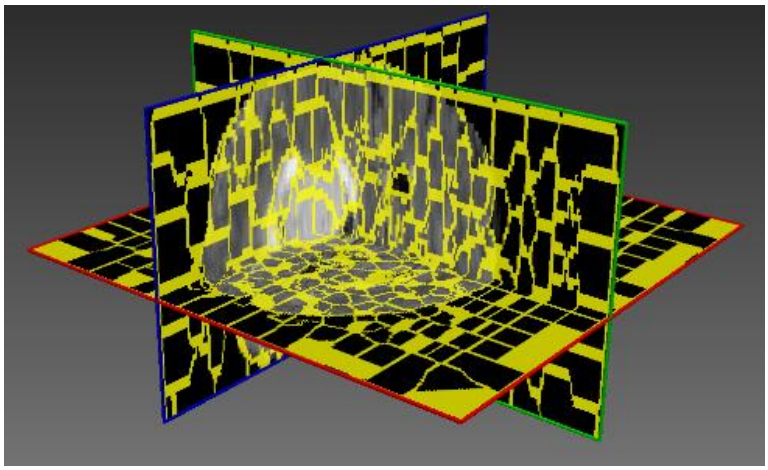
Figure 5-7 shows the supervoxel segmentation of a brain tumour using MRI FLAIR with two different initial grid sizes in a combination of axial, sagittal and coronal planes.



a



b



c

Figure 5-7 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).

Supervoxel segmentation of multimodal MRI data is not straightforward as tissue boundaries apparent on one MRI modality, for example, on T1-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. The hypothesis to tackle this problem and determine a multimodal supervoxel cluster is to adapt the supervoxel intensity distance Equation (5-4) in a multidimensional formation and apply this across all MRI modalities. In general, 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

$$d_c = \sqrt{\left(I_{Voxel,P_1} - I_{Center,P_1}\right)^2 + \cdots + \left(I_{Voxel,P_N} - I_{Center,P_N}\right)^2}, \tag{5-7}$$

where, $I_{Voxel,Pi}$ is the grey-level intensity corresponding to the *voxel* in protocol $P_i$. Figure 5-8 shows a schematic illustration of the multimodal supervoxel distances which were explained in Equation (5-7). It should be noted that different protocols have similar weighting in this formulation.

Figure 5-9 shows the framework for the multimodal supervoxel segmentation method.

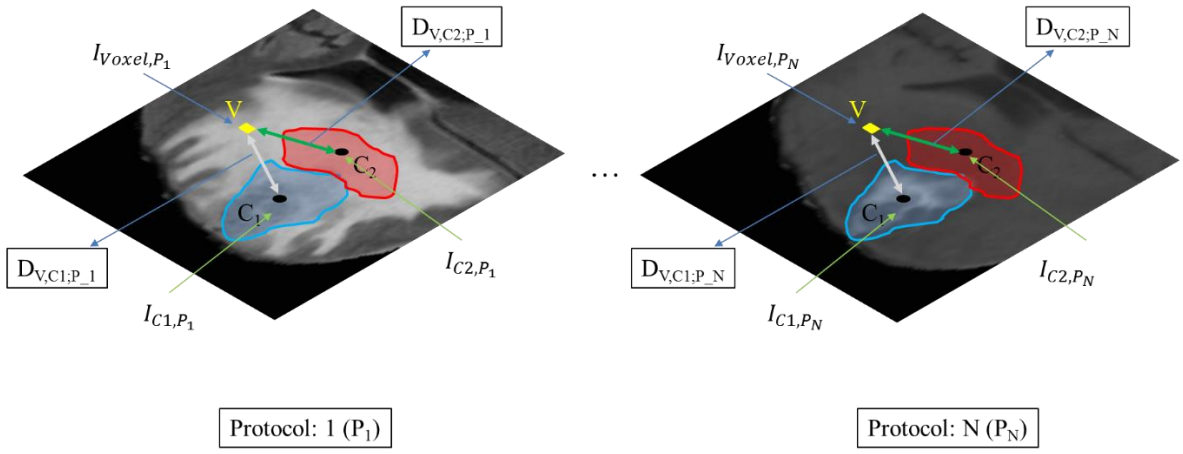



Figure 5-8 Schematic illustration of the distances for multimodal supervoxel segmentation from protocols: 1 and N. The distances are calculated for the voxel *V* (yellow point) from two adjacent supervoxels with centre $C_1$ (blue region) and $C_2$ (red region). $I_{Voxel,Pi}$ and $I_{Ck,Pi}$ represents the intensity of *V* and $C_k$ in protocol *i*. $D_{V,Ck;P\_i}$ is the total intensity distance between voxel *V* and $C_k$ in protocol *i*.

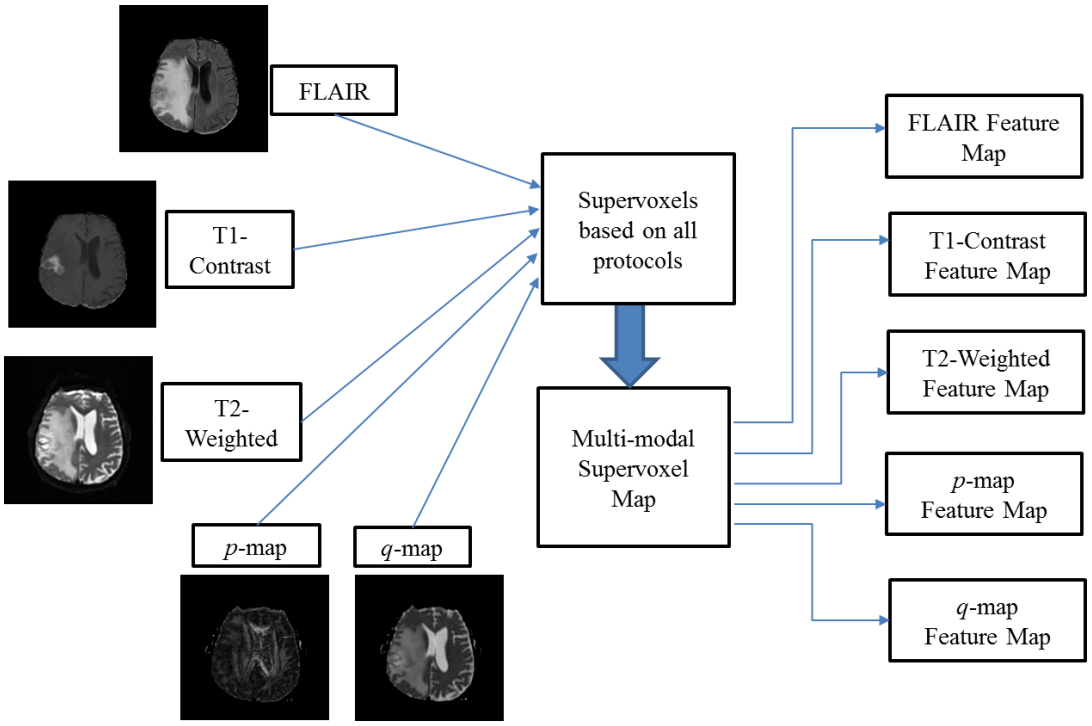



Figure 5-9 Framework of multimodal supervoxel segmentation.

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 other images from different protocols. The results do not depend on the majority, and even an edge in one single modality appears in the final SV segmentation. An example of this case is shown in Figure 5-10. The supervoxel map generated from the multimodal segmentation method is overlaid on both the FLAIR image (top row of the first column in Figure 5-10) and the *p*-map (bottom row of the first column in Figure 5-10). The middle and the last columns in Figure 5-10 show two corresponding close-up areas indicated in the FLAIR and the *p*-map images (yellow and orange rectangles). As can be seen, the middle column of the Figure 5-10 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 inverse effect can be seen in the right column of Figure 5-10. The multimodal clustering method for supervoxel calculation provides good image boundaries even when boundaries are not clear in one image modality.

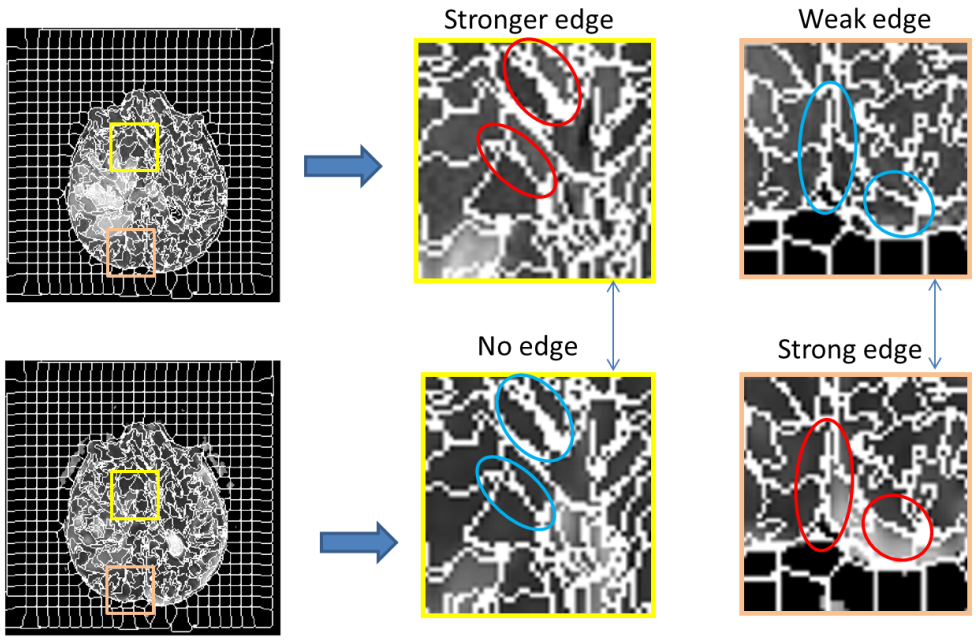



Figure 5-10 An example of using a multimodal approach to improve supervoxel boundaries by finding the edges which appear weak in one modality (blue ovals), but are apparent in the other modality (red ovals). (a) Upper image: FLAIR image overlaid by multimodal supervoxel segmentation, lower image: *p* 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.

At the end of supervoxel segmentation, some isolated voxels might appear. The label of these voxels is different from their surrounding voxels, which is considered as the noise of the supervoxels. A post-processing procedure is designed to reduce the isolated voxels. The label of the voxels connected to the isolated voxels are counted. Then the isolated voxel is relabelled to the major connected class.

A comparison of the supervoxel segmentation of the tumour core calculated from a single MRI modality (FLAIR) and from multimodal MRI (FLAIR, T1-weighted (with contrast), T2-weighted, *p*- and *q*-maps) is shown in Figure 5-11. As can be seen, there are misalignments between supervoxel's boundaries (computed from FLAIR) and the ground truth boundaries (see black ellipse in Figure 5-11 (f)), whilst multimodal supervoxels show improvement in boundary alignment to the tumour core (see black ellipse in Figure 5-11(i)).

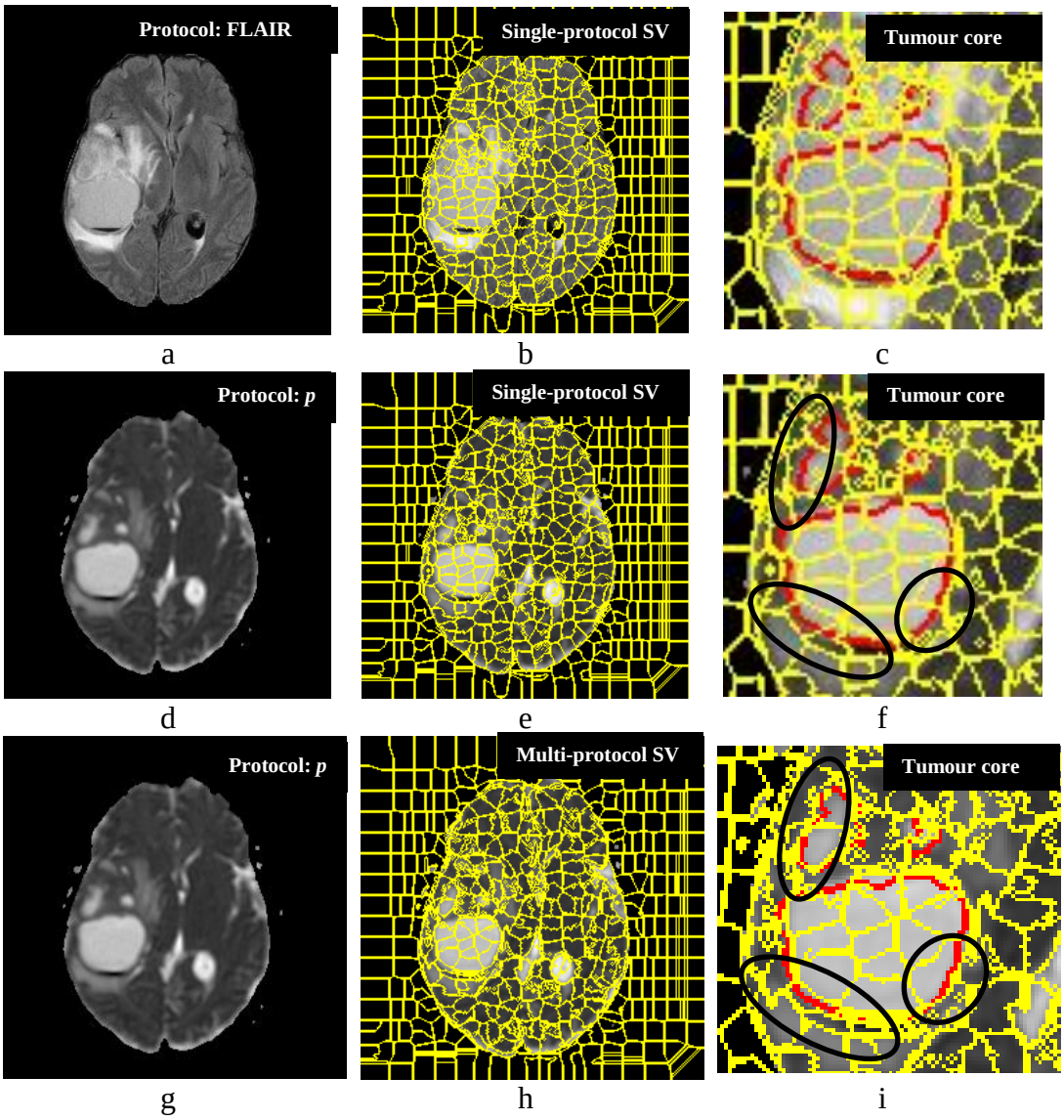



Figure 5-11 One comparison example of tumour core supervoxel segmentation (SV) using single modality and multimodal MRI approaches. (a) FLAIR, (b): overlay of the corresponding supervoxels calculated using single modality (FLAIR), (c): zoomed-in of (b) on tumour area (to show the details of the SV boundaries) and overlay of tumour core (ground truth from manual delineation shown in red); (d): protocol *p*-map, (e): Supervoxels calculated using single imaging modal (FLAIR) overlaid on image protocol *p*, (f): zoomed-in view of (e) on tumour area and overlay of tumour core (red). (g): protocol *p*, (h): Supervoxels calculated using multimodal (FLAIR, T1-contrast, T2-weighted, *p* and *q*-maps) overlaid on image protocol *p*-map. (i): zoomed-in of (h) on tumour area and overlay of tumour core (red). The boundaries surrounded by black ellipses in (f) and (i) 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.

### 5.2.4 Feature Extraction

Image features are calculated from each supervoxel in 3D space. Among the features that were used in Chapter 4 (Section 4.2.4), first order statistical features and texton features are utilised by extending to 3D. The first order features are similar to those which were discussed in Chapter 4 (Section 4.2.4).

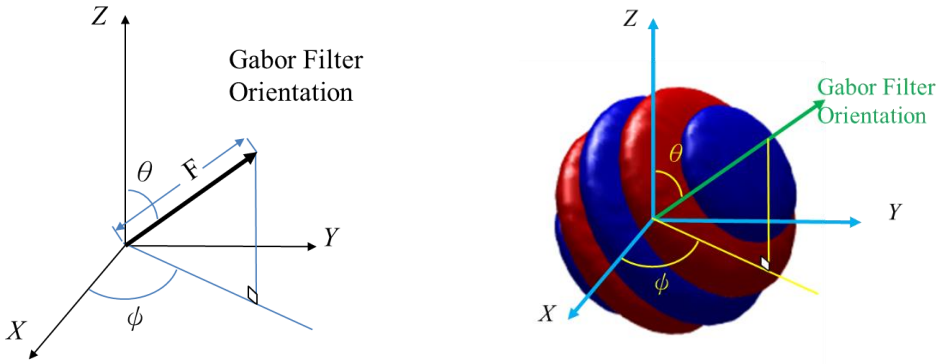
For the texton feature calculation performed in a 3D supervoxel, 3D Gabor filters [31] are used which are defined as

$$G(x, y, z; F, \theta, \varphi) = \frac{1}{(2\pi)^{3/2}\sigma^3}\exp\left(-\frac{x^2+y^2+z^2}{2\,\sigma^2}\right)\exp\left[i2\pi\left(F_x\,x + F_y\,y + F_z\,z\right)\right], \qquad (5\text{-}8)$$

where

$$\begin{aligned} F_x &= F\,\sin\theta\,\cos\varphi\,, \\ F_y &= F\,\sin\theta\,\sin\varphi\,, \\ F_z &= F\,\cos\theta\,. \end{aligned} \qquad (5\text{-}9)$$

In the Equations (5-8) and (5-9), $\sigma$ is the standard deviation of Gaussian envelope, $F$ is the radial centre frequency, $\theta$ is the orientation angle from the $Z$ axis, ψis the orientation angle of the projection on the X-Y plane from X axis. The orientation angles, $\theta$ and ψ, are shown in Figure 5-12. The sinusoid wavelength is $\lambda$ = 1 / F.




a b

Figure 5-12 Gabor filter orientation and the corresponding angles, i.e. $\theta$ and ψ. a) orientation axes illustration, b) a 3D sample for central frequency $F$, and angles $\theta$ and ψ. The blue and red parts correspond to the positive and negative values of Gabor filter, respectively.

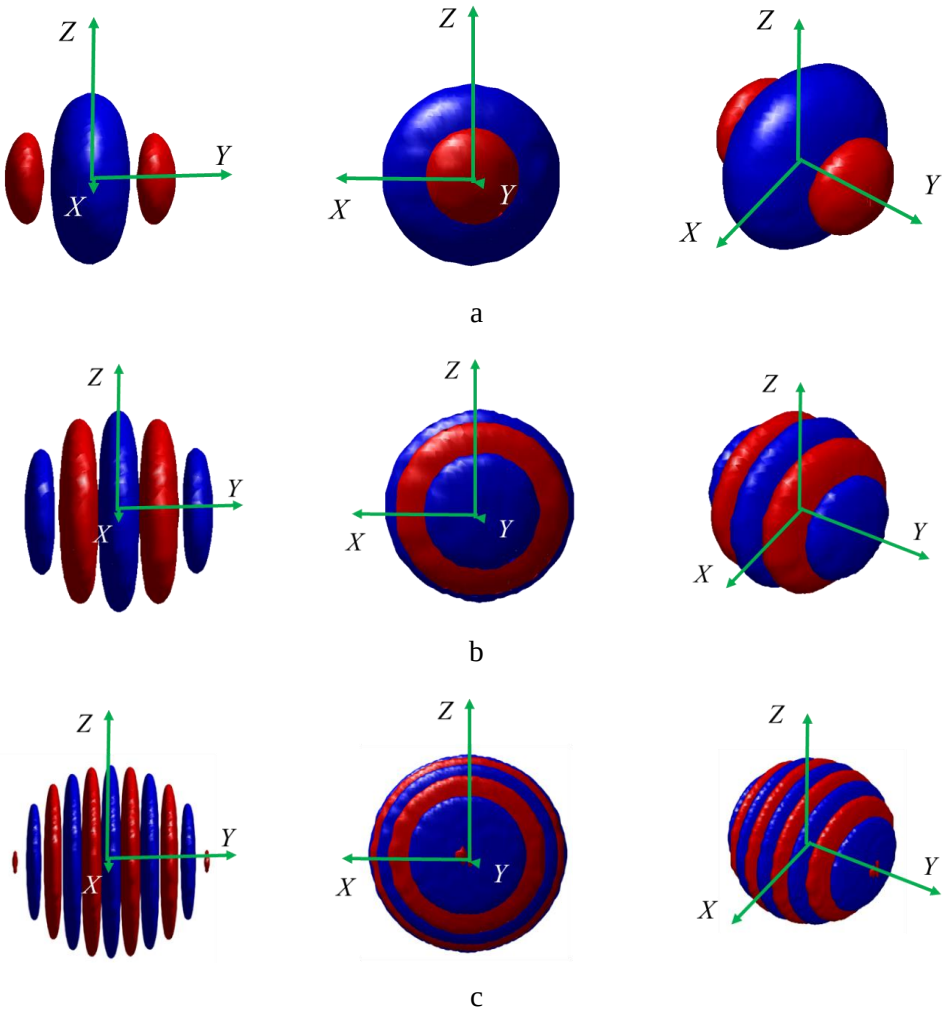


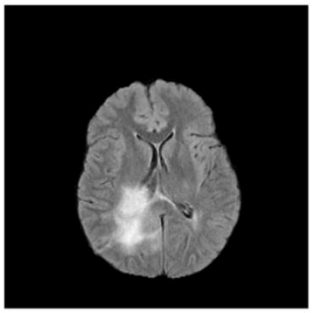

Figure 5-13 3D Gabor filters with different frequencies from various angle views. The blue and red parts correspond to the positive and negative values of Gabor filter, respectively. a) $\lambda$ = 1.2, b) $\lambda$ = 1.0, c) $\lambda$ = 0.8.

Figure 5-13 shows 3D Gabor filters with various frequencies from differnet angle vews.

To minimise the number of parameters for texton calculations, the orientation is varied in the *X*-*Y* plane. Therefore, the angle $\theta$ is fixed to 90 degrees and the filters are generated with different angles ψ. The reason for that is to arrange the filters in spatial dimensions of the MR images within the slice plane. The angle $\theta$ = 90 is selected since the resolution of the clinical data is higher in the *X*-*Y* plane, compared to the *Z* axis. Other angular features were not considered in this study to avoid increasing the parameters, complexity and computation time. It should be noted that excluding other $\theta$ angles will reduce the accuracy of feature extraction.

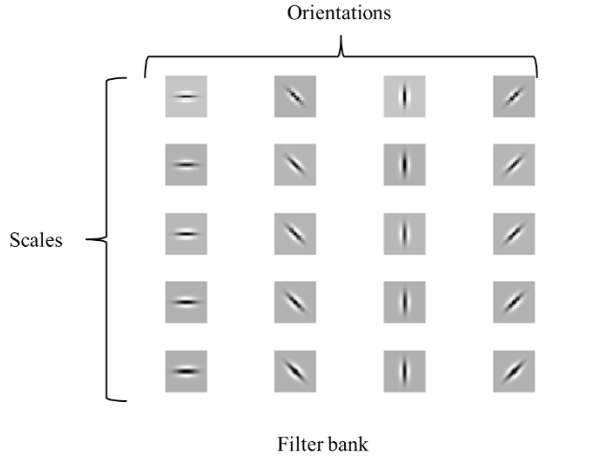
The Gabor filter parameters were chosen empirically. Six different filter directions were considered: [0°, 30°, 45°, 60°, 90°, 120°] with filter sizes from 0.3 to 1.5 at steps of 0.3. The

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. An example of the Gabor filter bank is shown in Figure 5-14.

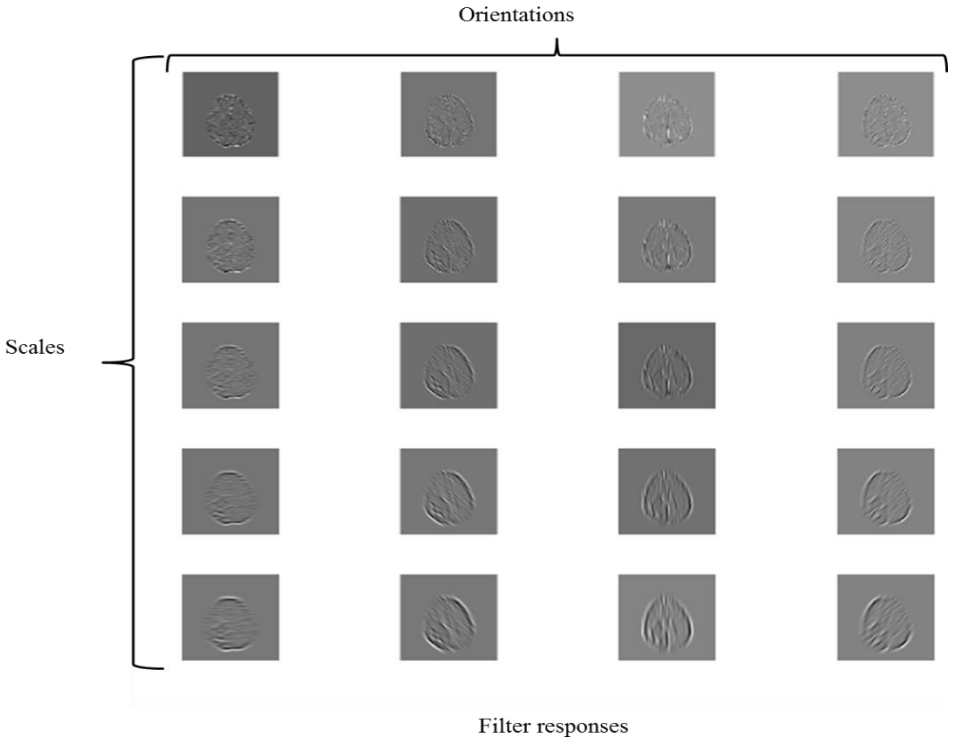


Original image

a




b




c

Figure 5-14 a) Example of Gabor filter bank with different filter size and directions and their responses. a) Original FLAIR image, b) Gabor filters with different filter size and directions: rows are corresponding to different filter sizes: [0.3, 0.6, 0.9, 1.2, 1.5] and columns represent different directions: [0$^{o}$, 45$^{o}$, 90$^{o}$, 135$^{o}$] c) the corresponding filter responses obtained by convolving the filters (b) with the original image (a).

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 $k_{texton} = 5$ representing tumour core, oedema and normal brain tissues. To reduce the computation time for clustering, the lowest number of clusters ($k_{texton} = 5$) that are capable of separating the tumour core and oedema from normal brain in the training set was chosen. Histograms of the texton parameters were then calculated for each supervoxel using the generated texton map.

Texton *ID* extraction was explained in Chapter 4. Here, the *k*-means clustering is performed on the whole image volume, instead of slice-by-slice computation. The histogram of the texton *ID*s is calculated for each supervoxel which is considered as a texton feature vector of that SV. It should be noted that the *ID*s are sorted ascendingly for each protocol based on the average intensity of the group of pixels within each cluster. Figure 5-15 shows examples of texton *ID* histograms for supervoxels of normal brain and tumour.

The extracted features and their numbers are summarised in Table 5-1. In total, there are 21 features for each MR image, so there are 105 features across the multimodal MRI data (FLAIR, T1-contrast, T2-weighted, *p*- and *q*-maps). The feature calculations are performed on supervoxels and extracted for each individual MR modality. Then, the extracted features from each modality are concatenated based on the corresponding superpixel to form the final multimodal feature vector.

Table 5-1 Summary of the features and their corresponding numbers which are used for the proposed 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 |

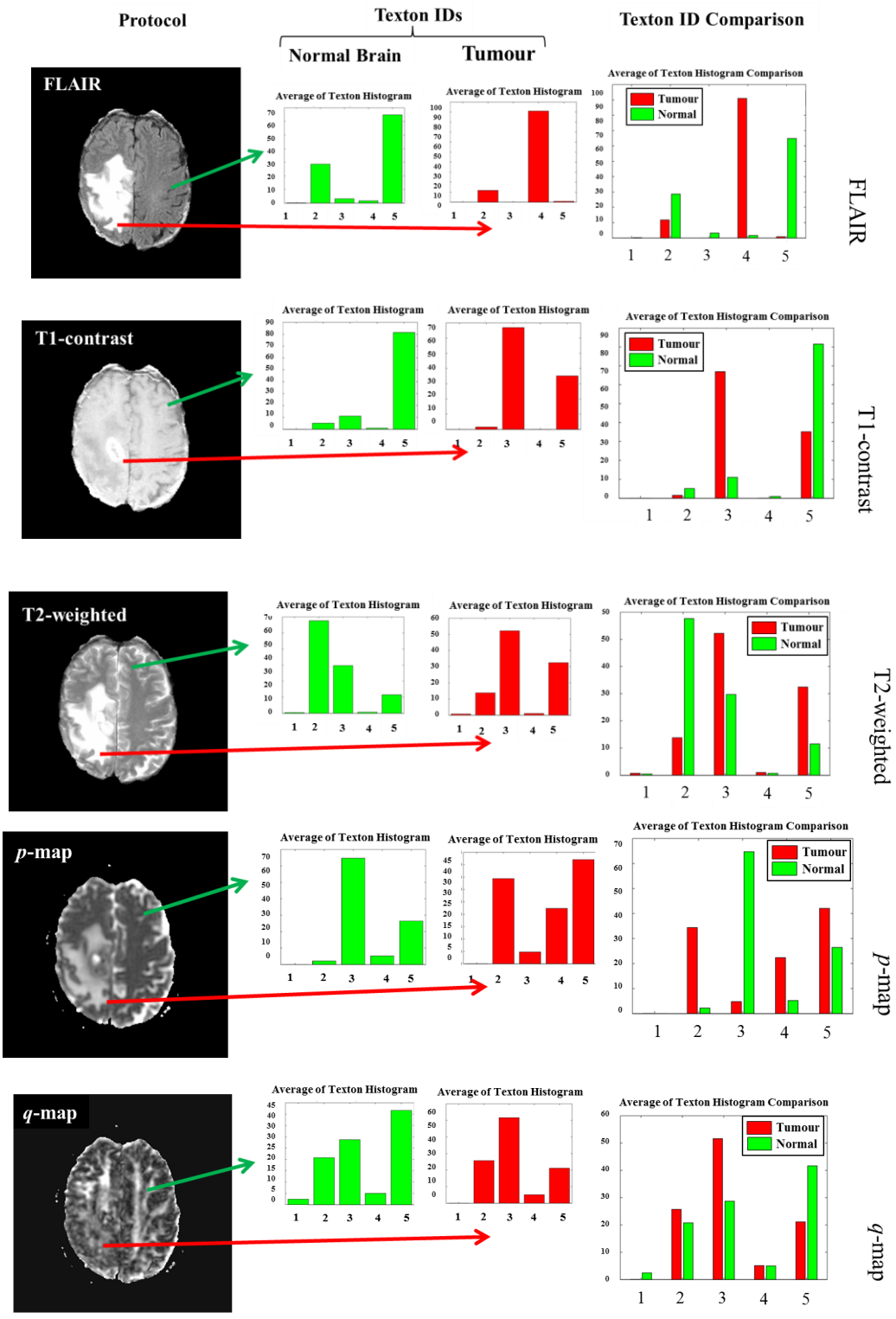



Figure 5-15 Texton *ID* histogram for tumour and normal brain. Average of the texton histogram of the regions inside the corresponding regions.

### 5.2.5 Classification of the Supervoxels

In the proposed multimodal method, all supervoxels within the brain region 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.

In Chapter 4, ERT was used for classification of superpixels for the features extracted from one single modality (i.e. FLAIR). In this chapter, the features are extracted from several protocols (5 protocols for the clinical datasets and 4 protocols for the BRATS datasets). Therefore, the number of feature increases. On the other hand, all the feature will be fed to the classifier, which will increase the complexity and dimensionality of the classification problem. To decrease the computation time, RF will be used since is it is faster to perform, compared to the ERT.

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 ($k_{attribute}$) selected to perform the random splits for a specific number of features $N_{feature}$ is $k_{attribute} = \sqrt{N_{feature}}$. For single modality and multimodal experiments, 5 and 10 attributes are selected, respectively. The reason for selecting these parameters will be discussed in the Section 5.3.2.

In the training stage, the supervoxels are split into three classes: normal brain tissue, tumour core and oedema. Supervoxels that 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. The proposed multimodal 3D supervoxel method is referred as SV_RF.

## 5.3 Experiments and Results

This section will describe the evaluation results of the proposed automatic multimodal brain tumour segmentation method. First, the datasets will be mentioned, followed by the experimental setting and evaluation methods. Thereafter, the results will be explained and the statistical analysis will be explained**.**

### 5.3.1 Dataset and Implementation

Two datasets were analysed that were also used in Chapter 4:

1. The clinical dataset for training and validation of the algorithm. This dataset consists of 11 patients with the acquisition protocols including C-MRI (FLAIR, T1-contrast, T2-weighted) and DTI (*p*- and *q*-map). The details of this dataset were explained earlier in Chapter 2 (Section 2.6.1).

2. The publicly available MICCAI BRATS 2013 dataset for further comparison and assessment of the robustness of the method. This dataset consists of 30 brain tumour patients (20 high-grade and 10 low-grade glioma) entered the study, which include conventional MRI protocols. The details of this dataset were explained earlier in Chapter 2 (Section 2.6.2).

**Evaluation**

For the clinical dataset, both tumour core and complete tumour (including tumour core and oedema) were evaluated, while the evaluation for the BRATS dataset was provided using VSD standard combination, which are tumour core (including necrosis, enhancing and non-enhancing) and complete tumour.

DSC is used to evaluate the overlap ratio between the segmentation results and the manual segmented gold standard which was explained in Chapter 2 (Section 2.7.2 and Equation (2-19)).

The leave-one-out approach is used to train and test the models.

**Experimental Setting**

To evaluate the performance of the proposed multimodal supervoxel classification method, three experiments are performed using different MRI modalities for the tumour parts including core and oedema:

1. FLAIR only
   In the first experiment, supervoxels are calculated based on a single modality, i.e. FLAIR image only. As discussed in Section 5.2.4 and shown in Table 5-1, for each supervoxel, there are 21 features extracted from each modality. Therefore, in total 21 features for FLAIR only were extracted from both datasets.

2. C-MRI data
   The second experiment was designed to evaluate the multimodal supervoxel based segmentation method. Supervoxels are calculated using Equation (5-6) based on

different MRI modalities, i.e. C-MRI data in both datasets. 63 features for the clinical C-MRI data (e.g. FLAIR, T2-weighted and T1-contrast), and 84 features for BRATS C-MRI data (e.g. FLAIR, T1-weighted, T2-weighted and T1-contrast) were extracted based on the corresponding supervoxel map.

3. C-MRI+DTI
   The third experiment was designed to evaluate the effect of adding the information from DTI modalities to the proposed multimodal segmentation method. The supervoxels are calculated using Equation (5-6) based on different MRI modalities, i.e. C-MRI+DTI (*p*- and *q*-maps) for the clinical datasets. In total, 105 features were extracted from the corresponding protocols based on the supervoxel map. Since the publicly available BRATS dataset does not include the DTI protocols, this experimental set is not applicable for it. Therefore, the experiment is only conducted for the clinical dataset and compared to C-MRI experiments.

The random forest classification is performed in each of the three experiments to classify each supervoxel into normal brain tissue and tumour. RF parameters $N_{tree}$ and $D_{tree}$ are chosen the same for all the experiments.

### 5.3.2 Parameter Selection

In the case of 2D superpixel calculation, which was presented in the previous single modality work (Soltaninejad *et al.*, 2016) and explained in Chapter 4 (Section 4.3.5), the optimal initial superpixel size of 5 was suggested. In the case of 3D supervoxels, the z direction is determined from the slice thickness and image resolutions based on Equation (5-2). As described in Chapter 2 (Section 2.6.1), the clinical dataset had different resolutions and all multimodal MRI data were co-registered to DTI with voxel dimensions 0.9375 mm × 0.9375 mm × 2.8 mm, while the resolution for the BRATS dataset was 1 $mm^3$ isotropic voxel dimensions. Therefore, due to the different resolutions between the two datasets, the supervoxel initial sizes were selected to be 8 × 8 × 3 for the clinical data, and 5 × 5 × 5 for the BRATS data. The compactness factor, *m*, was selected by visually inspecting the supervoxel boundaries and area in different views. i.e. coronal, axial and sagittal for some of the training cases. The value of $m = 0.05$ (Equation (5-6)) presented coherent boundaries. Since $d_s$ in Equation (5-6) is different for 2D and 3D spaces, different values were obtained for *m* in both spaces.

To select the optimum RF parameters, different ranges of the number of trees and depth were assessed on the clinical data. To select the optimal RF parameters (i.e. number of trees and depth), the *k*-fold validation was used with $k = 4$. For the clinical dataset, the folds are sets of

[3, 3, 3, and 2] randomly selected non-repetitive datasets, and for the BRATS dataset, the folds are sets of [8, 8, 7, and 7] randomly selected non-repetitive datasets. Classification accuracy was calculated for the testing fold in each iteration with different $N_{tree}$ and $D_{tree}$. Values of classification accuracy were averaged over all folds to determine the effects of $N_{tree}$ and $D_{tree}$, which are presented in Figure 5-16 and Figure 5-17, respectively. Figure 5-16 and Figure 5-17 show that a RF with $N_{tree}$ = 50 and $D_{tree}$ = 15 provides an optimum generalisation and accuracy. It should be noted that $N_{tree}$ = 100 also provided optimal results. However, due to minimisation of computational costs, $N_{tree}$ = 50 was selected. These optimal parameters were also directly used in the analysis of the BRATS dataset.

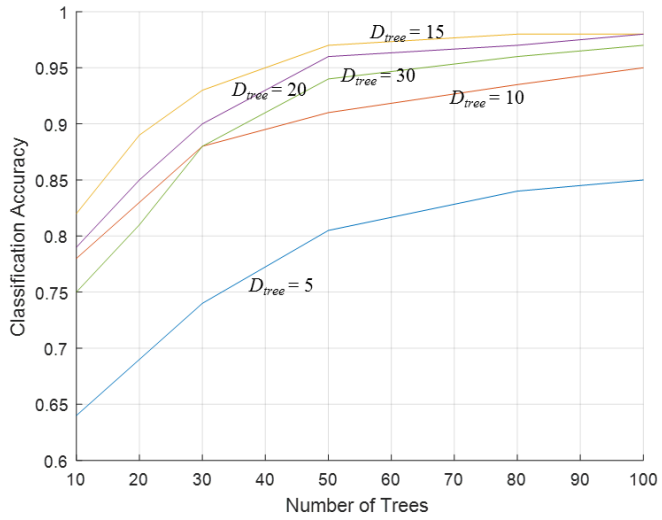



Figure 5-16 Effect of number of trees on RF classification accuracy with different depths.

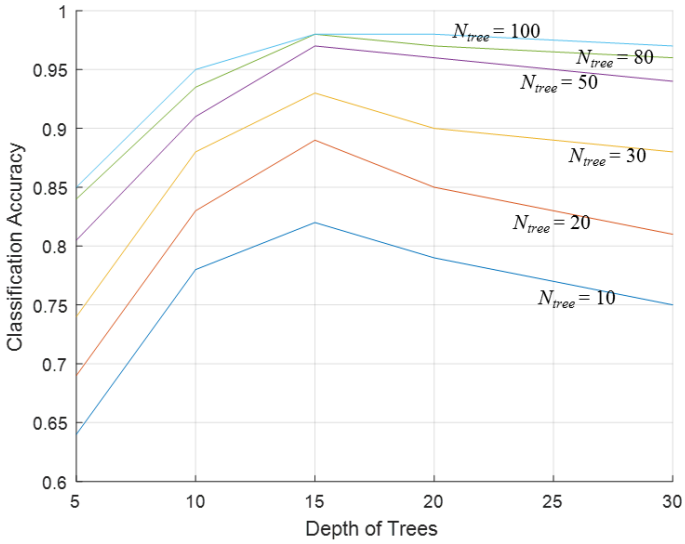



Figure 5-17 Effect of tree depth on RF classification accuracy with different numbers of trees.

The importance of the features which are extracted from different protocols are quantified by investigating the distribution of the features in the RF. The ratio of feature is obtained by calculating the proportion of the nodes related to a specific feature type (i.e. features form each individual protocols) with respect to all the nodes in the forests. Table 5-2 presents the ratio of features which are selected from each acquisition protocol using the RF from two multimodal experiments, i.e. C-MRI and C-MRI+DTI. The results show that most of the features (61%) were selected from the FLAIR, which represents the importance of FLAIR for tumour segmentation. When DTI is added to the feature space, 24% of features are selected from its protocols, from those 16% is relate to *p*-map and 8% to *q*-map. The presence of DTI also slightly reduces the proportion of corresponding features from the experiment with C-MRI modalities alone. The experimental results presented in the next sections (Sections 5.3.4 and 5.3.6) confirm that that *p*- and *q*- maps improve the overall segmentation of tumour core.

Table 5-2 Ranking of the features from each individual protocol in different multimodal experiments on the clinical dataset, based on their repetition in nodes of the forests of a RF with $N_{tree}$ = 50 number of trees and $D_{tree}$ = 15.

| **Experiment** | **FLAIR** | **T1-contrast** | **T2-weighted** | ***p*-map** | ***q*-map** |
|---|---|---|---|---|---|
| C-MRI | 0.61 | 0.15 | 0.24 | - | - |
| C-MRI+DTI | 0.49 | 0.09 | 0.18 | 0.16 | 0.08 |

### 5.3.3 Supervoxel Classification Results

Table 5-3, Table 5-4 and Table 5-5 show the results of supervoxel classification using the clinical dataset for the three experiments: FLAIR only, C-MRI, and C-MRI+DTI, respectively. The results are presented separately for each tumour part (core and oedema) and complete tumour. For the tumour core, precision, sensitivity and BER are 69.49%, 65.39% and 0.18 for FLAIR only, while for C-MRI the corresponding values are 73.64%, 69.67% and 0.15 and for C-MRI+DTI they are 83.44%, 74.62% and 0.13. For oedema, precision, sensitivity and BER are 84.17%, 79.28% and 0.11 for FLAIR only, while for C-MRI the corresponding values are 85.63%, 80.59% and 0.10 and for C-MRI+DTI they are 88.53%, 84.57% and 0.08. For the complete tumour, for FLAIR only, precision, sensitivity and BER are 88.16%, 81.88% and 0.09, while for C-MRI the corresponding values are 89.54%, 83.66% and 0.09 and for C-MRI+DTI they are 92.22%, 86.25% and 0.07.

Table 5-3 Classification results (average values over all the LOO-CV) for superpixels using single MRI protocol (FLAIR).

| No | ID | Core | | | Oedema | | | Whole | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Precision | Sensitivity | BER | Precision | Sensitivity | BER | Precision | Sensitivity | BER |
| **1** | 37227 | 87.32 | 84.51 | 0.08 | 81.97 | 72.62 | 0.14 | 92.46 | 83.79 | 0.09 |
| **2** | 37182 | 79.14 | 68.03 | 0.16 | 72.44 | 62.25 | 0.20 | 77.95 | 58.15 | 0.21 |
| **3** | 37197 | 70.56 | 64.76 | 0.18 | 76.99 | 86.98 | 0.07 | 79.53 | 86.03 | 0.07 |
| **4** | 37230 | 78.43 | 74.12 | 0.13 | 78.37 | 83.84 | 0.10 | 84.25 | 88.96 | 0.06 |
| **5** | 37253 | 91.32 | 64.25 | 0.18 | 88.29 | 82.57 | 0.09 | 92.40 | 91.86 | 0.04 |
| **6** | 37243 | 67.50 | 59.44 | 0.20 | 73.70 | 74.91 | 0.14 | 80.37 | 79.50 | 0.11 |
| **7** | 37256 | 59.85 | 67.78 | 0.16 | 90.85 | 84.24 | 0.08 | 92.72 | 88.17 | 0.06 |
| **8** | 37302 | 63.14 | 63.33 | 0.19 | 93.66 | 79.11 | 0.11 | 94.76 | 78.82 | 0.11 |
| **9** | 37394 | 57.27 | 53.65 | 0.24 | 84.56 | 89.56 | 0.06 | 88.95 | 89.75 | 0.05 |
| **10** | 37396 | 58.45 | 61.15 | 0.19 | 92.53 | 71.23 | 0.15 | 93.55 | 71.56 | 0.14 |
| **11** | 37218 | 51.42 | 58.25 | 0.21 | 92.53 | 84.76 | 0.08 | 92.79 | 84.14 | 0.08 |
| **Mean** | **All** | **69.49** | **65.39** | **0.18** | **84.17** | **79.28** | **0.11** | **88.16** | **81.88** | **0.09** |
| **STD** | **All** | **13.05** | **8.38** | **0.04** | **7.93** | **8.18** | **0.04** | **6.38** | **9.81** | **0.05** |

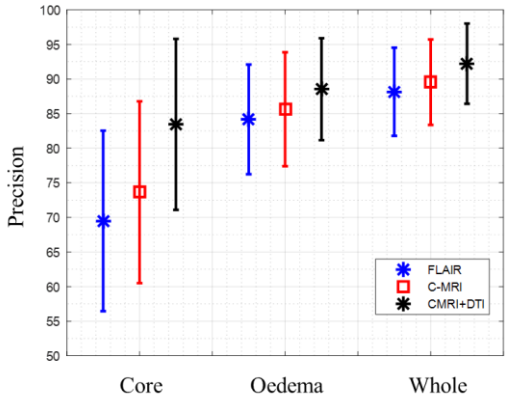
Table 5-4 Classification results (average values over all the LOO-CV) for superpixels using conventional MRI protocols (FLAIR, T1-contrast, T2-weighted).

| No | ID | Core | | | Oedema | | | Whole | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Precision | Sensitivity | BER | Precision | Sensitivity | Precision | Sensitivity | BER | BER |
| **1** | 37227 | 91.20 | 86.36 | 0.07 | 82.64 | 74.84 | 0.13 | 93.45 | 85.02 | 0.08 |
| **2** | 37182 | 82.61 | 70.65 | 0.15 | 73.66 | 62.84 | 0.20 | 79.69 | 63.21 | 0.19 |
| **3** | 37197 | 72.66 | 68.53 | 0.16 | 77.29 | 87.34 | 0.07 | 81.36 | 88.89 | 0.06 |
| **4** | 37230 | 82.78 | 77.62 | 0.11 | 81.64 | 86.74 | 0.07 | 86.62 | 90.97 | 0.05 |
| **5** | 37253 | 96.00 | 67.33 | 0.16 | 90.78 | 84.29 | 0.08 | 93.13 | 92.20 | 0.04 |
| **6** | 37243 | 71.43 | 61.11 | 0.19 | 74.26 | 76.51 | 0.14 | 80.80 | 80.34 | 0.11 |
| **7** | 37256 | 61.54 | 72.73 | 0.14 | 91.95 | 85.10 | 0.08 | 93.06 | 89.59 | 0.06 |
| **8** | 37302 | 71.35 | 66.67 | 0.17 | 95.08 | 79.27 | 0.11 | 95.38 | 79.77 | 0.10 |
| **9** | 37394 | 63.78 | 59.86 | 0.20 | 85.77 | 91.59 | 0.05 | 92.65 | 91.51 | 0.05 |
| **10** | 37396 | 62.37 | 71.77 | 0.14 | 93.72 | 71.82 | 0.14 | 93.62 | 72.13 | 0.14 |
| **11** | 37218 | 54.35 | 63.76 | 0.18 | 95.12 | 86.09 | 0.07 | 95.19 | 86.65 | 0.07 |
| **Mean** | **All** | **73.64** | **69.67** | **0.15** | **85.63** | **80.59** | **0.10** | **89.54** | **83.66** | **0.09** |
| **STD** | **All** | **13.14** | **7.59** | **0.04** | **8.24** | **8.44** | **0.04** | **6.18** | **9.16** | **0.05** |

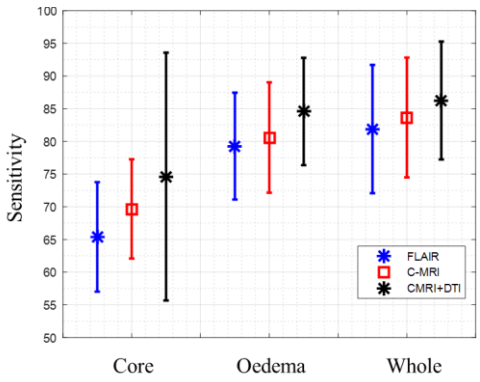
Table 5-5 Classification results (average values over all the LOO-CV) for superpixels using MRI conventional protocols plus DTI (FLAIR, T1-contrast, T2-weighted, *p* and *q*)

| No | ID | Core | | | Oedema | | | Whole | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Precision | Sensitivity | BER | Precision | Sensitivity | Precision | Sensitivity | BER | BER |
| **1** | 37227 | 97.01 | 91.55 | 0.04 | 86.00 | 81.13 | 0.10 | 95.21 | 89.83 | 0.05 |
| **2** | 37182 | 87.50 | 21.21 | 0.39 | 77.46 | 66.23 | 0.18 | 83.88 | 65.76 | 0.18 |
| **3** | 37197 | 81.23 | 75.48 | 0.12 | 78.47 | 88.49 | 0.07 | 79.08 | 86.87 | 0.07 |
| **4** | 37230 | 98.78 | 88.64 | 0.06 | 89.16 | 91.84 | 0.05 | 93.55 | 95.60 | 0.02 |
| **5** | 37253 | 100.00 | 86.66 | 0.07 | 91.38 | 86.98 | 0.07 | 94.48 | 95.57 | 0.02 |
| **6** | 37243 | 92.58 | 83.33 | 0.08 | 80.37 | 88.21 | 0.08 | 90.93 | 89.52 | 0.07 |
| **7** | 37256 | 70.00 | 77.78 | 0.11 | 91.45 | 85.33 | 0.08 | 91.33 | 87.74 | 0.06 |
| **8** | 37302 | 78.54 | 74.07 | 0.13 | 98.67 | 82.22 | 0.09 | 97.68 | 81.78 | 0.09 |
| **9** | 37394 | 75.57 | 72.73 | 0.14 | 87.90 | 96.49 | 0.02 | 95.60 | 94.67 | 0.03 |
| **10** | 37396 | 70.00 | 77.78 | 0.11 | 96.52 | 75.60 | 0.12 | 96.53 | 76.17 | 0.12 |
| **11** | 37218 | 66.67 | 71.54 | 0.14 | 96.41 | 87.77 | 0.06 | 96.19 | 85.28 | 0.08 |
| **Mean** | **All** | **83.44** | **74.62** | **0.13** | **88.53** | **84.57** | **0.08** | **92.22** | **86.25** | **0.07** |
| **STD** | **All** | **12.36** | **18.95** | **0.09** | **7.37** | **8.21** | **0.04** | **5.80** | **9.02** | **0.05** |

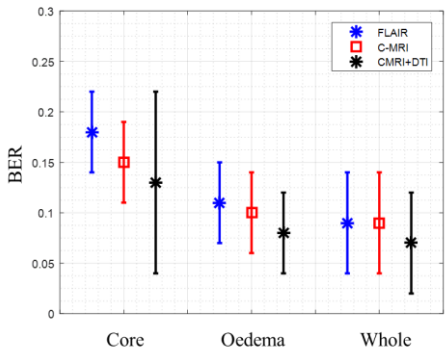
The average and standard deviation of the classification measure for the tumour tissues from all three experiments are plotted in Figure 5-18. It can be seen that the average precision and sensitivity increases by adding more protocols to the experiment. Also, BER decreases using C-MRI+DTI. However, for the whole tumour, BER is the same for FLAIR and C-MRI experiments. This shows that the FLAIR protocol is successful in segmenting the whole tumour (confirms the contribution of Chapter 4). However, for further segmentation of the tumour tissues subtypes, adding other protocols increases the accuracy. It also can be seen in Figure 5-18 that "tumour core" has a wide range of results (large black lines in Figure 5-18) for all the classification measures. The reason is that some tumour cases have a small size compared to the supervoxel size, which results in low classification measures.




a




b




c

Figure 5-18 Summary of the classification results for supervoxels from Single modality (FLAIR), multimodal C-MRI (FLAIR, T1-contrast and T2-weighted) and C-MRI+DTI (FLAIR, T1-contrast, T2-weighted, p- and *q*-maps) on clinical dataset. a) precision, b) sensitivity, c) BER.

### 5.3.4 Segmentation Results

Table 5-6 shows Dice score overlap measure between the annotated ground truth and the automated method using the three experiment sets on the clinical datasets. Results show significant improvement in the segmentation of the tumour core using the C-MRI+DTI approach with a DSC of 0.78 compared to C-MRI (DSC = 0.67) and the single FLAIR image (DSC = 0.54). This illustrates that adding the information from DTI sequences to the C-MRI increases the tumour segmentation accuracy for multimodal procedure.

Table 5-6 Dice score comparison for the segmentation of tumour core, oedema and complete tumour in clinical dataset using single protocol (FLAIR), C-MRI (FLAIR, T1-contrast, T2-weighted) and C-MRI+DTI (FLAIR, T1-contrast, T2-weighted, *p* and *q*-maps).

| **No** | **ID** | **FLAIR** | | | **FLAIR, T1-contrast, T2-weighted** | | | **FLAIR, T1-contrast, T2-weighted, p and q** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **Core** | **Oedema** | **Whole** | **Core** | **Oedema** | **Whole** | **Core** | **Oedema** | **Whole** |
| 1 | 37227 | 0.79 | 0.63 | 0.75 | 0.84 | 0.69 | 0.77 | 0.91 | 0.71 | 0.79 |
| 2 | 37182 | 0.55 | 0.66 | 0.70 | 0.60 | 0.69 | 0.72 | 0.84 | 0.73 | 0.77 |
| 3 | 37197 | 0.63 | 0.70 | 0.71 | 0.68 | 0.70 | 0.74 | 0.76 | 0.71 | 0.73 |
| 4 | 37230 | 0.65 | 0.73 | 0.78 | 0.76 | 0.77 | 0.82 | 0.85 | 0.86 | 0.91 |
| 5 | 37253 | 0.56 | 0.81 | 0.82 | 0.62 | 0.83 | 0.83 | 0.68 | 0.85 | 0.85 |
| 6 | 37243 | 0.65 | 0.72 | 0.75 | 0.72 | 0.73 | 0.76 | 0.83 | 0.81 | 0.85 |
| 7 | 37256 | 0.53 | 0.85 | 0.86 | 0.74 | 0.86 | 0.87 | 0.86 | 0.85 | 0.86 |
| 8 | 37302 | 0.42 | 0.85 | 0.85 | 0.58 | 0.86 | 0.86 | 0.62 | 0.87 | 0.87 |
| 9 | 37394 | 0.34 | 0.82 | 0.83 | 0.59 | 0.83 | 0.85 | 0.70 | 0.89 | 0.91 |
| 10 | 37396 | 0.41 | 0.86 | 0.86 | 0.68 | 0.85 | 0.86 | 0.83 | 0.86 | 0.88 |
| 11 | 37218 | 0.34 | 0.83 | 0.84 | 0.52 | 0.85 | 0.87 | 0.67 | 0.86 | 0.87 |
| **Mean** | **All** | **0.54** | **0.77** | **0.79** | **0.67** | **0.79** | **0.81** | **0.78** | **0.82** | **0.84** |
| **STD** | **All** | **0.14** | **0.08** | **0.06** | **0.10** | **0.07** | **0.06** | **0.09** | **0.07** | **0.06** |

Figure 5-19 shows examples of segmentation of the complete tumour using C-MRI and C-MRI+DTI for three different cases with grade IV tumours. It is noted that, there are some false positive regions (FPs) in the segmented masks shown in Figure 5-19 (c2 and c3) by using C-MRI protocols, this is due to the wrongly classified supervoxels, while adding DTI protocols reduces some of the FPs, leading to more accurate segmentation.

Figure 5-21 shows examples of segmentation of tumour core using C-MRI and C-MRI+DTI for the same three cases as in Figure 5-19. It can be seen that using DTI protocols with conventional MRI presents more accurate segmentation results. In Figure 5-21 (c1 and c3), there are areas of tumour core which are missed by using C-MRI protocols, while those tumour

areas can be detected by adding DTI protocol, as shown in Figure 5-21 (d1 and d3). This demonstrates an improvement of the segmentation accuracy by combining both conventional C-MRI and DTI.

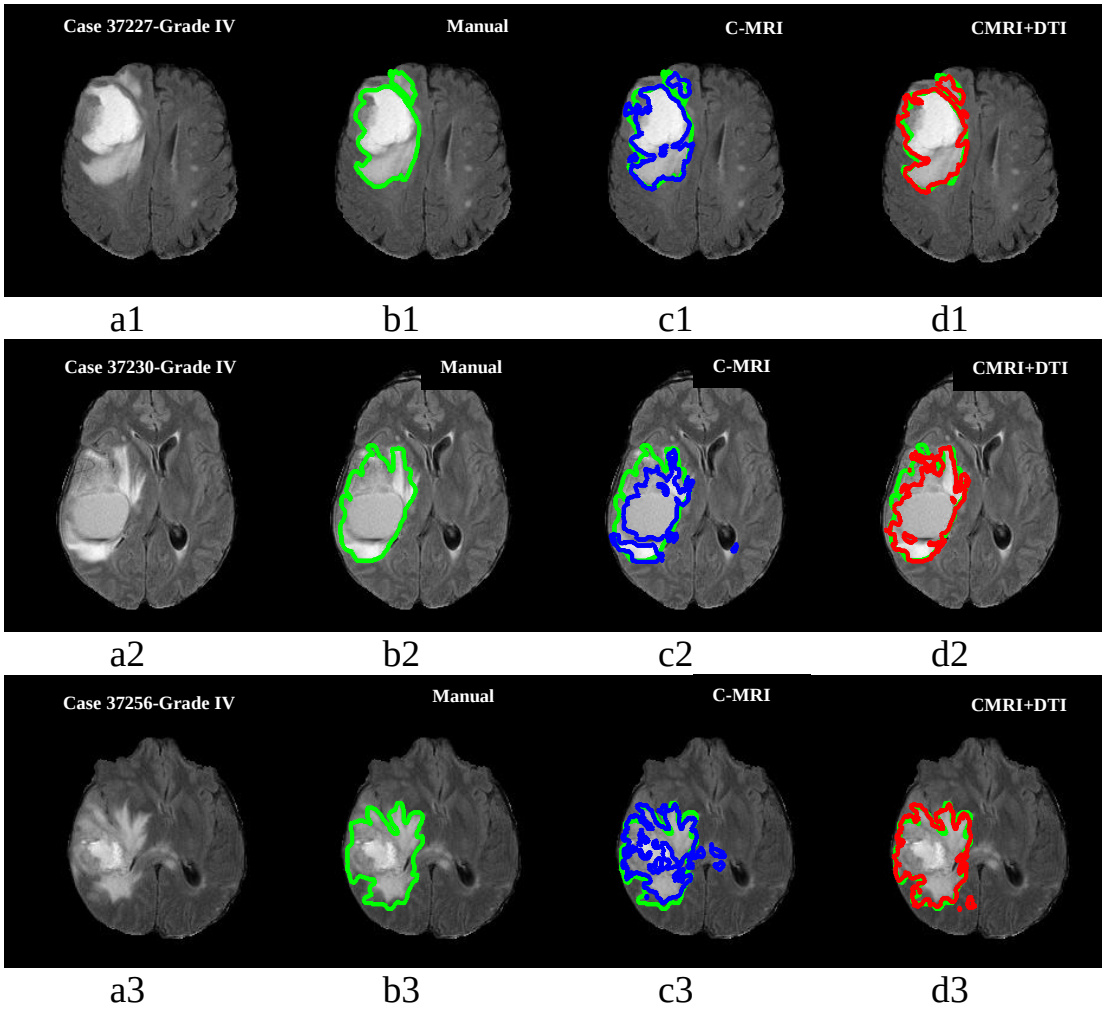



Figure 5-19 Comparison example of segmentation of complete tumours using C-MRI and C-MRI+DTI for three different cases with grade IV tumours. a1-a3) FLAIR image (a1: case 37227, a2: case 37230, and a3: case 37256), b1-b3) manual segmentation c1-c2) segmentation using C-MRI d1-d3) segmentation using C-MRI+DTI.

Figure 5-20 shows the 3D graphical view of the complete tumour segmentation volumes from the corresponding cases in Figure 5-19. The segmentation surfaces using C-MRI (blue) and C-MRI+DTI (red) are separately overlaid on the ground truth (green). As it can be seen, C-MRI segmentation results in false positive volumes for all cases, which are eliminated by adding DTI. Under-segmentation can be seen in Figure 5-20 by comparing the green surface areas. The green surface shows that the automatic segmentation surface is inside the ground truth. Although it is difficult to see the content of under-segmentation, the slices in Figure 5-19

provide an estimation for it. The blue and red surfaces show exact match of the segmentation surfaces in addition to over-segmentation. As it can be seen Figure 5-20, under-segmentation of C-MRI method for case numbers 37227 and 37230 is clearly more than the C-MRI+DTI approach.

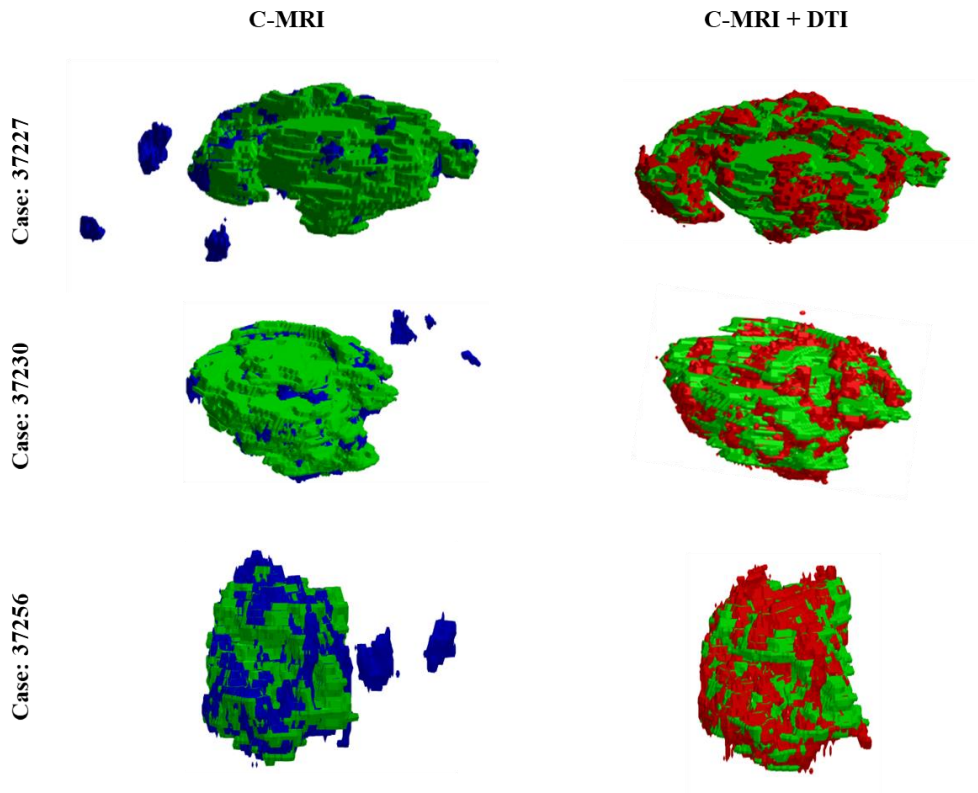



Figure 5-20 The 3D graphical representation of the complete tumour surfaces from the correponding cases in Figure 5-19. The segmentation surfaces using C-MRI (blue) and C-MRI+DTI (red) are overlaid on the ground truth (green).

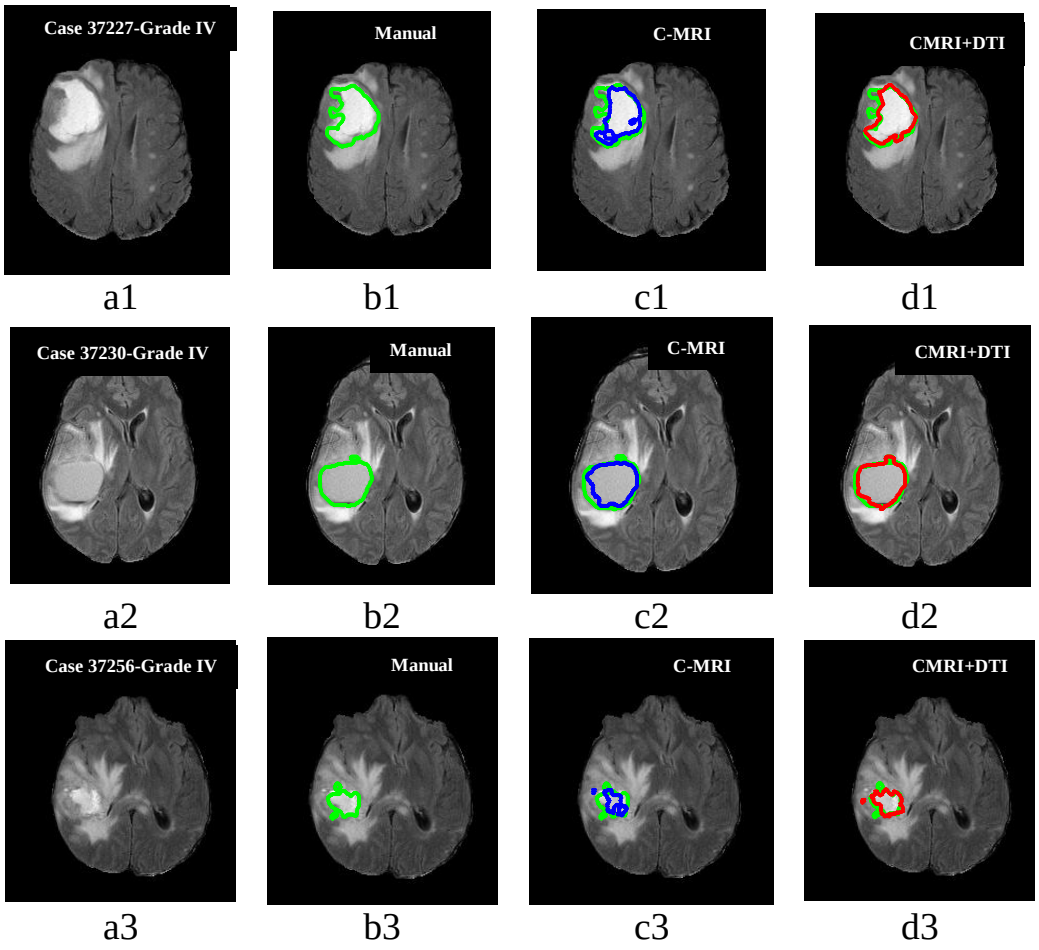



Figure 5-21 Segmentation results for the core of tumours using C-MRI and C-MRI+DTI for three different cases with grade IV tumours. a1-a3) FLAIR image (a1: case 37227, a2: case 37230, and a3: case 37256), b1-b3) manual segmentation c1-c2) segmentation using C-MRI d1-d3) segmentation using C-MRI+DTI.

Figure 5-22 shows the 3D graphical view of the tumour core segmentation volumes from the corresponding cases in Figure 5-21. The segmentation surfaces using C-MRI (blue) and C-MRI+DTI (red) are separately overlaid on the ground truth (green). As it can be seen, both methods did not produce any isolated false positive regions. Therefore, both C-MRI and DTI protocols provide the required features to separate the tumour core from normal brain tissues. However, adding DTI protocols provide a closer match of the surfaces of segmented volumes and ground truth. The surfaces of C-MRI + DRT segmentation (red) in Figure 5-22 cover more area compared to the C-MRI only (blue). The large area of ground truth surface (green) in the C-MRI results shows the method results in under-segmentation.

The case 37256 in Figure 5-22 and Figure 5-21 is a sample of small tumour tissue subtype. As it can be seen in Figure 5-22, a small part of the tumour core was not segmented by any of the methods. This is a major failure of supervoxel-based methods. The small sized tissues are not

detectable if their size is smaller than the supervoxels or if they are included partially in a few number of supervoxels.

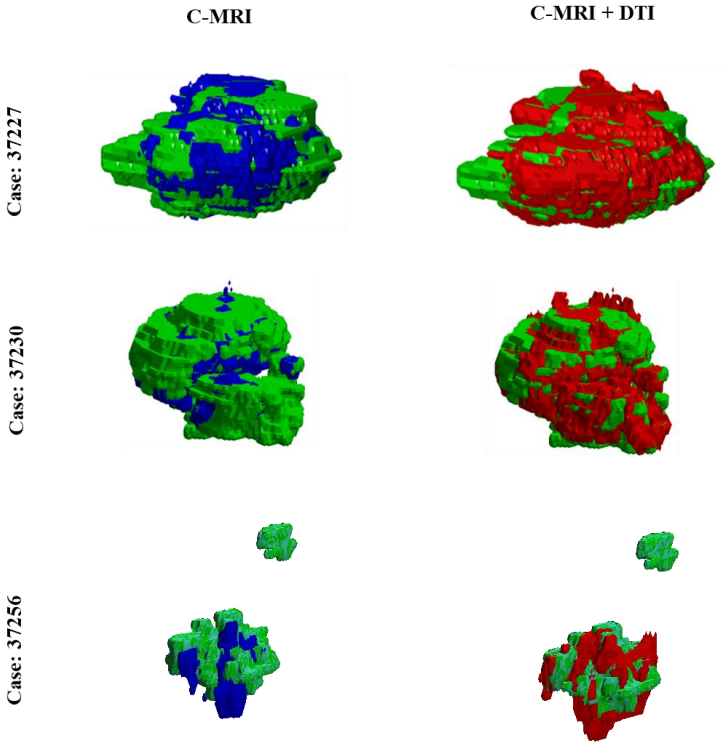



Figure 5-22 The 3D graphical representation of the tumour core surfaces from the correponding cases in Figure 5-21. The segmentation surfaces using C-MRI (blue) and C-MRI+DTI (red) are overlaid on ground truth (green).

### 5.3.5 Evaluation on Public BRATS 2013 Dataset

To evaluate the robustness of the proposed multimodal segmentation method, it was also applied to the BRATS 2013 training dataset. The details of the BRATS dataset were explained in Chapter 2 (Section 2.6.2). This dataset includes conventional MRI protocols, i.e. FLAIR, T1-weighted, T2-weighted and T1-contrast. Since no DTI protocols are available, only the multimodal aspect of the proposed SV_RF method is evaluated by comparing the segmentation using C-MRI (FLAIR, T1-weighted, T2-weighted and T1-contrast) with that using the single imaging modality (FLAIR only).

To investigate the robustness of parameter tuning, the parameters used for feature extraction from the BRATS dataset are selected similar to those which were used for the clinical datasets (Section 5.3.2). In the case of supervoxel segmentation, the only parameter that is different from analysis of the clinical dataset is the initial superpixel size, since the two datasets have different voxel dimensions. The voxel dimension for all BRATS data is 1 mm × 1 mm × 1

mm, so the initial subvolumes are cubes with the same dimensions. The supervoxel size is selected 5 mm × 5 mm × 5 mm for segmenting both oedema and tumour core, considering small tumours in some images. Table 5-7 presents the average evaluation results using SV_RF for supervoxel classification of tumour core and oedema against the rest of tissues separately, and classification of complete tumour against the healthy the tissue using single modality of FLAIR and multimodal approach on C-MRI including FLAIR, T1-weighted, T1- contrast and T2-weighted imaging.

Table 5-7 Classification results for supervoxels from FLAIR protocols of BRATS 2013 dataset.

| ID | Core | | | Oedema | | | Whole | | |
|---|---|---|---|---|---|---|---|---|---|
| | Precision | Sensitivity | BER | Precision | Sensitivity | BER | Precision | Sensitivity | BER |
| LG-01 | 94.12 | 78.11 | 0.11 | 99.61 | 87.59 | 0.06 | 99.31 | 85.67 | 0.07 |
| LG-02 | 96.66 | 92.76 | 0.04 | 96.42 | 93.74 | 0.03 | 99.60 | 97.31 | 0.01 |
| LG-04 | 99.51 | 90.63 | 0.05 | 60.98 | 75.76 | 0.12 | 99.55 | 92.93 | 0.04 |
| LG-06 | 98.37 | 90.95 | 0.05 | 78.79 | 86.67 | 0.07 | 98.62 | 93.45 | 0.03 |
| LG-08 | 90.00 | 86.54 | 0.07 | 96.92 | 96.18 | 0.02 | 98.89 | 97.27 | 0.01 |
| LG-11 | 94.23 | 92.39 | 0.04 | 98.98 | 90.68 | 0.05 | 99.47 | 91.30 | 0.04 |
| LG-12 | 94.80 | 91.95 | 0.04 | 96.28 | 95.39 | 0.02 | 99.44 | 97.27 | 0.01 |
| LG-13 | 96.33 | 80.15 | 0.10 | 98.51 | 81.89 | 0.09 | 98.87 | 82.21 | 0.09 |
| LG-14 | 88.64 | 79.59 | 0.10 | 93.44 | 89.06 | 0.05 | 92.38 | 85.84 | 0.07 |
| LG-15 | 87.80 | 87.80 | 0.06 | 97.71 | 85.91 | 0.07 | 96.71 | 89.18 | 0.05 |
| HG-01 | 92.55 | 95.60 | 0.02 | 95.72 | 92.01 | 0.04 | 98.87 | 98.25 | 0.01 |
| HG-02 | 97.15 | 95.88 | 0.02 | 89.06 | 95.00 | 0.03 | 98.87 | 97.77 | 0.01 |
| HG-03 | 99.25 | 95.38 | 0.02 | 84.29 | 95.68 | 0.02 | 99.76 | 98.37 | 0.01 |
| HG-04 | 95.06 | 88.51 | 0.06 | 96.63 | 95.95 | 0.02 | 99.75 | 97.10 | 0.01 |
| HG-05 | 96.12 | 92.96 | 0.04 | 98.89 | 82.66 | 0.09 | 99.79 | 88.62 | 0.06 |
| HG-06 | 94.26 | 92.80 | 0.04 | 98.49 | 93.76 | 0.03 | 99.93 | 95.69 | 0.02 |
| HG-07 | 95.76 | 96.17 | 0.02 | 96.18 | 82.97 | 0.09 | 98.73 | 93.05 | 0.03 |
| HG-08 | 98.83 | 90.58 | 0.05 | 93.52 | 92.87 | 0.04 | 99.69 | 93.79 | 0.03 |
| HG-09 | 97.55 | 96.05 | 0.02 | 96.00 | 87.65 | 0.06 | 99.93 | 95.14 | 0.02 |
| HG-10 | 91.00 | 88.35 | 0.06 | 93.55 | 85.29 | 0.07 | 93.89 | 89.78 | 0.05 |
| HG-11 | 96.70 | 94.47 | 0.03 | 97.24 | 90.31 | 0.05 | 99.69 | 95.71 | 0.02 |
| HG-12 | 90.79 | 90.79 | 0.05 | 97.87 | 77.97 | 0.11 | 95.93 | 87.41 | 0.06 |
| HG-13 | 81.13 | 93.48 | 0.03 | 96.77 | 85.71 | 0.07 | 94.05 | 97.53 | 0.01 |
| HG-14 | 77.48 | 90.70 | 0.05 | 99.63 | 87.04 | 0.06 | 99.79 | 90.78 | 0.05 |
| HG-22 | 98.44 | 96.33 | 0.02 | 96.42 | 90.65 | 0.05 | 99.79 | 95.62 | 0.02 |
| HG-24 | 96.63 | 95.82 | 0.02 | 94.76 | 78.70 | 0.11 | 98.17 | 91.17 | 0.04 |
| HG-25 | 94.07 | 92.50 | 0.04 | 93.75 | 84.68 | 0.08 | 94.83 | 90.66 | 0.05 |
| HG-26 | 92.04 | 86.81 | 0.07 | 97.71 | 80.33 | 0.10 | 95.70 | 84.86 | 0.08 |
| HG-27 | 98.00 | 84.19 | 0.08 | 88.10 | 79.74 | 0.10 | 97.80 | 85.02 | 0.08 |
| HG-22 | 91.43 | 92.44 | 0.04 | 98.01 | 84.19 | 0.08 | 99.77 | 89.99 | 0.05 |
| **Mean** | **93.82** | **90.69** | **0.05** | **94.01** | **87.53** | **0.06** | **98.25** | **92.29** | **0.04** |
| **STD** | **5.08** | **4.99** | **0.02** | **7.77** | **5.91** | **0.03** | **2.12** | **4.68** | **0.02** |

Table 5-8 shows the evaluation of SV_RF for supervoxel classification of tumour segmentation using multimodal approach on C-MRI protocols including FLAIR, T1-weighted, T1-contrast and T2-weighted.

Table 5-8 Classification results for superpixels from MRI Multi protocols (FLAIR, T1, T1-contrast and T2-weighted) of BRATS 2013 dataset.

| ID | Core | | | Oedema | | | Whole | | |
|---|---|---|---|---|---|---|---|---|---|
| | Precision | Sensitivity | BER | Precision | Sensitivity | BER | Precision | Sensitivity | BER |
| LG-01 | 92.68 | 87.36 | 0.06 | 100.00 | 93.15 | 0.03 | 99.68 | 92.86 | 0.04 |
| LG-02 | 98.36 | 95.50 | 0.02 | 96.48 | 97.94 | 0.01 | 99.89 | 98.77 | 0.01 |
| LG-04 | 99.75 | 95.50 | 0.02 | 91.49 | 97.73 | 0.01 | 100.00 | 94.95 | 0.03 |
| LG-06 | 100.00 | 95.59 | 0.02 | 96.43 | 87.10 | 0.06 | 100.00 | 94.89 | 0.03 |
| LG-08 | 94.44 | 87.93 | 0.06 | 96.30 | 98.48 | 0.01 | 97.88 | 97.37 | 0.01 |
| LG-11 | 98.01 | 93.46 | 0.03 | 98.85 | 99.42 | 0.00 | 99.78 | 98.90 | 0.01 |
| LG-12 | 98.29 | 93.67 | 0.03 | 98.65 | 99.10 | 0.00 | 100.00 | 97.58 | 0.01 |
| LG-13 | 99.06 | 89.77 | 0.05 | 99.45 | 89.66 | 0.05 | 100.00 | 88.06 | 0.06 |
| LG-14 | 95.45 | 88.25 | 0.06 | 97.18 | 100.00 | 0.00 | 98.26 | 92.62 | 0.04 |
| LG-15 | 97.65 | 96.51 | 0.02 | 99.36 | 96.30 | 0.02 | 100.00 | 97.58 | 0.01 |
| HG-01 | 98.64 | 97.06 | 0.01 | 99.33 | 98.67 | 0.01 | 99.26 | 98.18 | 0.01 |
| HG-02 | 99.75 | 98.26 | 0.01 | 98.61 | 100.00 | 0.00 | 99.57 | 98.52 | 0.01 |
| HG-03 | 99.72 | 98.54 | 0.01 | 99.46 | 98.40 | 0.01 | 99.84 | 98.67 | 0.01 |
| HG-04 | 99.61 | 95.46 | 0.02 | 98.69 | 99.83 | 0.00 | 99.88 | 98.07 | 0.01 |
| HG-05 | 100.00 | 94.32 | 0.03 | 99.34 | 87.57 | 0.06 | 99.80 | 89.20 | 0.05 |
| HG-06 | 97.99 | 96.57 | 0.02 | 98.59 | 98.48 | 0.01 | 99.63 | 98.88 | 0.01 |
| HG-07 | 99.12 | 96.15 | 0.02 | 98.92 | 97.87 | 0.01 | 99.27 | 97.16 | 0.01 |
| HG-08 | 99.57 | 96.45 | 0.02 | 97.77 | 97.33 | 0.01 | 100.00 | 97.73 | 0.01 |
| HG-09 | 99.29 | 98.39 | 0.01 | 98.37 | 93.77 | 0.03 | 99.87 | 97.58 | 0.01 |
| HG-10 | 96.88 | 89.42 | 0.05 | 100.00 | 100.00 | 0.00 | 97.71 | 92.09 | 0.04 |
| HG-11 | 98.93 | 98.03 | 0.01 | 99.72 | 97.80 | 0.01 | 99.90 | 98.63 | 0.01 |
| HG-12 | 94.11 | 94.59 | 0.03 | 98.08 | 86.44 | 0.07 | 98.44 | 94.74 | 0.03 |
| HG-13 | 100.00 | 97.92 | 0.01 | 97.78 | 100.00 | 0.00 | 98.91 | 98.91 | 0.01 |
| HG-14 | 96.27 | 94.16 | 0.03 | 100.00 | 97.37 | 0.01 | 99.62 | 97.05 | 0.01 |
| HG-22 | 99.67 | 98.37 | 0.01 | 99.28 | 97.53 | 0.01 | 100.00 | 98.52 | 0.01 |
| HG-24 | 98.84 | 97.15 | 0.01 | 99.17 | 99.17 | 0.00 | 99.32 | 98.31 | 0.01 |
| HG-25 | 97.40 | 94.01 | 0.03 | 98.82 | 95.80 | 0.02 | 98.88 | 94.27 | 0.03 |
| HG-26 | 97.59 | 95.81 | 0.02 | 98.70 | 92.14 | 0.04 | 98.90 | 94.83 | 0.03 |
| HG-27 | 99.33 | 92.65 | 0.04 | 95.71 | 85.11 | 0.07 | 99.49 | 91.74 | 0.04 |
| HG-22 | 99.18 | 95.58 | 0.02 | 98.83 | 94.61 | 0.03 | 99.89 | 95.96 | 0.02 |
| **Mean** | **98.19** | **94.75** | **0.03** | **98.31** | **95.89** | **0.02** | **99.46** | **96.09** | **0.02** |
| **STD** | **1.90** | **3.24** | **0.02** | **1.72** | **4.49** | **0.02** | **0.66** | **3.00** | **0.01** |

Table 5-7 and Table 5-8 show that the classification performances for different tumour parts (e.g. core, oedema, complete tumour) using multimodal C-MRI have been significantly improved compared to that using single FLAIR protocol. For tumour core, the overall supervoxel classification results using FLAIR protocol are precision of 93.82%, sensitivity of 90.69% and BER of 5%, whilst, of 98.19%, 94.75% and 3% for multi-protocol, respectively. For oedema, the overall classification results are precision of 94.01%, sensitivity of 87.53% and BER of 6% for single-protocol, and of 98.31%, 95.89% and 2% for multi-protocol, respectively. The overall classification results for the complete tumour are precision of 98.25%, sensitivity of 92.29% and BER of 4% for single-protocol, and of 99.46%, 96.09% and 2% for multi-protocol, respectively.

Table 5-9 shows the DSC overlap measures between the ground truth and segmented tumours using both single and multi-protocols, on the BRATS dataset. This demonstrates that using multi-protocol approach presents better overlap measures for tumour core, oedema, and

complete tumour, compared to that using FLAIR only, with mean of overlap for core of 0.8, oedema of 0.89, complete tumour of 0.89 using multiprotocol, against 0.79, 0.85, and 0.8, respectively using FLAIR only.

Table 5-9 Comparison results for DSC between manual annotation and the automated segmentation using a single protocol (FLAIR) and multi-protocol (FLAIR, T1-weighted, T1-contrast and T2-weighted) of BRATS 2013.

| Case No | Grade/ID | Single-Protocol | | | Multi-Protocol | | |
|---|---|---|---|---|---|---|---|
| | | Core | Oedema | Whole | Core | Oedema | Whole |
| 1 | LG-01 | 0.42 | 0.83 | 0.79 | 0.76 | 0.86 | 0.83 |
| 2 | LG-02 | 0.65 | 0.84 | 0.91 | 0.86 | 0.87 | 0.92 |
| 3 | LG-04 | 0.83 | 0.52 | 0.88 | 0.92 | 0.78 | 0.91 |
| 4 | LG-06 | 0.78 | 0.61 | 0.86 | 0.85 | 0.78 | 0.89 |
| 5 | LG-08 | 0.59 | 0.87 | 0.88 | 0.68 | 0.91 | 0.90 |
| 6 | LG-11 | 0.55 | 0.84 | 0.84 | 0.71 | 0.94 | 0.92 |
| 7 | LG-12 | 0.69 | 0.84 | 0.89 | 0.85 | 0.89 | 0.89 |
| 8 | LG-13 | 0.54 | 0.79 | 0.78 | 0.76 | 0.85 | 0.83 |
| 9 | LG-14 | 0.63 | 0.73 | 0.69 | 0.86 | 0.86 | 0.80 |
| 10 | LG-15 | 0.67 | 0.79 | 0.78 | 0.82 | 0.89 | 0.88 |
| 11 | HG-01 | 0.55 | 0.87 | 0.90 | 0.55 | 0.95 | 0.92 |
| 12 | HG-02 | 0.81 | 0.71 | 0.89 | 0.94 | 0.85 | 0.92 |
| 13 | HG-03 | 0.82 | 0.78 | 0.91 | 0.89 | 0.90 | 0.92 |
| 14 | HG-04 | 0.67 | 0.88 | 0.89 | 0.90 | 0.93 | 0.92 |
| 15 | HG-05 | 0.64 | 0.78 | 0.81 | 0.85 | 0.86 | 0.84 |
| 16 | HG-06 | 0.54 | 0.88 | 0.89 | 0.87 | 0.93 | 0.93 |
| 17 | HG-07 | 0.68 | 0.76 | 0.82 | 0.88 | 0.90 | 0.91 |
| 18 | HG-08 | 0.69 | 0.85 | 0.89 | 0.73 | 0.92 | 0.92 |
| 19 | HG-09 | 0.70 | 0.85 | 0.91 | 0.82 | 0.90 | 0.93 |
| 20 | HG-10 | 0.65 | 0.69 | 0.77 | 0.80 | 0.90 | 0.83 |
| 21 | HG-11 | 0.70 | 0.85 | 0.91 | 0.89 | 0.93 | 0.94 |
| 22 | HG-12 | 0.54 | 0.68 | 0.77 | 0.77 | 0.75 | 0.81 |
| 23 | HG-13 | 0.66 | 0.73 | 0.79 | 0.89 | 0.94 | 0.88 |
| 24 | HG-14 | 0.61 | 0.87 | 0.88 | 0.78 | 0.95 | 0.93 |
| 25 | HG-15 | 0.68 | 0.85 | 0.89 | 0.82 | 0.93 | 0.92 |
| 26 | HG-22 | 0.65 | 0.77 | 0.86 | 0.76 | 0.94 | 0.91 |
| 27 | HG-24 | 0.66 | 0.80 | 0.84 | 0.77 | 0.91 | 0.88 |
| 28 | HG-25 | 0.56 | 0.79 | 0.82 | 0.62 | 0.88 | 0.91 |
| 29 | HG-26 | 0.68 | 0.68 | 0.80 | 0.73 | 0.78 | 0.86 |
| 30 | HG-27 | 0.60 | 0.83 | 0.88 | 0.72 | 0.91 | 0.91 |
| **Mean** | **All** | **0.65** | **0.79** | **0.85** | **0.80** | **0.89** | **0.89** |
| **STD** | **All** | **0.09** | **0.09** | **0.06** | **0.09** | **0.05** | **0.04** |

Figure 5-23 shows the segmentation overlays of the complete tumour for BRATS 2013 dataset using single- and multi-protocols. Figure 5-25 shows the overlay of the tumour core on the same cases and slices as shown in Figure 5-23. It is noted that the segmentation results are presented for axial view and are overlaid on the FLAIR protocol.

Some irregular boundaries can be seen in Figure 5-23 (a2) and (c3). It should be noted that they are 2D presentation of a 3D volumetric segmentation. If a supervoxel is labelled as normal brain, but contains some voxels related to tumour, it will paper as a hole in one slice. Therefore, the irregular boundaries are the holes that are created as a consequence of misclassification of a supervoxel.

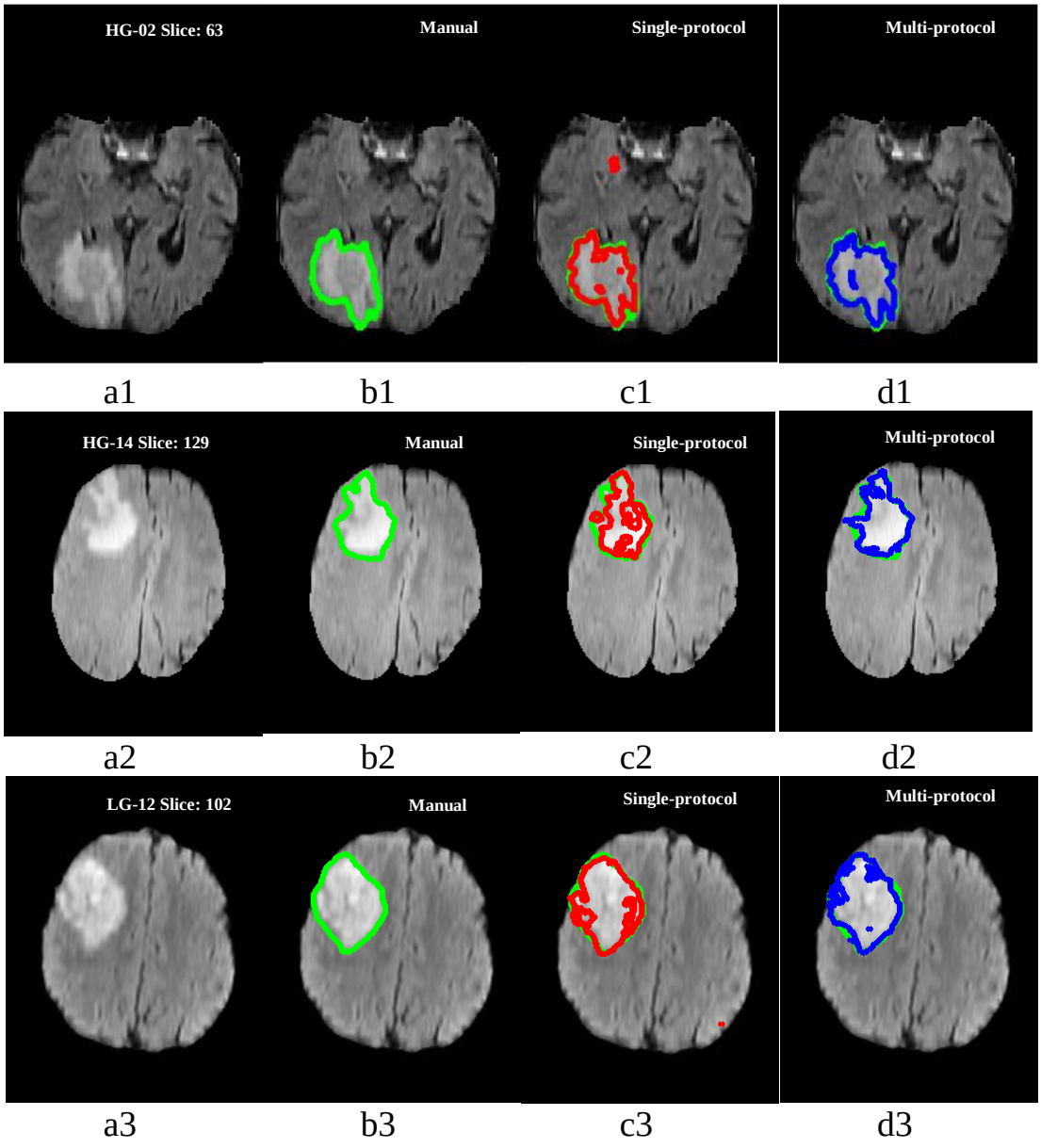



Figure 5-23 Segmentation results overlaid on the ground truth (complete tumour including oedema and core), using single (FLAIR) and multi-protocol (C-MRI including FLAIR, T1-weighted, T1-contrast and T2-weighted); a1)-a3) FLAIR image, b1)-b3) manual segmentation (green) c1)-c3) segmentation using FLAIR (red) d1)-d3) segmentation using conventional MRI (blue).

Figure 5-24 shows the 3D graphical view of the complete tumour segmentation volumes from the corresponding cases in Figure 5-23. The segmentation surfaces using single-protocol (red) and multi-protocol (blue) are separately overlaid on the ground truth (green). As it can be seen, the number of isolated false positive volumes decreases by adding more protocols to the segmentation procedure.

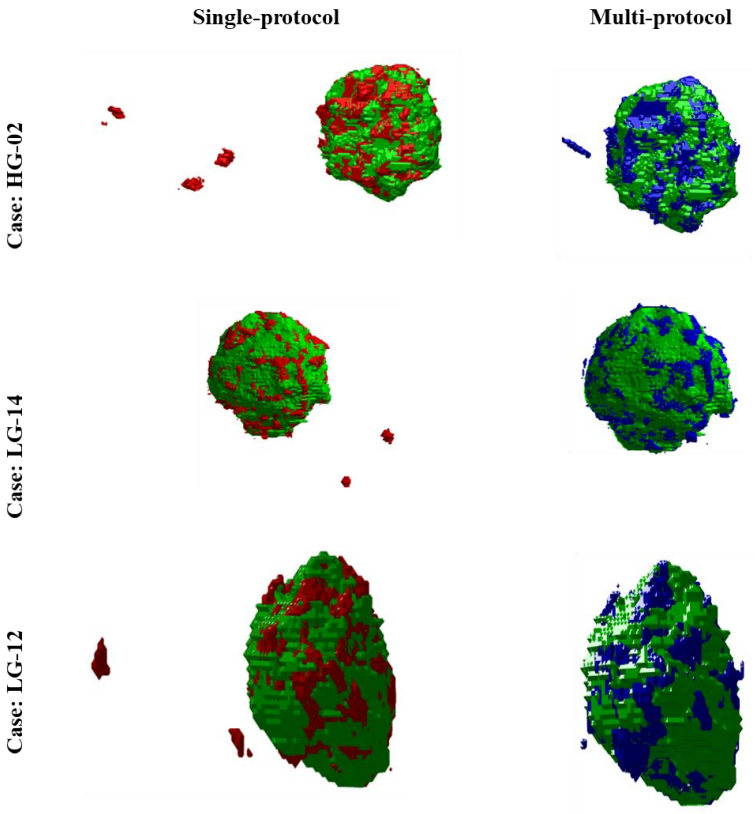



Figure 5-24 The 3D graphical representation of the complete tumour surfaces from the correponding cases in Figure 5-23. The segmentation surfaces using single-protocol (red) and multi-protocol (blue) are overlaid on ground truth (green).

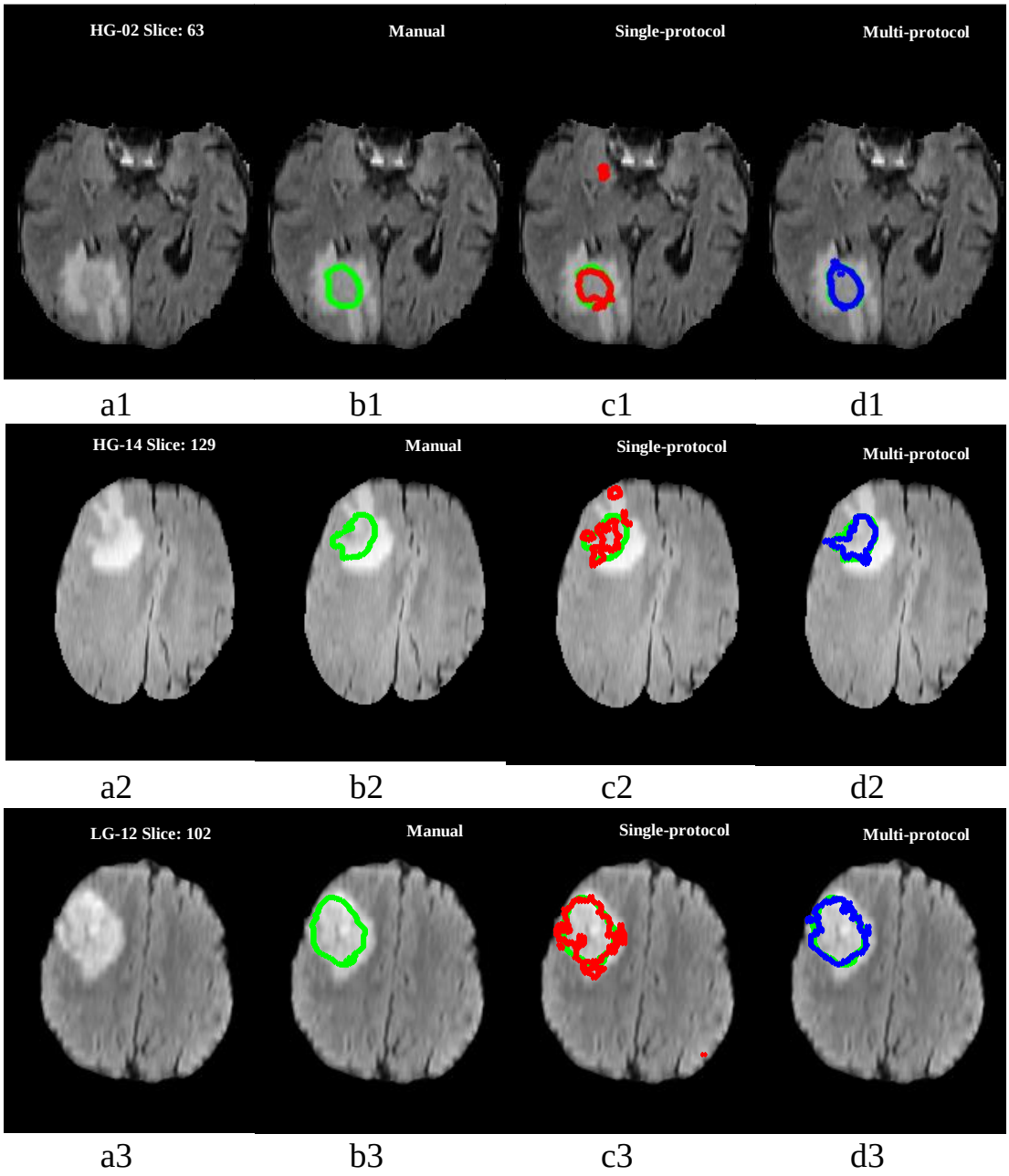



Figure 5-25 Segmentation results overlaid on the ground truth (tumour core), using single (FLAIR) and multi-protocol (C-MRI including FLAIR, T1, T1-contrast and T2-weighted) for the same three cases shown Figure 5-23. a1)-a3) FLAIR image, b1)-b3) manual segmentation (green) c1)-c3) segmentation using FLAIR (red) d1)-d3) segmentation using conventional MRI (blue)).

Figure 5-26 shows the 3D graphical view of the tumour core segmentation volumes from the corresponding cases in Figure 5-25. The segmentation surfaces using single-protocol (red) and multi-protocol (blue) are separately overlaid on the ground truth (green). As it can be seen, the number of isolated false positive volumes decreases by adding more protocols to the segmentation procedure. However, the under-segmentation comparison is difficult in the 3D

volumes. As can be seen in Figure 5-25, single-protocol approach results in under-segmentation only for the case HG-14.

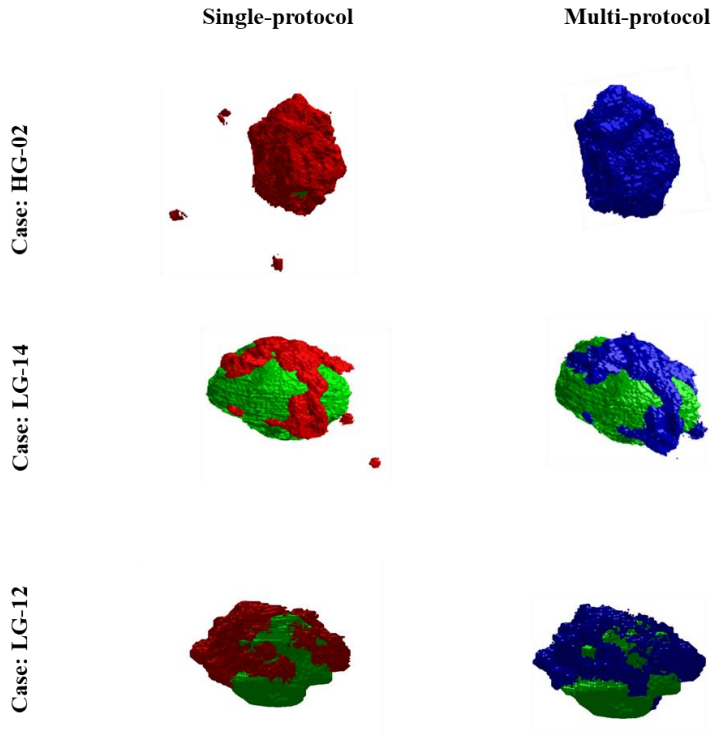
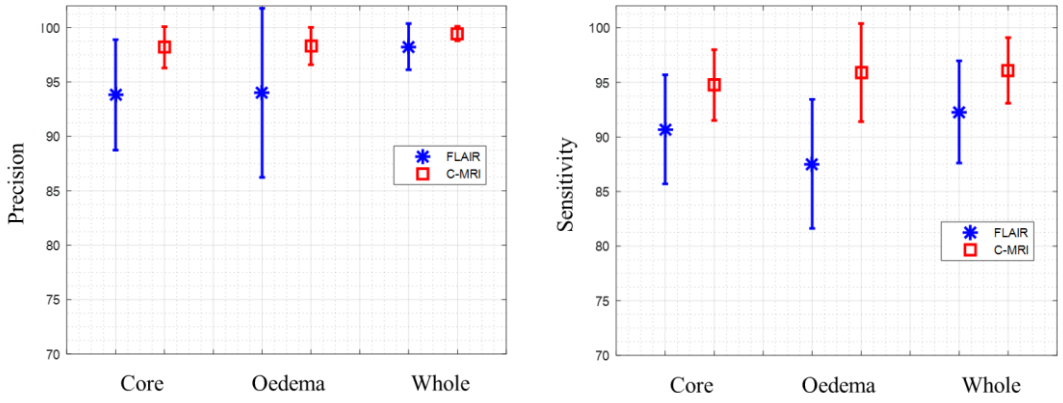
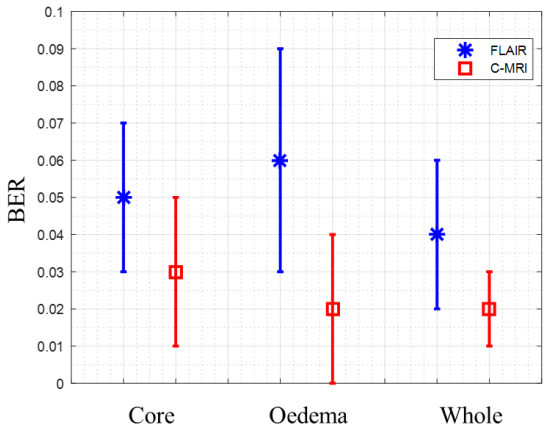



Figure 5-26 The 3D graphical representation of the tumour core surfaces from the correponding cases in Figure 5-25. The segmentation surfaces using single-protocol (red) and multi-protocol (blue) are overlaid on ground truth (green).

It can be seen from Figure 5-23 that using the multi-protocol procedure results in better and more accurate segmentation compared to single-protocol. Figure 5-25 shows using multi-protocol procedure has very accurate segmentation for the tumour core part. In Figure 5-25 (c1) some parts of normal brain are detected as tumour core and in Figure 5-25 (c2 and c3) some parts of oedema are wrongly classified as tumour core, using FLAIR only, which have been improved in Figure 5-23 (d1,d2, and d3) using C-MRI. The ratio of supervoxels for the whole tumour class versus the rest of classes is higher compared to other tissue classes. Therefore, the range of sensitivity and BER for the whole tumour is smaller compared to core and oedema.




a b




c

Figure 5-27 Average classification results for supervoxels from single modality (FLAIR) and multimodal C-MRI of the BRATS 2013 dataset. a) precision, b) sensitivity, c) BER.

### 5.3.6 Statistical Analysis

To investigate the differences in both DSC and classification measures of precision, sensitivity and BER between the single- and multimodal approaches, Wilcoxon signed-rank tests at a 95% confidence level was performed on both the clinical dataset and the BRATS 2013.

Table 5-10 shows Wilcoxon signed-ranks test statistical results for complete tumour segmentation for the DSC and classification measures using the different imaging protocols on the clinical dataset (N=11).

Table 5-10 Wilcoxon signed-ranks test statistical parameters results for the segmentation overlap measure of DSC and the classification measures using FLAIR only, C-MRI, and C-MRI+DTI) on the clinical dataset (11 subjects).

| **Complete 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 |

As can be noted in Table 5-10, there is a statistically significant improvement in DSC and in classification measures of precision, sensitivity, BER, when using the C-MRI+DTI multimodal data compared to C-MRI or FLAIR alone.

Table 5-11 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 should be noted that the dataset does not contain DTI images.

Table 5-11 Wilcoxon signed-ranks test statistical parameters results for the segmentation overlap measure of DSC and the classification measures using FLAIR only, and C-MRI on BRATS dataset (30 subjects).

| **Complete 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 |

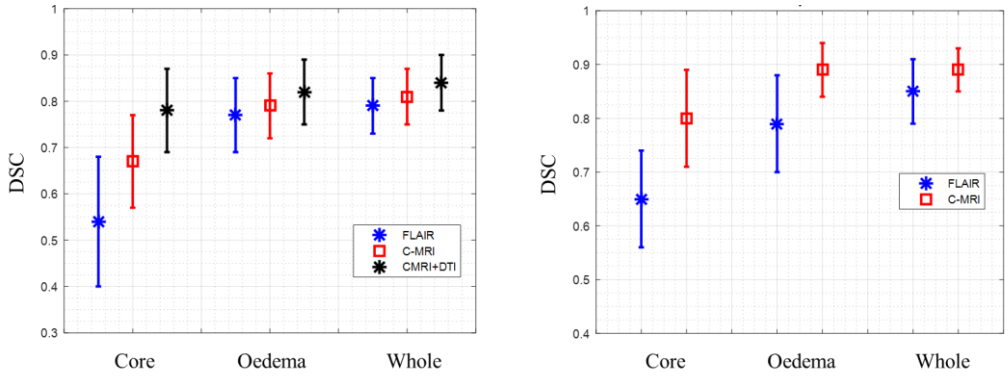
Finally, the results from the two different datasets were combined in a single group containing either FLAIR or C-MRI (N=41) and the corresponding Wilcoxon signed-ranks test statistics are presented in Table 5-12. This test was designed to assess the robustness of the method for application in a dataset with various acquisition parameters and image properties. The results indicate a statistically significant improvement in Dice scores and all classification measures when using the C-MRI protocol, instead of the FLAIR image alone.

Table 5-12 Wilcoxon signed-ranks test statistical parameters results for the segmentation overlap measure of DSC and the classification measures using FLAIR only, and C-MRI, on both the clinical and BRATS 2013 dataset (41 subjects).

| **Complete 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 |

## 5.4 Discussion

The calculation of distances for supervoxel segmentation were based on SLIC (Achanta *et al.*, 2012) which was originally developed for natural images using 2D regular arrays without considering pixel resolutions. Whilst, the 3D clinical dataset is anisotropic, with different voxel resolutions along each dimension. To address this problem, the distance formulations in the supervoxel calculation (Achanta *et al.*, 2012) were adopted to be applicable for MR data with different acquisition parameters as shown in Equation (5-5). The modified supervoxel formulation should be capable of being used for the data with different acquisition parameters mentioned in Chapter 2 (Section 2.6) and Section 5.3.1. Two different sets of data with different voxel dimensions and slice thickness were used to evaluate the supervoxel method. The 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. $8 \times 8 \times 3$). Whilst, the BRATS dataset has been interpolated to 1 $mm^3$ isotropic resolution, so initial supervoxels are defined to be cubic (e.g. $5 \times 5 \times 5$). The supervoxel segmentation boundary for the 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 the clinical data. The results in Table 5-6 and Table 5-9 confirm this, and show the overall segmentation of tumour for the clinical dataset has the average of 0.84 with standard deviation 0.06, whereas for the BRATS dataset they are 0.89 and 0.04, respectively. Figure 5-28 presents an overall comparison of the DSC overlap measure of each tumour part (i.e. whole, oedema and core) using different sets of modalities (FLAIR, C-MRI and C-MRI+DTI).




a b

Figure 5-28 Comparison results for Dice overlap ratio between manual annotation and the automated segmentation using: a) single modality (FLAIR), multimodal C-MRI, and C-MRI+DTI of the clinical dataset, b) single modality (FLAIR) and multimodal C-MRI of BRATS 2013 dataset.

Figure 5-29 compares the segmentation of the overall tumour from different experiments combining Figure 5-19 and Figure 5-21. Several supervoxels are wrongly classified, e.g. false positive regions (FPs), in the segmented masks when using FLAIR and C-MRI images (see Figure 5-29 (c2 and c3)) whereas adding the DTI image modalities reduces these FPs, leading to a more accurate segmentation. For example, in Figure 5-29 (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 Figure 5-29 (d1 and d3). This demonstrates an improvement in segmentation accuracy using both C-MRI and DTI.

To evaluate the robustness and generality of the proposed multimodal SV_RF method, it was applied to the BRATS 2013 multimodal dataset. However, this dataset does not contain DTI protocols *p*- and *q*-map. Therefore, the single modal (FLAIR) is compared against the multimodality C- 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.

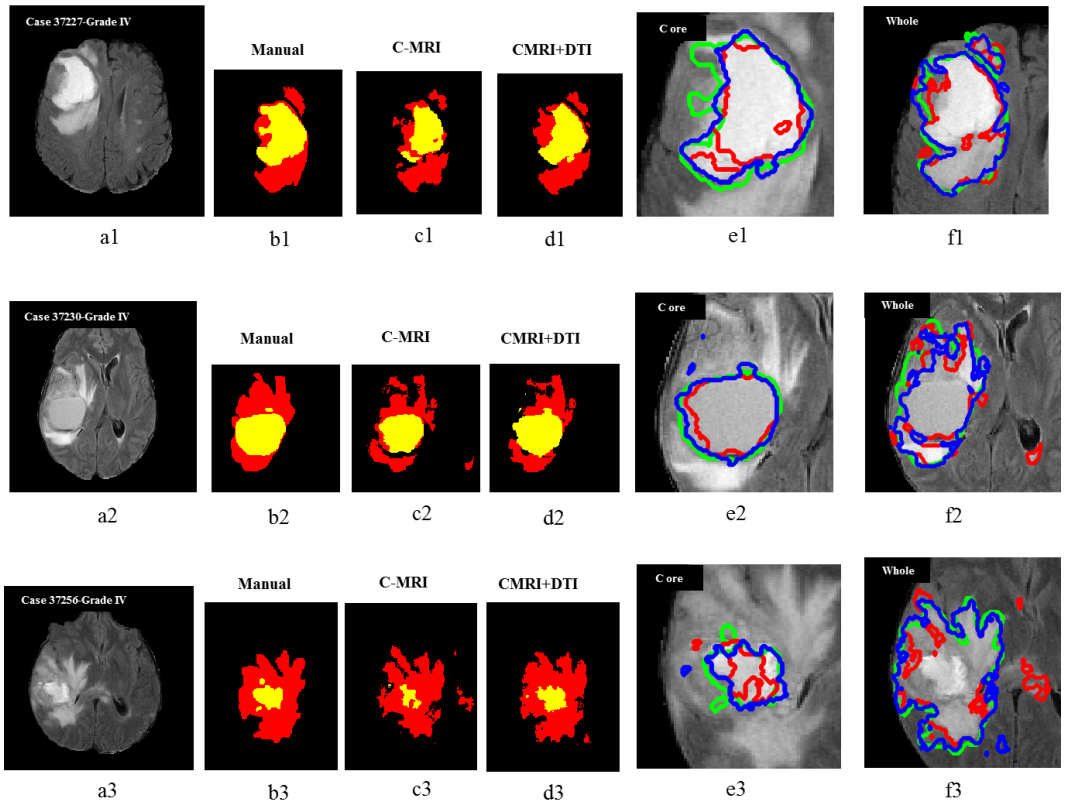



Figure 5-29 Overall comparison tumour segmentation. A) FLAIR image, B) manual segmentation of the core (yellow region) and oedema (red region) C) segmentation using conventional MRI, D) segmentation using C-MRI+DTI, E) comparison of both methods C-MRI (red), plus DTI (blue) and manual (green) segmentation for core (zoomed in), F) comparison of both methods C-MRI (red), C-MRI+DTI (blue) and manual (green) segmentation for oedema (zoomed in).

A zoomed-in image of the overlay of the tumour cores (shown in Figure 5-25) is depicted in Figure 5-30. 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-contrast. As can be seen in Figure 5-30, the information from protocol T1-contrast improves the segmentation of tumour core, as the tumour core has more clear boundaries in this protocol. The homogenous region in the FLAIR image (Figure 5-30 (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-weighted improves the segmentation accuracy (blue contour in Figure 5-30 (d)). The false positive region (shown in red in Figure 5-30 (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.

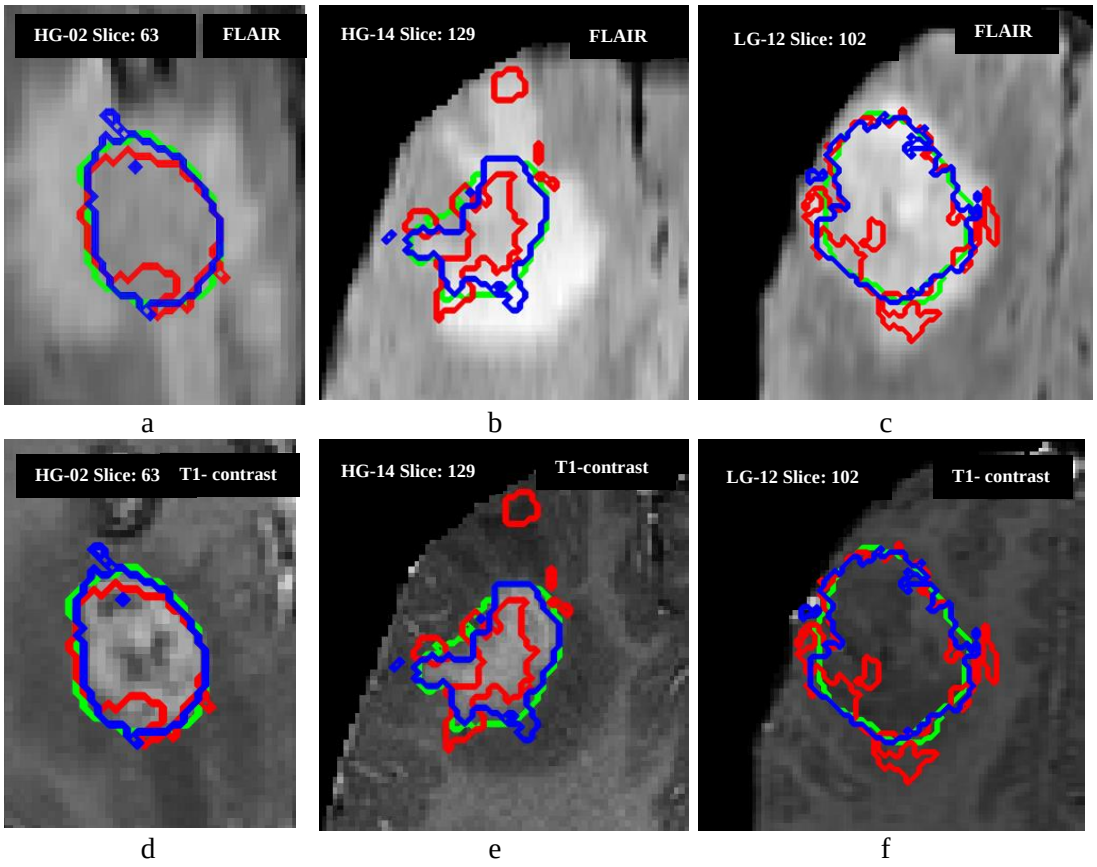


Figure 5-30 Comparison between single-modal and multimodal segmentation of the core. a-c) FLAIR, d-f) T1-contrast. Green: manual ground truth, red: single-modal, blue: multimodal.

Despite the advantages of supervoxel segmentation in providing homogenous patches of the volume, there are limitations which should be considered while using this technique. One limitation is the minimum size for supervoxels regarding its parameters and image characteristics. Therefore, 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 5-6 are relatively low, which is due to very small tumour cores for those data. Those cores only contain a limited number of supervoxels, unlike oedema which is usually large and encompasses many supervoxels.

The results of the multimodal SV_RF on the BRATS 2013 dataset and the best scores in 2012 and 2013 challenges from other groups (Menze *et al.*, 2015) are presented in Table 5-13. The method proposed by Tustison *et al.* (Tustison *et al.*, 2014) was the winner of the on-site BRATS 2013 challenge. Although the testing dataset for evaluation of multimodal SV_RF is different with their dataset, it provides a comparable scale to current experiments. To fairly evaluate the proposed SV_RF method, the results are also compared with the best scores that achieved from the same clinical training dataset from the other groups. Reza *et al.* (Reza and

Iftekharuddin, 2013) used the training clinical data to evaluate their method and obtained the best results for the same data as used for SV_RF method. They used different texture features and considered all the voxels in the image, which provided the best results amongst all the methods. Although their method provides slightly better results, it is time consuming since the computations were performed for all the voxels. Other methods which used RF classifier and were explained in Chapter 3 are also included in the Table 5-13, i.e. Festa *et al.* (Festa *et al.*, 2013), Geremia *et al.* (Geremia *et al.*, 2013), and Meier *et al.* (Meier *et al.*, 2014b). Bauer *et al.* used the most number of feature types (i.e. 44 feature types), among them four are statistical intensity-based. Festa *et al.* split the training set to half for normal brain and half for tumour and oedema to tackle the unbalanced problem. The proposed SV_RF used all the sample points which has the actual ratio of classes. The average of the top 10 best results which used the same training dataset of BRATS 2013 according to their website ("BRATS :: The Virtual Skeleton Database Project," n.d.) is also presented in Table 5-13. It should be noted that, as the VSD evaluation protocol (described in Chapter 2, Section 2.7.1) does not provide the evaluation metrics for oedema separately, they are not mentioned in the comparison table. For the same reason, the results of oedema segmentation using the proposed SV_RF method are also excluded in Table 5-13.

Overall, the comparison results in Table 5-13 demonstrate a good performance of the proposed multimodal SV_RF method for segmentation both of tumour core and complete tumour, with Dice scores of 0.80 and 0.89, respectively.

Table 5-13 Dice score comparison of the proposed multimodal SV_RF with other methods which used BRATS 2013 training dataset (MICCAI 2012 and 2013).

| **Work** | **Method** | **Whole** | **Tumour Core** |
|---|---|---|---|
| Tutison (Tustison *et al.*, 2014) | RF and MRF | 0.87 | 0.78 |
| Reza (Reza and Iftekharuddin, 2013) | RF and texture features | 0.92 | 0.91 |
| Festa (Festa *et al.*, 2013) | Local context features and RF | 0.79 | 0.62 |
| Bauer (Bauer *et al.*, 2012) | RF and CRF | 0.68 | 0.48 |
| Geremia (Geremia *et al.*, 2013) | Spatially adaptive RF | 0.83 | 0.62 |
| Meier (Meier *et al.*, 2014b) | Appearance and context features with RF and CRF | 0.83 | 0.66 |
| Top 10 average | | 0.87 | 0.78 |
| The proposed multimodal SV_RF | RF and multimodal supervoxel | 0.89 | 0.80 |

## 5.5 Limitations

The multimodal supervoxel segmentation method proposed in this chapter considers the imaging protocols with the same weight. Therefore, all the protocols will have the same impact on calculating the SV boundaries. For annotation of the tumour tissue subtypes some protocols provide more accurate boundaries, compared to the others. For example, FLAIR provides clearer boundaries for oedema, while T1-contrast provide a better representation for the tumour core. The proposed method does not consider this impact. It might be more accurate to consider the impact of different protocols by introducing a weight factor in the supervoxel formulations.

Another limitation of supervoxel-based methods is the segmentation of small targets. For instance, if a tumour core volume is less than the average supervoxel size, the proposed method might fail in assigning the correct class to that supervoxel. On the other hand, if a tumour volume is partially split between several supervoxels, again the proposed method might fail to segment it correctly. An example of this scenario is the case 37256 from the clinical dataset that contains two separate volumes of tumour core. As it can be seen in Figure 5-22, both supervoxel-based methods failed to segment the small part of the tumour core.

The supervoxel compactness factor parameter, *m*, was chosen by visually inspecting the image slices. Supervoxel is a 3D method that requires volumetric evaluation for selecting the parameters. Therefore, using an automatic measurement of the distance between the ground truth surface and the SV boundaries in 3D is a more accurate approach. This can be a future direction for selecting a more accurate optimal value for *m*.

## 5.6 Conclusion

A supervised learning based method is proposed for segmentation of tumour in multimodal MRI brain tumour images. Supervoxels were calculated using information fusion from multimodal MRI images. A novel histogram of texton descriptors based on supervoxels, calculated using a set of 3D Gabor filters with different sizes and orientations was developed. A random forest 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. Also, 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. The proportion of the features selected from each protocol using the RF for the segmentation and classification of the tumour were computed. The features extracted from the DTI protocols were found to be included as 24% of the total features, which represents the further improvement of the segmentation and classification performance by combining the *p*- and *q*-map protocols with the C-MRI. In addition, the proposed supervoxel method has also been evaluated on the BRATS 2013 dataset which also presents accurate and robust results.

This work has been submitted to the journal of Computers in Biology and Medicine (CBM) and currently is in a revision stage.

Despite the promising performance of the SV_RF method, there is a limitation in supervoxel minimum size which makes the method difficult for segmentation of very small regions (e.g. tumour core). Also, hand designed features need a large amount of experiments and parameter optimisation to produce promising segmentations. The next chapter will investigate the machine learned features and application of them with hand designed features. The further segmentation of tumour core subtypes, i.e. necrosis, non-enhanced and enhancing core, will also be investigated at the voxel level.

# Chapter 6

# FCN-based Brain Tumour Tissue Segmentation

## 6.1 Introduction

Most of the classification-based brain tumour segmentation techniques used hand-designed features which are fed into a classifier such as RF, SVM, etc. (Gotz *et al.*, 2014; Pinto *et al.*, 2015). Among the conventional classifiers, RF presents the best segmentation results (Gotz *et al.*, 2014; Menze *et al.*, 2015). However, one limitation of these types of methods is that a large number of features are required in order to provide better description of the different types of classes (tissues) in the images. Therefore, it results in a high dimensional problem which makes the process more complicated and time consuming. Furthermore, many experiments and optimisation should be performed to obtain the optimal parameters for feature extraction and designing the optimal classifier.

To overcome this problem, another variant of discriminative approaches that use a CNN has attracted significant attention in recent years. CNNs are the state-of-the-art methods which can efficiently perform classification of images. Later, CNNs were adopted for image segmentation. Several CNN-based methods have been developed for medical image analysis, especially for segmentation of brain tumours in MRI (Havaei *et al.*, 2017; Pereira *et al.*, 2016). A major drawback of CNN-based methods is that sufficient local dependencies are not considered for classification of the pixels. Recently, a few approaches have been suggested to overcome this limitation and provide more dense predictions at the pixel level. A most recent method is based on FCN which was proposed by Long *et al.* (Long *et al.*, 2015). They suggested fusing information from the intermediate convolutional layers to the output, so that both semantic and high resolution location information are considered. This will produce dense classification that results in a partially end-to-end pixel-by-pixel segmentation. The output layer of this FCN provides feature maps with the same size of the input image can be considered as the machine-learned features generated from the local and global information. On the other hand, some hand-designed feature extraction methods consider the spatial features and local dependencies of the pixel classes. As discussed in Chapters 4 and 5, texton features are amongst them and provides strong descriptors of local dependencies in both the spatial and frequency domains. The main hypothesis of this chapter is that combination of those hand-designed features with the machine-learned features may provide more local information, and thus improve the semantic segmentation from the FCN.

As discussed in Chapter 3, the largest challenge in the field of brain tumour segmentation is to develop an efficient and automated procedure to accurately segment different tumour parts in a multiclass scheme. The RF classifier alongside with hand-designed features, especially texton, have shown the promise in segmentation of a tumour from the normal brain tissue (Chapter 4). The multimodal SV_RF has demonstrated good performance in solving the multiclass segmentation problem (e.g. segmentation of tumour core and oedema from normal brain (Chapter 5)). The motivation of this chapter is to tackle this problem by investigating the hybrid method for accurate segmentation of complex tumour structures. Several experiments were also designed to investigate the machine-learned and hand-designed features to understand the efficiency of the combination.

The VGG16 architecture was selected for implementation of the FCN since the methods that were based on VGG16 provided promising performance in many recent semantic segmentations of medical and nonmedical tasks. The score maps extracted from the last deconvolution layer of the FCN were used to localise the tumour area and were also used as feature extraction. Due to the coarse segmentations using FCN, texton based features were then used to take into account the pixel neighbourhood for more accurate classification. To evaluate the robustness of the method, it was applied on the publicly available BRATS 2013 and 2017 dataset. The segmentation results presented here were reported by the blind test which was provided by the VSD online system.

Yao *et al.* (Yao *et al.*, 2016) proposed a CNN method based on Gabor filters that feed the responses of Gabor filters as an input to the network. The original images were applied to a set of Gabor filters with different orientations. Then, the filter responses were merged to form an input array.

Luan *et al.* (Luan *et al.*, 2017) proposed a method by fusing the Gabor filters to the architecture of CNN. The aim of their study was to incorporate the capability of Gabor filers in dealing with spatial transformation into deep CNN architectures to increase their robustness to spatial changes, i.e. orientation and scale. Gabor filters were used as orientation and convolutional filters. For each orientation, a layer was considered in the architecture of the network. The weights of the Gabor convolutional layers were updated using back-propagation, similar to conventional CNNs.

The proposed method in this thesis applies k-means clustering to extract texton features from the input image, which are used to form a feature representation based on connectivity of the pixels.

The contributions of this chapter can be listed as follows:

- A novel framework is proposed to combine two conceptually similar but operationally different mechanisms (i.e. hand-designed and machine-learned) for feature representation. The principle of these two methods is to simulate a mechanism that is able to decompose high order visual features to elementary visual features.
- The connection between the traditional feature engineering and deep learning (FCN) is explored in a new perspective, where the hand-designed features are obtained from a shallow network (with designable filters) while the machine-learned features are extracted from a deep network (with trainable filters).
- A novel scheme is proposed to address the limitations of the FCN (e.g. local dependencies are not sufficiently taken into account) in combination with texton features. The local dependencies are implicitly encoded into the feature representation.

## 6.2 Methods

The deep CNN was explained in Chapter 3. The following sections will explain the concepts and architecture of FCN, which is a variation of deep networks, in detail.

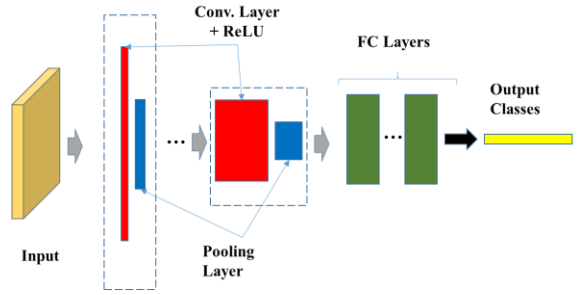
### 6.2.1 From CNN to FCN

As discussed in Chapter 3, the standard CNN structure mainly consists of two blocks, i.e. convolution layers and fully connected layers (Figure 6-1 (a)). This type of architecture takes a fixed size input and produces a single label as output. The FC can only take fixed-size inputs. The main contribution of changing a CNN to FCN is replacing the FC by CONVs, hence the architecture become a "fully convolutional network" (Figure 6-1 (b)). Typically, the last CONV has a 1×1 output. None of the layers have a predefined input size. As an interesting result, the network can handle inputs with any arbitrary size.

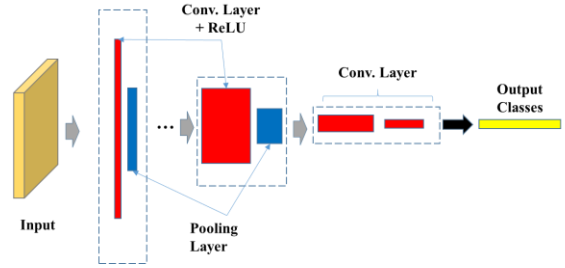
The first FCN was proposed by Matan *et al.* (Matan *et al.*, 1992) which was one-dimensional. Wolf and Platt (Wolf and Platt, 1993) proposed two-dimensional FCN. The main concern of deeper FCN in image segmentation is that several CONVs will affect the final resolution of the output. Long *et al.* (Long *et al.*, 2015) proposed a deep FCN for dense pixel classification, which fuses multi-resolution network layers to incorporate different resolutions of features. They suggested adding deconvolution layers (convolution layers with upsampling) to produce output with the same size of the input image (Figure 6-1 (c)). Therefore, their FCN generates pixelwise classification as output.

### 6.2.2 FCN for Dense Predication

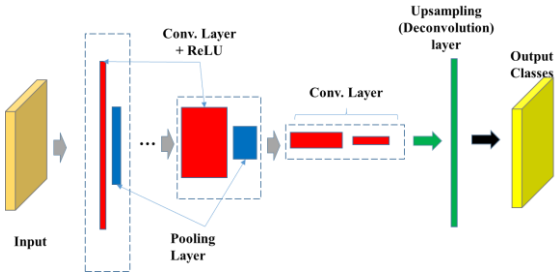
Each layer of data in a CNN generally can be considered as a 3D array of size $d_i$, $i \epsilon \{1,2,3\}$. In case of MRI multimodal data $d_1$ and $d_2$ are the spatial dimensions of images and $d_3$ is the number of channels which is the number of protocols in MRI data. It should be noted that the inputs of FCN are 2D slices of the MRI images. Each element in a deeper layer of a network is related to a specific region of the input image via the network connections, known as the receptive field of that element. The CNN substructures; i.e. CONV, POOL and activation function work on local input image regions, therefore they depend on relative spatial coordinates which make the CNN translation invariant.




a




b




c

Figure 6-1 Schematic architectures of the convolutional networks. a) standard CNN takes fixed size input in FC layer, b) standard FCN (classification) takes input with any size and the output is the feature vector or classification value, c) modified FCN for dense pixel segmentation takes input with any size and produces output with the same size.

In a particular layer, for the data vector $x_{ij}$ at location *(i, j)*, the output is calculated using

$$y_{ij} = f_{ks}\left(\{x_{si+\delta i,sj+\delta j}\}_{0\leq\delta i,\delta j\leq k}\right), \tag{6-1}$$

where $f_{ks}$ is the layer type, $k$ is the kernel size, and $s$ is the stride or subsampling factor. The layer type can be one of the following: convolution matrix, average pooling, spatial max for max pooling, nonlinearity for the activation function, or other types of layers. In general, deep networks compute a general nonlinear function.

The loss function is summed over the spatial dimension of the last layer, i.e.

$$l(x;\theta) = \sum_{ij} l'(x_{ij};\theta). \tag{6-2}$$

In this case, the gradients of the loss function are the sum of the gradients of each spatial component. The gradient descent on $l(x;\theta)$, which is computed over the image, is the same as gradient descent over $l'(x;\theta)$, which takes all the final layer receptive field as one batch. If these receptive fields overlap significantly, then the feedforward and backpropagation are more efficient when calculated layer by layer on the whole image compared to computing them independently patch by patch.

The FCs of a typical network have fixed dimensions, which can operate on input with fixed size. Therefore, FCs generate nonspatial outputs by throwing away the spatial coordinates. The replaced CONVs can be convolved with kernels that enable them to cover the entire input region and produce spatial outputs. Therefore, they can be suitable candidates for semantic segmentations. By providing the ground truth at the level of each output pixel, the forward and backward passes will be straightforward. The advantage of these type of networks is that they can take any input size. Whilst the weak point is the output dimensions are reduced by subsampling. The output size is reduced by a factor equal to the stride of the output layer's receptive fields which leads to coarse output results.

To address this problem, a deep stack of features is added to the final layer of activations. These new layers of features contain information from the shallower and deeper layers and are combined to provide both local and global features. The shallower layers (i.e. closer to the input) have higher resolution with more of the local information. The deeper layers (i.e. closer to the output) provide lower resolution features that have more semantic global information. Figure 6-2 shows the schematic illustration of using the multi-resolution information from different layers to produce dense output.

In order to perform dense prediction from the coarse segmentation (output image), upsampling or interpolation can be used. In the case of bilinear interpolation, each output is computed based on the nearest four inputs by using a linear map. Upsampling with factor *f* is equivalent to convolution with fractional input stride of *1/f*. It can also be explained as deconvolution or backward convolution with an output stride of *f*. The deconvolution layer is trainable i.e. a nonlinear upsampling can be learned by attaching the deconvolution layer and activation function.

To include the higher resolution information from the different level of layers, "skip layers" are added to the network. The schematic illustration of the concept and the skip layers are shown in Figure 6-2. Since these skip layers have different resolutions, they need to be interpolated and accumulated to produce the desired dense output. More details on the skip layers and upsampling layer of the FCN architecture will be explained in the next section.

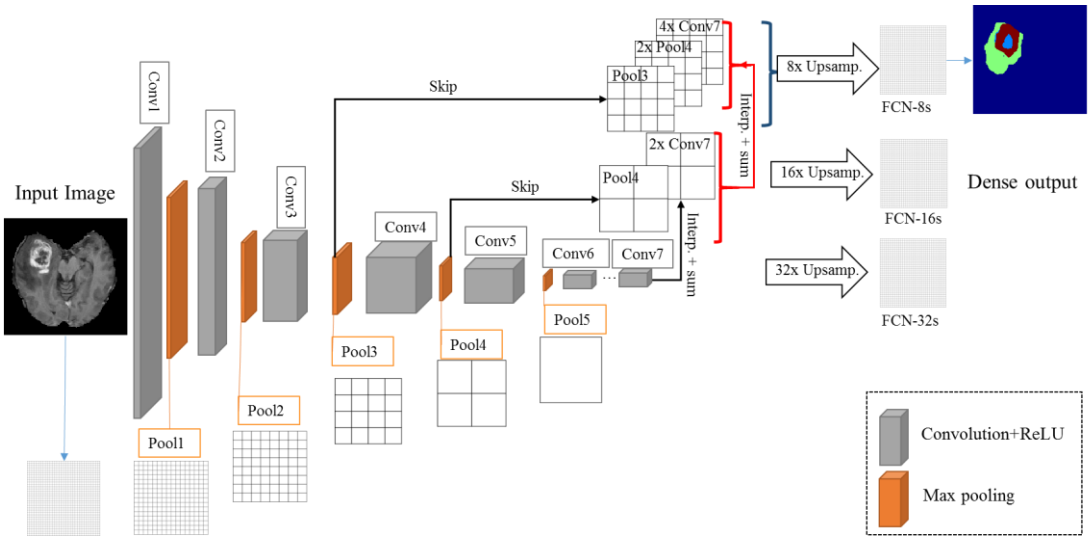



Figure 6-2 Schematic illustration of how different resolution feature information are considering from different depth of layers.

### 6.2.3 FCN Architecture

The details of the FCN architecture of VGG16 were shown in Figure 6-2. The VGG16 network is implemented using the Caffe model (Jia *et al.*, 2014). The network is modified by eliminating the final classifier layer and converting all the FC layers to convolution layers. Then a CONV of size 1 × 1 with channel dimension 5 (i.e. number of output classes) is added to predict the scores for each pixels of multimodal MRI images which produces the coarse

segmentation output. A deconvolution layer is added in order to upsample the coarse output bilinearly to the pixel-wise fine segmentation.

The output stride is divided in half by predicting from a 16-pixel stride layer. A 1 × 1 convolution is added to the *POOL4* to further predict the classes. The corresponding output is combined with the output of *CONV7* at stride 32 by adding a 2x upsampling layer. The upsampling layer is initially a bilinear interpolation and its parameters can be learned later. The stride 16 predictions are then upsampled back to the image. This structure is called FCN-16. The initial parameters of *POOL4* are set to zero, so the starting prediction of the network is unmodified.

The same procedure is applied on *POOL3* with a 2x upsampling which is combined with *POOL4* and *CONV7* which results in FCN-8s architecture. To produce finer predictions, it is recommended to decrease the strides of the pooling layers. This procedure is a difficult task in VGG16 structure. To set the *POOL5* stride to 1, the convolutionalised *FC6* should have the kernel size 14 × 14. This will increase the computational cost and those large filters makes the learning process more difficult.

**FCN for Brain Tumour Segmentation**

In this chapter, the FCN-8s architecture in (Long *et al.*, 2015) is modified and adopted for segmentation of brain tumour in multimodal MRI images, where the VGG16 (Simonyan and Zisserman, 2014) is employed as a CNN classification net. The architecture of FCN based on VGG16 for brain tumour segmentation is presented in Figure 6-3. In the FCN architecture, initially, the classification net is transformed to be a fully convolutional net, then adding an upsampling or de-convolutional layer to it for pixel-wise predictions. The FCN training is an end-to-end supervised learning procedure and the image segmentation is performed using a pixel-wise prediction/classification. The FCN-8s constructed from FCN-16s skip net and FCN-32s coarse net.

The predictions at shallow layers are produced using a skip layer that combines coarse predictions at deep layers to improve segmentation details. More specifically in the experiments, the FCN-8s is implemented by fusing predictions of shallower layer (*POOL3*) with 2x upsampling of the sum of two predictions derived from *POOL4* and last layer. Then the stride 8 predictions are upsampled back to the image.

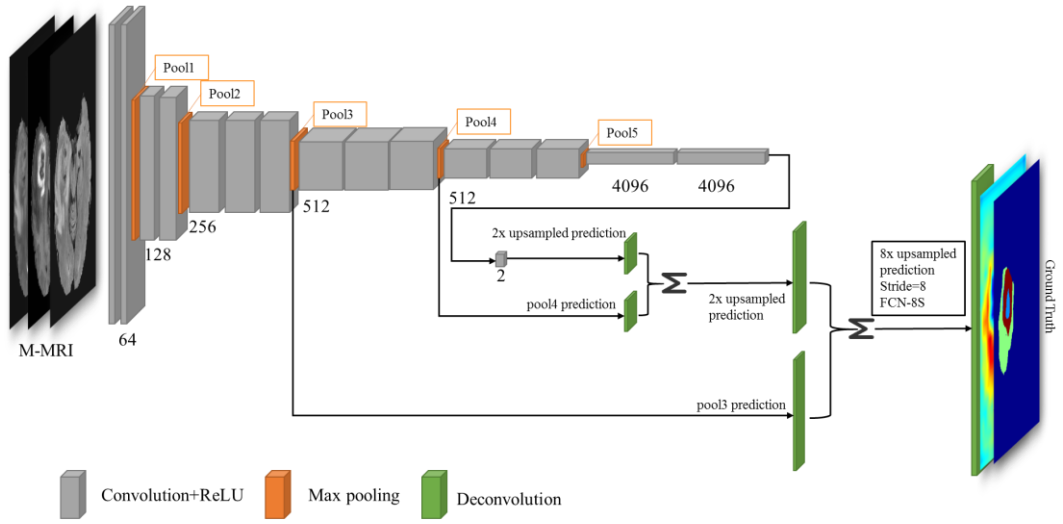



Figure 6-3 The detailed architecture of the FCN used for segmentation of brain tumour in multimodal MRI.

**Pretrained VGG16**

The VGG model is designed based on natural image with intensities with bit depth = 8 (i.e. the intensity range is [0 255]). The VGG model has also the capability of being further trained by adding more datasets to the existing pretrained model. The initialisation using a pretrained model decreases the training time. Furthermore, the first layers of any CNN learn more abstract features of the images which is common between all kind of images. Therefore, a pretrained model on other images will improve the model accuracy for new images. In this thesis, due to the insufficient training MRI data, a pre-trained model on natural colour images is used. The DSC measure 0.64 is obtained using a model without pretraining for the whole tumour, which is a poor segmentation. It should be noted that the pre-trained model architecture was designed for 2D images. Therefore, a 2D architecture was implemented in this thesis to use the pretrained model parameters accordingly.

Using a pretrained model results in faster training procedure and provides more accurate results in a specific number of epochs. However, it should be noted that the data used for the pretrained model was different from MRI. This is still a disadvantage that should be considered the training of the model is not completely end-to-end. Therefore, the term “partially” end-to-end is used since the model parameters were initiated from a model that was trained on a different set of image data.

In order to use the pretrained VGG model for the case of medical MR images, some modifications should be taken into account. The intensities of the images are normalised into the range of a natural image which is [0, 255]. The pretrained VGG model is trained on RGB images which has three channels. Therefore, three protocols were selected to build the input data. Regarding the importance of the protocols which were discussed in Chapter 5 (Section 5.3.2), FLAIR, T1-contrast and T2-weighted are selected. Due to the lack of sufficient available DTI MRI dataset, *p* and *q*-map protocols were not used in this stage of the thesis. However, regarding to the experiments in Chapter 5, incorporating the DTI protocols may result in better segmentation of the tumour tissue subtypes.

**Segmentation of MRI Tumour using FCN**

The final segmentation is obtained by max voting to the final score maps of the FCN. The results of dense classification output (segmentation masks) are generated using FCN-8s network and shown in Figure 6-4.

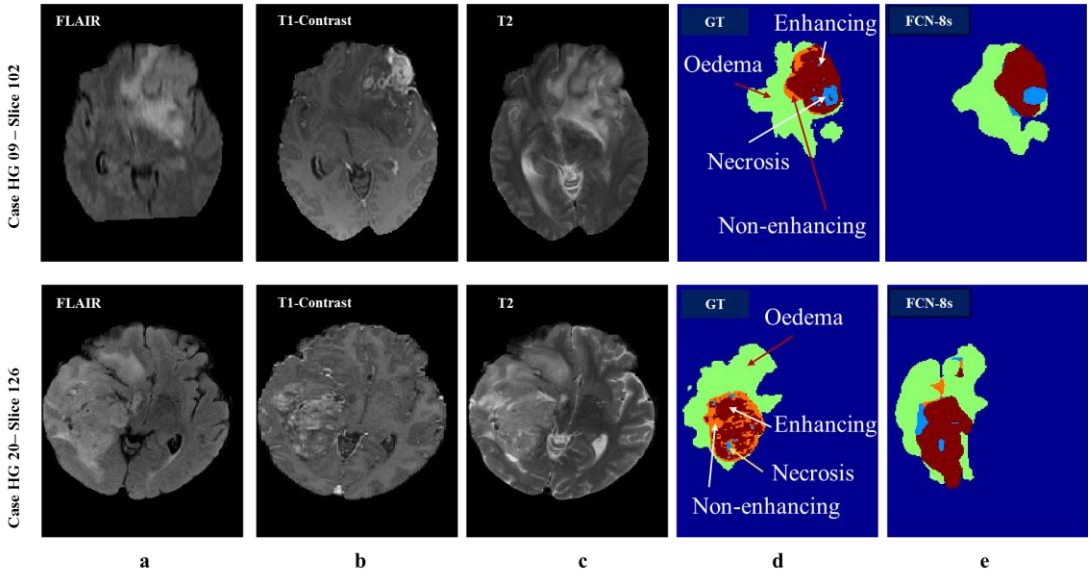



Figure 6-4 Comparison of the multimodal MRI segmentation output of the FCN-8s network with the ground truth.

Comparing the segmentation output (Figure 6-4(e)) with the ground truth (Figure 6-4 (d)), it can be seen that the FCN-8s was successful to globally detect the tumour structures. However, the lack of the spatial regularisation of FCN results in label disagreement between similar pixels and diminishes the spatial consistency for segmentation and results in coarse

segmentation. Incorporating more information from the local dependencies of the pixel (e.g. hand-designed features) may improve the results by providing finer segmentation. In the next section a hybrid method will be introduced which considers the three-dimensional connectivity (neighbourhood system) to compensate the FCN limitation of coarse segmentation.

### 6.2.4 Fusing Hand-designed Features with FCN via a Hybrid Method

As discussed in Chapters 4 and 5, the texton features based on Gabor filters provide strong descriptors which represent the local dependencies in the spatial and frequency domains. Fusing the texton features with the FCN based features will incorporate more local information to the final segmentation.

**Overview of the method**

The method is comprised of four major steps (pre-processing, CNN design, Texton map extraction, RF classifier) that are depicted in Figure 6-5. In the pre-processing stage, intensity histograms of the different modalities are normalised. The images are then fed into the FCN system to create the score maps for the classes. The texton maps are also extracted directly from the images. Both machine-learned features (i.e. from FCN) and hand-designed features (i.e. from texton map) are then combined and fed into a RF classifier to classify each MRI image pixel into normal brain tissues and different tumour structures.

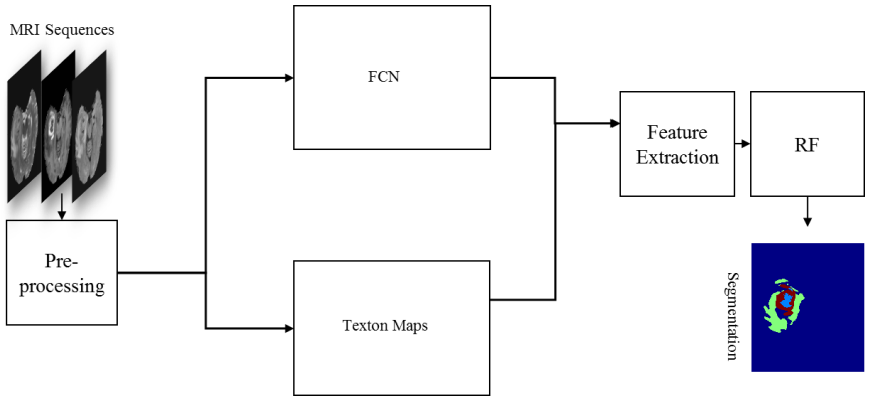



Figure 6-5 The overall flowchart of the hybrid method which uses hand-designed and machine-learned features for automatic brain tumour segmentation in MRI images.

**Preprocessing**

The preprocessing stage includes intensity normalisation and histogram matching, which have been discussed in Chapter 4 (Section 4.2.2). In this chapter, the same procedure is performed for the training dataset. For the testing dataset, the data are normalised and matched directly to the average template obtained from the training dataset.

**Feature Extraction form the FCN Layers**

The machine-learned features are the score maps generated from the FCN for each individual output class. The final segmentation output of the FCN was obtained by assigning the label of maximum values of the final score maps after the deconvolution layer. Those score maps can also be considered as the feature maps and include all the hierarchies of the features from coarser (lower resolution) to finer (higher resolution). The feature (score) maps and their location in the FCN architecture are illustrated in Figure 6-6.

The number of score maps is the number of classification labels including the background. For example, in the case of the BRATS dataset standard labellings (or the score maps) are five, including:

1. Background and normal brain
2. Necrosis
3. Oedema
4. Non-enhancing tumour
5. Enhancing tumour

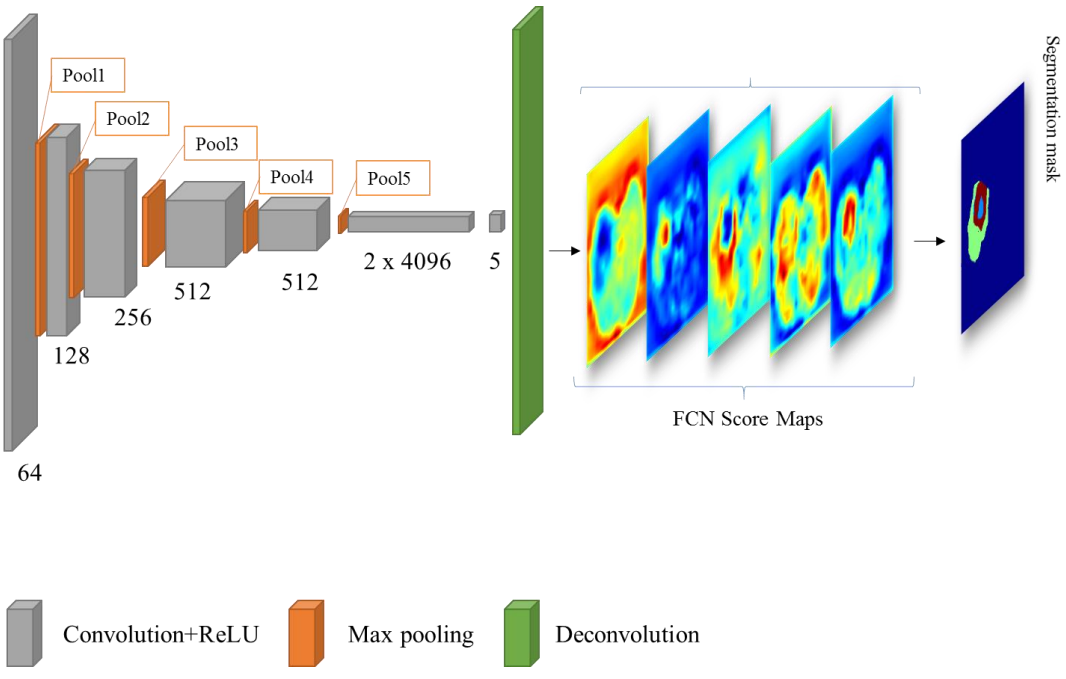



Figure 6-6 The score maps are extracted from the deconvolution layer from the FCN.

For each pixel, a five-dimensional feature vector is constructed, with the value of each element in the feature vector is that equivalent to the value of each score map layer for that corresponding pixel. Figure 6-7 shows the FCN based score maps for each class layer and the ground truth for a testing image. Figure 6-7 also shows the protocols which were used to generate the corresponding score maps.

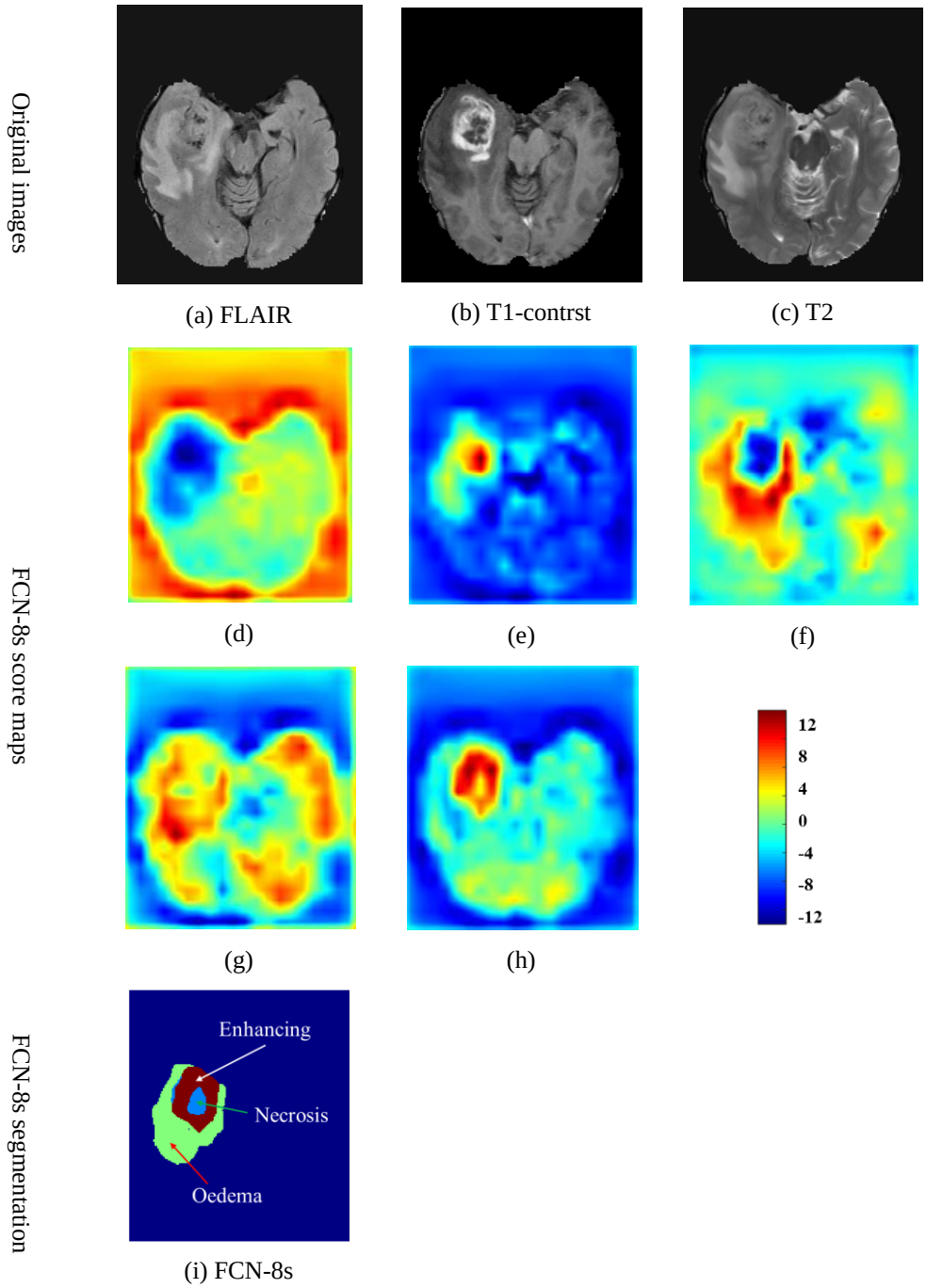



Figure 6-7 The FCN-based score maps generated from the multimodal MRI images. The figures are shown for challenge case number HG-0309 (ID: 17604) and slice 59. a-c) Original MRI: a) FLAIR, b) T1-contrast, and c) T2-weighted. d-h) FCN score maps for classes: d) background and normal brain, e) necrosis, f) oedema, g) non-enhancing tumour, and h) enhancing tumour. i) The segmentation mask using FCN-8s.

**Spatial Feature Extraction**

The coarse results of the FCN-based segmentation are due to the down-sampling that occurs in the pooling layers. Given that the local dependency is not sufficiently considered in the FCN, a new pipeline of the texton based feature descriptor derived from the convolutional feature is proposed. Textons are used since they are powerful features that represent the local dependencies and neighbourhood system information (Section 4.2.5) in all phases, i.e. convolution, *k*-means clustering and histogram calculation. The texton maps are created using the 3D method discussed in Section 5.2.4. The number of maps is equal to the number of MRI protocols ($N_{protocol}$) which are used for the segmentation. It is noted that, 3 protocols are used regarding to the machine-learned FCN based feature calculation (Section 6.2.3), whereas in the previous work (i.e. SV_RF method in Chapter 5), 4 protocols were used (Section 5.2.4). The reason for using 3 protocols is that the pretrained model is based on three-channel natural images. The T1-weighted protocol is excluded since it provides less information on the tumour subtypes compared to T1-contrast. Figure 6-8 shows the texton maps for the corresponding slices presented in Figure 6-7.

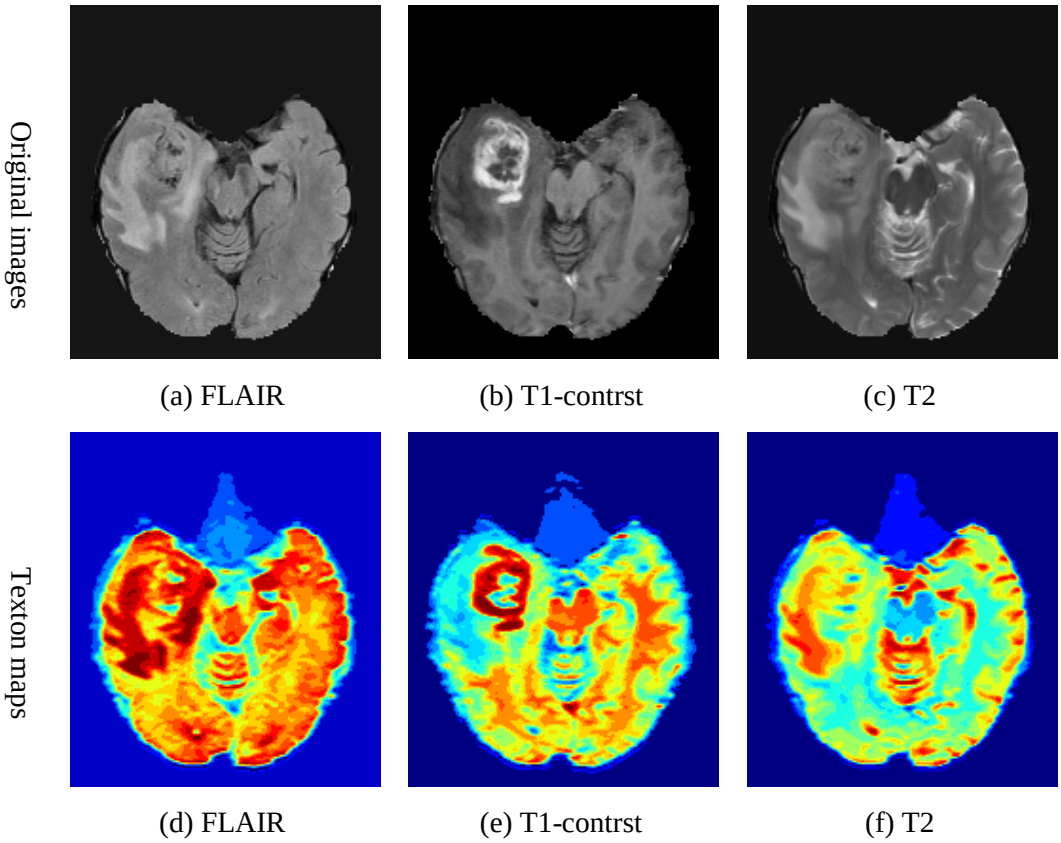



Figure 6-8 The texton maps generated from the M-MRI images. The figures are shown for challenge case number HG-0309 (ID: 17604) and slice 59. a-c) Original MRI: a) FLAIR, b) T1-contrast, and c) T2-weighted. d-f) texton feature map extracted separately from the protocols: d) FLAIR, e) T1-contrast, and f) T2-weighted.

It can be seen in Figure 6-8-(d) to (f), that the texton maps created detailed presentations of the local boundaries of the tumour parts. By comparing Figure 6-8-(a) to Figure 6-8-(d), the oedema region is clearly distinguishable from the map extracted from the FLAIR protocol. Also, Figure 6-8-(e) presents an obvious separation of the tumour core parts from enhancing tumour, which is the bright area in the T1-contrst image and the red/brown area in the T1c texton map. The necrosis is the blue area inside the tumour core which is equivalent to the dark area in the T1-contrast original image in Figure 6-8-(b).

Each pixel in the image is described by its intensity and neighbourhood pixels. For each pixel, its neighbourhood with size $n \times n$ pixel is represented by the histogram of texton while the centre pixel is represented by its normalised intensity. This descriptor of each pixel implicitly encodes the information that the centre pixel conditionally depends on its neighbourhood, thus incorporates the local dependencies into feature representation. The procedure of mapping the connectivity based on the texton *ID*s (in the neighbourhood window) to the histogram is illustrated in Figure 6-9.

Figure 6-10 and Figure 6-11 show how the connectivity based on the texton *ID*s of adjacent pixels in the neighbourhood provides feature representation at the pixel level. Figure 6-10 shows the T1-contrast image, FCN-based learned score map of the "enhancing" class, the texton map of T1-contrast, and the manual GT. The features (i.e. FCN-based score of the "enhancing" class and connected textons *ID* histogram) are extracted for two pixels with different GT labels (enhancing tumour and necrosis). The FCN-based score values for those pixels are 3.88 and 3.15 respectively. These values are close together regarding to the whole range of [-10, +12] for FCN-based scores of the "enhancing" class. Therefore, both pixels may be assigned to the same class if only FCN-based score feature is considered. Whilst, considering the neighbourhood system based on the texton *ID*s, the corresponding histogram provides a feature representation which makes those classes, i.e. enhancing and necrosis, more separable. Both cases are shown in the lower row of Figure 6-11.

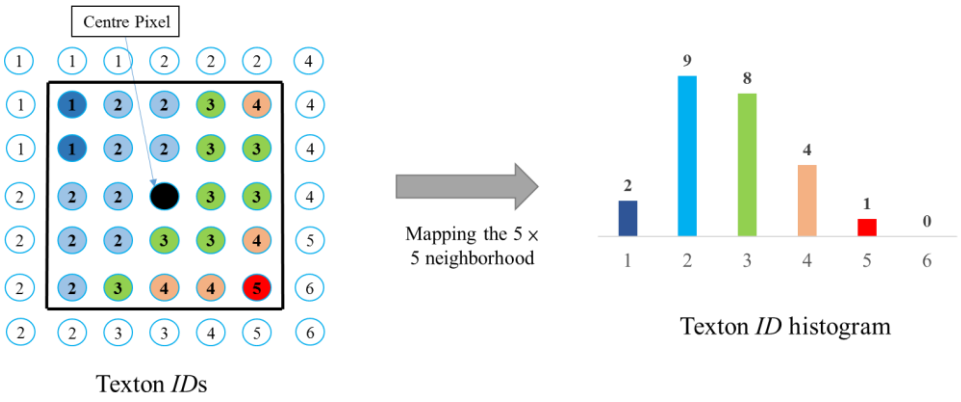



Figure 6-9 The connectivity of the adjacent pixels from the histogram of the texton *IDs* in a 5 × 5 neighbourhood of the centre pixel. The texton clusters are integers in the range [1, 6]. The texton *IDs* outside the neighbourhood are not counted. This is a simplified example to illustrate the procedure of texton histogram feature extraction from the pixel neighbourhood. The texton histogram values are zero for *IDs* from 6 to 16 in this example.

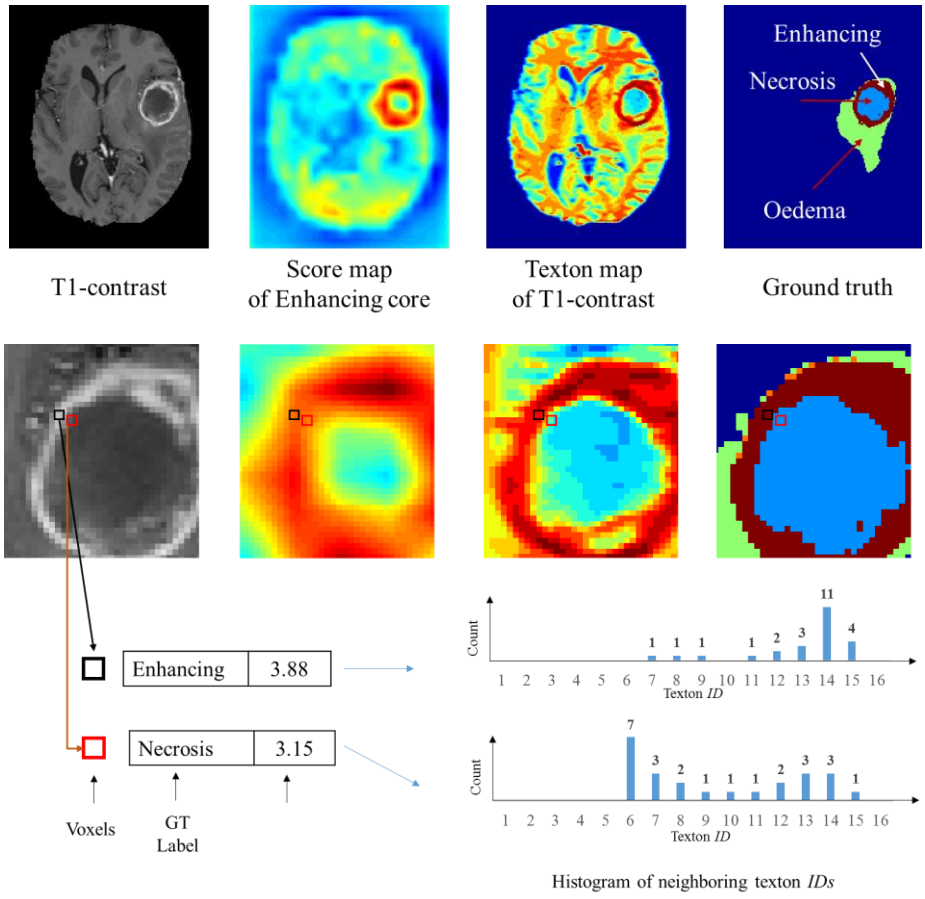



Figure 6-10 (Upper row) T1-contrast, FCN-based score map of enhancing core, texton map of T1-contrast, and ground truth; (middle row) the corresponding close up of the upper row images and two different pixels with GT labels: enhancing (black square) and necrosis (red square); (lower row) the features extracted for the corresponding pixels including FCN-based score of "enhancing", and connected textons *ID* histogram in a 5 × 5 neighbourhood around the centre pixel.

Figure 6-11 shows the texton map of the FLAIR image, FCN-based score map of the oedema class, and the feature representation of two pixels with different labels (oedema and normal brain). The patient image and slice is the same as Figure 6-10, but the selected pixels are different. FCN-based scores values for the corresponding pixels are 7.79 and 6.85, respectively. These values are very close compared to the range of the FCN scores of "oedema" for this patient case, which is in the range [-9, +10]. Considering the neighbourhood system based on the texton *ID*s, the corresponding histogram provides a feature representation which makes the classes, i.e. oedema and normal brain, more separable which is shown in the lower row of Figure 6-11.

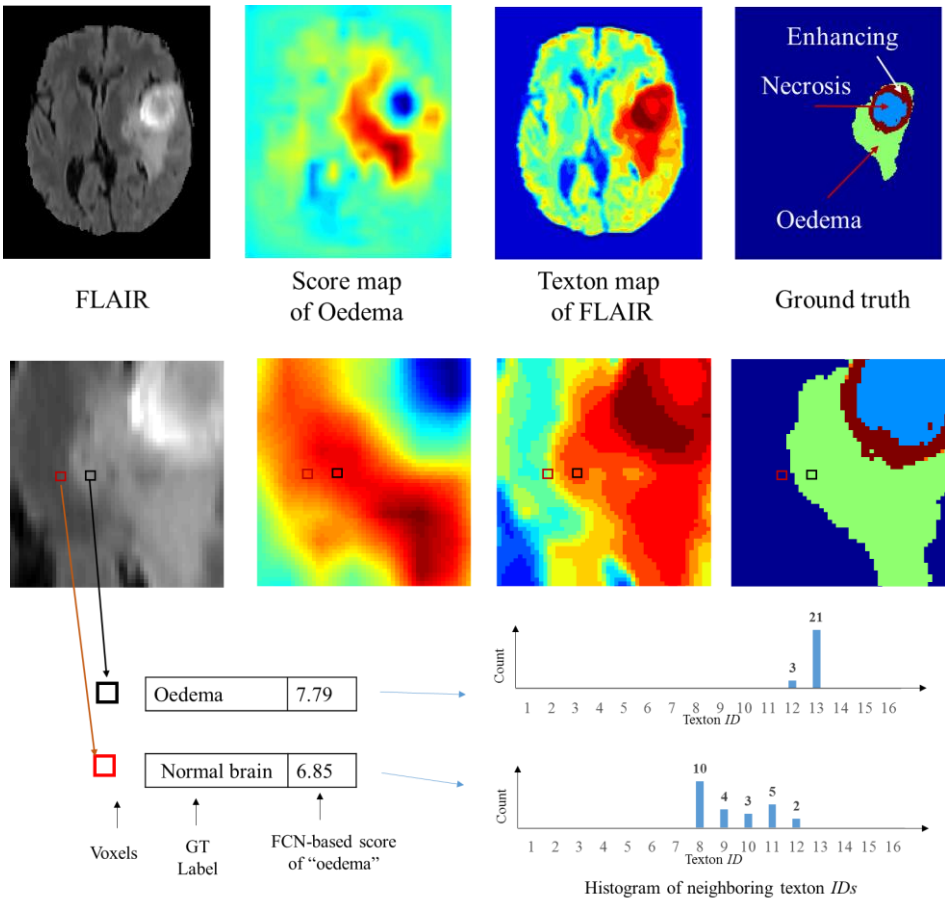



Figure 6-11 (Upper row) FLAIR, FCN-based score map of "oedema", texton map of FLAIR, and ground truth; (middle row) the corresponding close up of the upper row images and two different pixels with GT labels: oedema (black square) and normal brain (red square); (lower row) the features extracted for the corresponding pixels including FCN-based score of "oedema", and connected textons *ID* histogram in a 5 × 5 neighbourhood around the centre pixel.

In total 56 features were calculated which are summarised in Table 6-1. The features from FCN are chosen from each layer of the score maps for the corresponding pixel. Whilst, the texton features are extracted based on the histogram of texton map in a fixed size window of 5 × 5, centred at that pixel. The number of texton features is based on the number of texton clustering which is chosen $k_{texton}$ = 16 (Section 6.3.3).

Figure 6-12 shows a graphic presentation of feature vector generation for the hybrid method and presents how machine-learned and hand-designed features are integrated together.

Table 6-1 Details of the features which are used in the proposed method.

| Feature | Type | Number for each protocol | Total number |
|---|---|---|---|
| FCN | Machine-learned | - | 5 |
| Texton | Hand-designed | 16 | 48 |
| Normalised intensity | Hand-designed | 1 | 3 |
| Total | Combined | - | 56 |

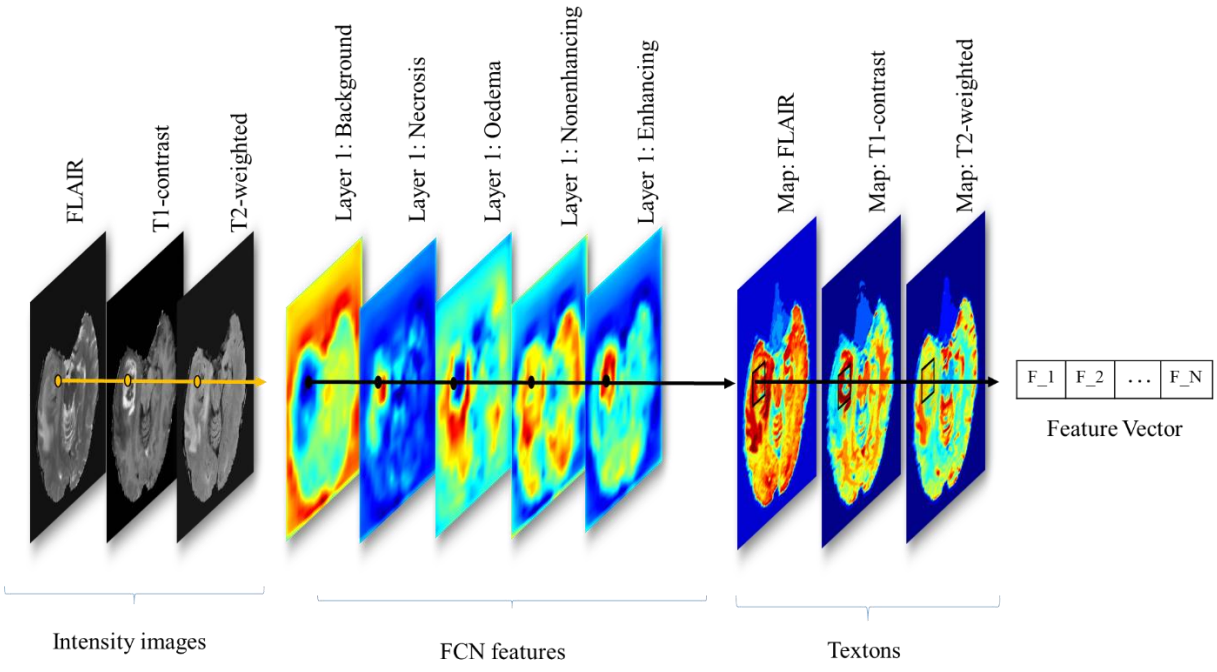



Figure 6-12 The detailed of feature vector generation for the hybrid method. Machine based features from FCN are extracted based on pixel and the hand-designed texton features extracted from the neighbourhood around the pixel and considers the local dependencies.

**RF and segmentation**

The RF classifier is applied locally on tumour candidate region which is detected by the FCN. The final output of the FCN creates segmentation masks which encompass the complete tumour region. This will reduce the number of pixels to be classified by RF. A confidence margin of 10 pixels in 3D space around the detected tumour area is selected by morphological dilation (Gonzalez and Woods, 2007). This is to ensure more parts of the tumour which may not detected by the FCN are included and also to include some normal brain parts to make a more balanced classification. For each pixel in this target area, a feature vector (i.e. 56 features for each pixel) is then fed to the RF for training.

The RF was explained in detail in Chapter 5. The RF parameters tuning will be discussed later in Section 6.3.2. The final segmentation mask is obtained based on the classes assigned for each pixel in the test dataset by mapping back the pixel estimated class from RF to the segmentation mask volume.

In the post-processing stage, the false positive related to the remaining skull area are eliminated using a connected component analysis. The connected components are considered the pixels which are connected in 3D with 26 adjacencies. The connected components which include less than 300 pixels are excluded from the segmentation mask. This value was obtained using the training set.

## 6.3 Experiments and Results

The experiments were evaluated on the publicly available MICCAI BRATS 2013 and 2017 dataset. The method was tested on an independent "challenge" dataset by uploading the results on the VSD system (using the required format) and obtaining the evaluation results comparison with other competitors from the organiser website. Quantitative evaluations of the proposed method were based on the overlap measures for segmented tumour vs ground truth which has been calculated and reported by the VSD website. In the following subsections, the dataset, parameter setting of the models and the evaluation of segmentation results will be described.

**Dataset**

For all of the experiments, the systems are trained based on the BRATS 2013 and 2017 training dataset. The training dataset consists of 30 (for BRATS 2013) and 285 (for BRATS 2017) patient MRI scans that were explained earlier in Section 2.6.1. The evaluation is based on application of the BRATS 2013 challenge (test) and 2017 validation dataset to the trained

system. The test dataset consists of 10 cases with HGG. The FCN architecture, i.e. VGG16, were designed for three layers of input image. Therefore, three protocols were used: FLAIR, T1-contrast and T2-weighted to train the FCN. The reason for selection of this configuration is that these three protocols contain the most information on all the tumour structures in the clinical settings which was also used in Chapter 5. The tumour type classes are provided by VSD system, i.e. oedema, necrosis, enhancing and non-enhancing tumour.

**Evaluation**

The evaluation measures are the standard segmentation which are used in the VSD system to compare the segmentation results with the gold standard (blind testing). DSC, PPV, and sensitivity are considered. The evaluation results were provided using the BRATS challenge standard combination which are as follow:

3. Complete tumour (oedema, necrosis, enhancing and non-enhancing)
4. Tumour core (necrosis, enhancing and non-enhancing)
5. Enhancing tumour

The segmented masks obtained from the challenge testing dataset are uploaded to the VSD website and evaluated by the corresponding online system (blind testing).

**Implementation**

The experiments were implemented on a PC with CPU Intel Core i7 and RAM 16 GB with the operating system windows 8.1. The FCN was implemented using MatCovNet toolbox (Vedaldi and Lenc, 2015). GPU GeForce gtx980i was used for reducing the training time of the FCN. Other sections of the proposed method were performed using MATLAB 2016b. RF open source code provided in (Taormina, n.d.), which is a specialised toolbox for RF classification based on MATLAB, was utilised for the classification.

**Experiment Setting**

To evaluate the performance of the proposed method, three sets of comparative experiments were investigated.

1. **FCN**
   In this scenario, a FCN is directly applied to the clinical dataset to segment the tumour regions. The FCN structure is designed based on (Long *et al.*, 2015). For this phase, no hand designing or parameter tuning is needed.
2. **FCN_RF**
   In this scenario, the score map of the trained FCN in the previous scenario is extracted. The score maps are generated by applying the FCN on the BRATS training set. The score

maps are considered as feature vectors and then fed into a RF classifier. In this phase, the RF parameters should be tuned for this specific set of features. The parameters are tree depth ($D_{tree}$) and the number of trees ($N_{tree}$). The RF parameters are obtained by 4-fold cross validation on the BRATS training dataset. This experiment was designed to assess the effect of local dependencies with and without textons.

3. **FCN_Texton_RF**
   This is the proposed method which was explained in Section 6.2.4. In this phase, the texton parameters also should be tuned and are the number of texton clusters ($k_{texton}$) and Gabor filters parameters, i.e. filter orientations, sizes, and wavelength of sinusoid coefficients.

### 6.3.1 FCN

Table 6-2 provides the evaluation results obtained by applying the FCN segmentation on the BRATS 2013 challenge dataset. Figure 6-13 shows the examples of segmented tumour parts in some cases from the BRATS 2013 challenge dataset by the method based on FCN only. The original images of FLAIR and T1-contrast protocols are shown and the segmentation results are presented with the coloured areas overlaid on the FLAIR image. As the ground truth is not accessible, it cannot be included it in Figure 6-13. Table 6-2 shows that that the sensitivity is very good for segmentation using FCN only but the DSC and PPV are below 0.80 for most of the cases. This explains that FCN can detect the area that includes the tumour, but it cannot accurately and locally detect its boundaries. In the third column of Figure 6-13 which shows the segmentation overlays based on FCN only method, the FCN over-segmented the tumour area especially the tumour core. Although it detects the location of the tumour but it was not accurate in determining the exact tumour boundaries.

Table 6-2 Segmentation results per case for BRATS 2013 challenge dataset using FCN only, evaluated by the VSD website system.

| Case | Dice score | | | Positive Predictive Value | | | Sensitivity | | |
|---|---|---|---|---|---|---|---|---|---|
| | Complete | Core | Enhancing | Complete | Core | Enhancing | Complete | Core | Enhancing |
| HG-0301 | 0.80 | 0.68 | 0.78 | 0.77 | 0.62 | 0.82 | 0.82 | 0.76 | 0.75 |
| HG-0302 | 0.73 | 0.75 | 0.65 | 0.63 | 0.62 | 0.69 | 0.87 | 0.94 | 0.63 |
| HG-0303 | 0.65 | 0.83 | 0.82 | 0.49 | 0.75 | 0.70 | 0.97 | 0.92 | 0.98 |
| HG-0304 | 0.80 | 0.70 | 0.50 | 0.71 | 0.55 | 0.35 | 0.91 | 0.98 | 0.90 |
| HG-0305 | 0.59 | 0.39 | 0.44 | 0.86 | 0.94 | 0.65 | 0.45 | 0.25 | 0.33 |
| HG-0306 | 0.84 | 0.80 | 0.67 | 0.84 | 0.88 | 0.65 | 0.84 | 0.73 | 0.69 |
| HG-0307 | 0.89 | 0.29 | 0.50 | 0.83 | 0.17 | 0.37 | 0.96 | 0.95 | 0.76 |
| HG-0308 | 0.83 | 0.81 | 0.58 | 0.75 | 0.72 | 0.43 | 0.95 | 0.93 | 0.89 |
| HG-0309 | 0.86 | 0.80 | 0.70 | 0.89 | 0.69 | 0.61 | 0.83 | 0.94 | 0.82 |
| HG-0310 | 0.81 | 0.69 | 0.76 | 0.80 | 0.63 | 0.78 | 0.82 | 0.77 | 0.74 |
| **Total** | **0.78** | **0.67** | **0.64** | **0.76** | **0.66** | **0.61** | **0.84** | **0.82** | **0.75** |

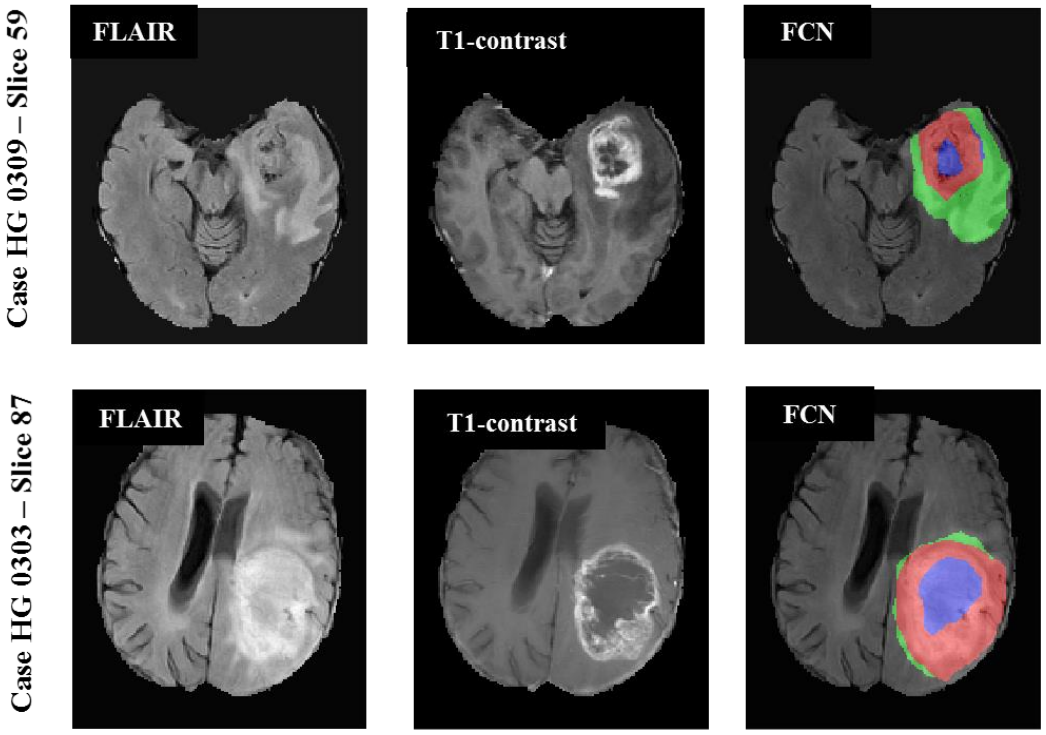



Figure 6-13 Segmentation results for some cases of BRATS 2013 challenge dataset. a) the original FLAIR images, b) T1-weighted-contrast, c) segmentation mask of FCN only overlaid on FLAIR image. Oedema: green, necrosis: blue, enhancing tumour: red.

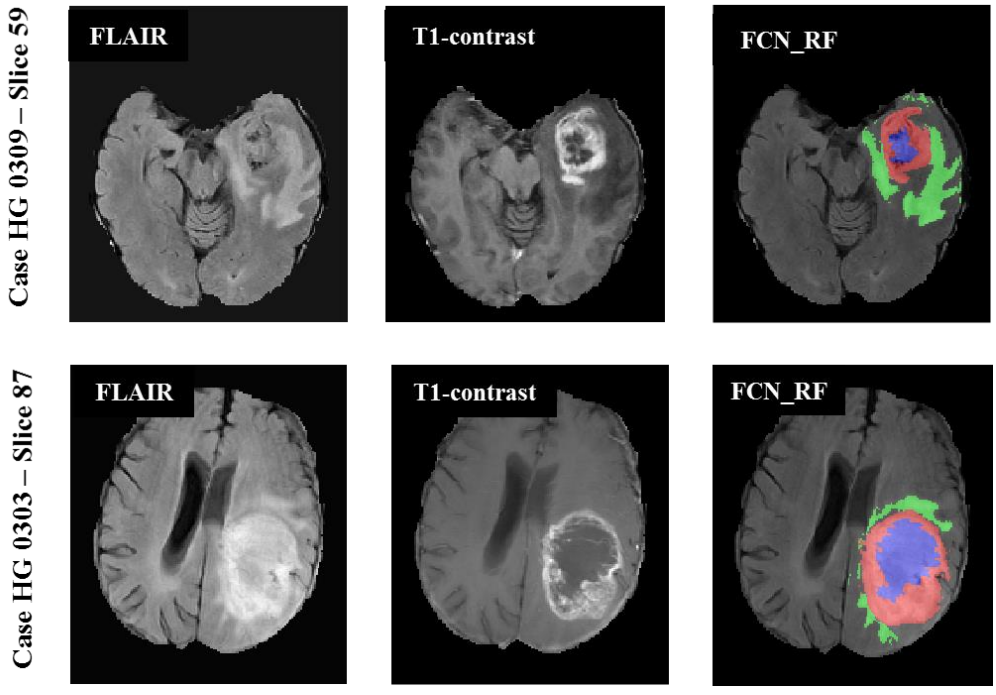
### 6.3.2 FCN_RF

RF parameters were tuned by examining different tree depths and number of trees on the BRATS 2013 training patient datasets. A 4-fold cross validation is used for evaluating the classification accuracy during the tuning process. The number of trees $N_{tree}$ = 50 with depth $D_{tree}$ = 15 provide an optimum generalisation and accuracy.

Table 6-3 provides the evaluation results obtained by applying the FCN_RF method on the BRATS 2013 challenge dataset. Figure 6-14 shows the examples (same as Figure 6-13) of the segmentation of the tumour using FCN_RF in the BRATS 2013 challenge dataset. The original images of FLAIR and T1-contrast protocols are shown and the segmentation results are presented with the coloured areas overlaid on the FLAIR image.

Table 6-3 Segmentation results per case for the BRATS 2013 challenge dataset using FCN_RF, evaluated by the VSD website system.

| Case | Dice score | | | Positive Predictive Value | | | Sensitivity | | |
|---|---|---|---|---|---|---|---|---|---|
| | Complete | Core | Enhancing | Complete | Core | Enhancing | Complete | Core | Enhancing |
| HG-0301 | 0.86 | 0.86 | 0.77 | 0.93 | 0.93 | 0.80 | 0.79 | 0.80 | 0.74 |
| HG-0302 | 0.87 | 0.69 | 0.82 | 0.88 | 0.95 | 0.89 | 0.87 | 0.54 | 0.76 |
| HG-0303 | 0.88 | 0.93 | 0.80 | 0.93 | 0.92 | 0.71 | 0.84 | 0.94 | 0.90 |
| HG-0304 | 0.77 | 0.73 | 0.62 | 0.84 | 0.66 | 0.56 | 0.71 | 0.82 | 0.70 |
| HG-0305 | 0.82 | 0.70 | 0.76 | 0.88 | 0.95 | 0.76 | 0.77 | 0.55 | 0.76 |
| HG-0306 | 0.87 | 0.79 | 0.68 | 0.97 | 0.98 | 0.94 | 0.79 | 0.66 | 0.53 |
| HG-0307 | 0.88 | 0.24 | 0.61 | 0.91 | 0.15 | 0.83 | 0.86 | 0.60 | 0.48 |
| HG-0308 | 0.91 | 0.91 | 0.70 | 0.92 | 0.95 | 0.60 | 0.90 | 0.88 | 0.85 |
| HG-0309 | 0.82 | 0.88 | 0.81 | 0.97 | 0.92 | 0.88 | 0.71 | 0.85 | 0.76 |
| HG-0310 | 0.83 | 0.89 | 0.81 | 0.94 | 0.95 | 0.84 | 0.75 | 0.83 | 0.77 |
| **Total** | **0.85** | **0.76** | **0.74** | **0.92** | **0.84** | **0.78** | **0.80** | **0.75** | **0.73** |




a b c

Figure 6-14 Segmentation results for some cases of BRATS 2013 challenge dataset. a) the original FLAIR images, b) T1-weighted-contrast, c) segmentation mask of FCN_RF overlaid on the FLAIR image. Oedema: green, necrosis: blue, enhancing tumour: red.

The results in Table 6-3 shows that FCN_RF slightly improved the DSC for the complete tumour and significantly improved the tumour core and enhancing part and also improved PPV for all areas. It means that the segmentation boundaries are now closer to the ground truth.

While sensitivity for the complete tumour decreases, which represents under-segmentation. The reason is that FCN_RF considers only the normalised intensity of pixels and it does not take into account the local dependencies.

### 6.3.3 FCN_Texton_RF

As described in Chapter 4, the Gabor filter parameters are chosen using exhaustive grid search. Six different filter directions were used: [$0^o$, $30^o$, $45^o$, $60^o$, $90^o$, $120^o$] to cover more orientations. Filter sizes were chosen in the range from 0.3 to 1.5 as: [0.3 0.6 0.9 1.2 1.5]. The wavelength of the sinusoid coefficients of the Gabor filters were chosen [0.8, 1.0, 1.2, 1.5].

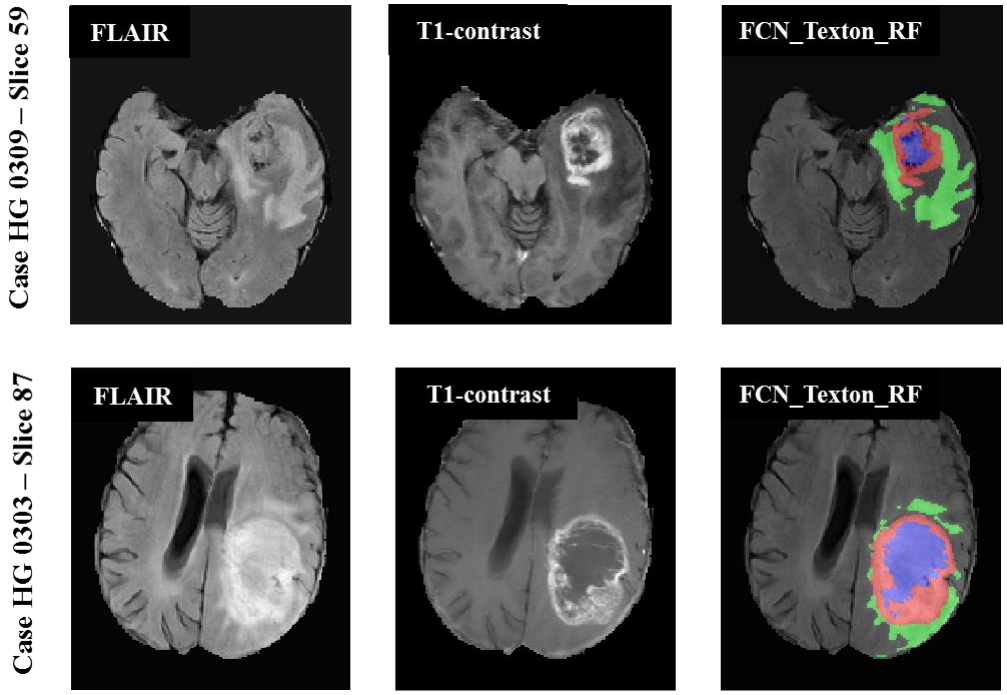
The number $k_{texton} = 16$ was selected as the optimal value for the number of clusters in texton map. Zheng *et al.* (Zhang *et al.*, 2016) proposed that the number $k_{texton} = 32$ is optimal for the texton map extraction for a complete image. In this thesis, the aim is to calculate the texton histogram in a bounding box neighbourhood. Using $k_{texton} = 32$ produce might produce very sparse histogram for a small neighbourhood window (e.g. 5 × 5). On the other hand, the number of clusters should be sufficient to separate all possible tissues in the image. In the case of MRI brain images, the complexity and variety of tissue intensities and textures should be considered. Therefore, the $k_{texton} = 16$ is chosen to allocate at least two *ID*s for each main tissue (i.e. WM, GM, CSF, oedema, necrosis, enhancing, non-enhancing, and other background). The texton map is created by assigning the cluster number to each pixel of the image and then sorted based on the average intensity values of the original image pixels inside each cluster. This is to organise/relabel the clusters as *k*-means assigns the initial cluster number randomly. The texton feature for each pixel is the histogram of textons in a neighbourhood window of 5 × 5 around that pixel.

As explained in Chapter 4, the optimal value for $k_{attribute}$ in the classification using RF is $k_{attribute} = \sqrt{N_{feature}}$ where $N_{feature}$ is the total number of features, in the current study $k_{attribute} = 7$. RF parameters were selected similar to the previous section, i.e. the number of trees $N_{tree} = 50$ with depth $D_{tree} = 15$.

Table 6-4 provides the evaluation results obtained by applying FCN_Texton_RF on the BRATS 2013 challenge dataset. Figure 6-15 shows the examples (same as Figure 6-13) of the segmentation of tumour structures in some of the BRATS 2013 challenge dataset by FCN_Texton_RF. The original images of FLAIR and T1-contrast protocols are shown and the segmentation results are presented with the coloured areas overlaid on the FLAIR image.

Table 6-4 Segmentation results per case for the BRATS 2013 challenge dataset using FCN_Texton_RF, evaluated by the VSD website system.

| Case | Dice score | | | Positive Predictive Value | | | Sensitivity | | |
|---|---|---|---|---|---|---|---|---|---|
| | Complete | Core | Enhancing | Complete | Core | Enhancing | Complete | Core | Enhancing |
| HG-0301 | 0.87 | 0.88 | 0.79 | 0.82 | 0.93 | 0.82 | 0.91 | 0.83 | 0.76 |
| HG-0302 | 0.86 | 0.70 | 0.82 | 0.78 | 0.98 | 0.91 | 0.96 | 0.55 | 0.75 |
| HG-0303 | 0.89 | 0.89 | 0.75 | 0.96 | 0.97 | 0.82 | 0.83 | 0.83 | 0.70 |
| HG-0304 | 0.86 | 0.64 | 0.57 | 0.79 | 0.60 | 0.81 | 0.94 | 0.69 | 0.44 |
| HG-0305 | 0.90 | 0.74 | 0.76 | 0.86 | 0.97 | 0.85 | 0.95 | 0.60 | 0.70 |
| HG-0306 | 0.91 | 0.84 | 0.63 | 0.92 | 0.95 | 0.95 | 0.90 | 0.75 | 0.47 |
| HG-0307 | 0.91 | 0.53 | 0.59 | 0.94 | 0.38 | 0.45 | 0.87 | 0.86 | 0.84 |
| HG-0308 | 0.89 | 0.92 | 0.71 | 0.97 | 0.96 | 0.60 | 0.81 | 0.88 | 0.85 |
| HG-0309 | 0.88 | 0.91 | 0.81 | 0.92 | 0.98 | 0.92 | 0.84 | 0.85 | 0.72 |
| HG-0310 | 0.89 | 0.91 | 0.84 | 0.86 | 0.96 | 0.89 | 0.91 | 0.87 | 0.79 |
| **Total** | **0.89** | **0.80** | **0.73** | **0.88** | **0.87** | **0.80** | **0.89** | **0.77** | **0.70** |




a b c

Figure 6-15 Segmentation results for some cases of the BRATS 2013 challenge dataset. a) the original FLAIR images, b) T1-weighted-contrast, c) segmentation mask of the FCN_Texton_RF method overlaid on the FLAIR image. Oedema: green, necrosis: blue, enhancing tumour: red.

As can be seen in Table 6-4, adding the texton-features to the pipeline improves the overlap measure for complete tumour and increases the sensitivity while it slightly decreases the PPV.

Therefore, FCN_Texton_RF improves the overlap measure while maintaining a balance between sensitivity and PPV.

The average and standard deviation values of the DSC for the three above mentioned experiments are compared in Figure 6-16. The values are compared separately for each tumour part; i.e. complete tumour, core and enhancing. The results show that using RF instead of the last classification layer of the FCN slightly improves the segmentation in the FCN_RF experiment. The reason is that the normalised intensity of the pixels is taken into account for classification. Adding the hand-designed features (e.g. texton histogram) to the classification pipeline significantly improves the segmentation. As shown in Figure 6-16, there is an evident difference between segmentation overlap measures for all the tumour structures using both machine-learned and hand-designed features. The result is not good for the segmentation of the enhancing tumour using only machine-learned features regardless of the classifier type (e.g. FCN or RF). Whilst the segmentation results based on both feature types, i.e. hand-designed and machine-learned are consistent for all tumour structures.

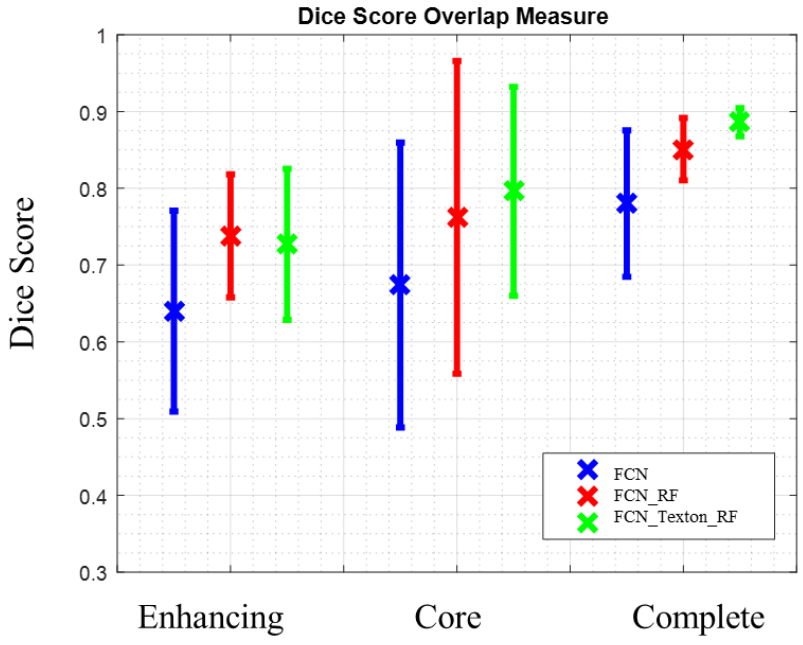



Figure 6-16 Comparison of DSC (average and standard deviation) for the three experiments separated for whole, tumour core and enhancing tumour.

The average and standard deviation values of the PPV for the corresponding experiments are compared in Figure 6-17. The values are compared separately for each tumour part; i.e.

complete tumour, core and enhancing. The results show that the PPV of FCN_Texton_RF is higher than the other two experiments for the tumour core. FCN_RF provides the highest PPV for the complete tumour. However, the PPV is 0.88 for the proposed method and is good compared to the state-of-the-art (Section 6.4). PPV values are not good for all the tumour structures using FCN only. It shows that FCN_RF provides more accurate segmentation compared to FCN only. This also provides evidence to the fact that FCN provides coarser segmentation.

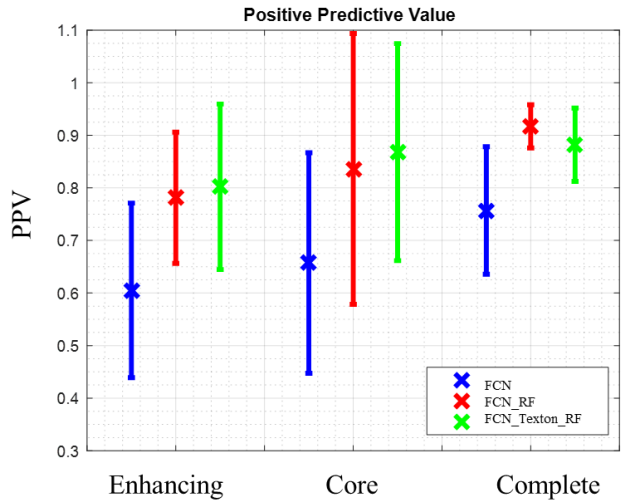



Figure 6-17 Comparison of PPV (average and standard deviation) for the three experiments separated for whole, tumour core and enhancing tumour.

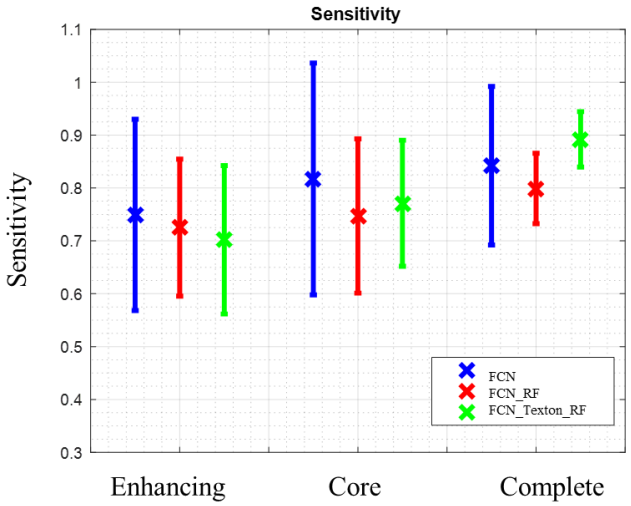



Figure 6-18 Comparison of Sensitivity (average and standard deviation) for the three experiments separated for whole, tumour core and enhancing tumour.

The average and standard deviation values of sensitivity measures for the corresponding experiments are compared in Figure 6-18. The values are again compared separately for each tumour part; i.e. complete tumour, core and enhancing. The results show that the sensitivity of FCN_Texton_RF for the complete tumour has the highest value. The FCN only provides better sensitivity for the tumour core and enhancing core compared to the other two experiments which means that it was able to encompass the tumour structure although it provides coarse segmentation. FCN_RF presents the lowest mean sensitivity for complete and tumour core.

### 6.3.4 Evaluation on BRATS 2017 Dataset

The proposed FCN_Texton_RF method was also implemented on BRATS 2017 challenge dataset. The model was trained on 285 training sets (both HGG and LGG). Then the model was tested on 46 validation datasets. The results were uploaded into the Centre for Biomedical Image Computing and Analytics (CBICA) image processing portal ("Penn Imaging – Home," 2017). Table 6-5 provides the evaluation results obtained by applying the FCN_Texton_RF method on the BRATS 2017 validation dataset. The results of this thesis appear as the team name "LoVE" (Laboratory of Vision Engineering) on the CBICA portal leaderboard ("MICCAI-BraTS 2017 Leaderboard," 2017). The precision measure was not provided by the portal. The Dice score and sensitivity measures for the whole tumour are 0.86 and 0.83, respectively, and are comparable to the BRATS 2013 results. However, the segmentation results on the BRATS 2017 are a little lower than BRATS 2013 experiments. This can be due to the modification in the labelling that was merging non-enhancing with necrosis. The other reason can be that more complicated and challenging tumour images were added to the 2017 challenge. The Dice score and sensitivity measures for the tumour core are 0.86 and 0.83, respectively, which are close to, but lower than, the BRATS 2013 results. The corresponding results of the enhancing tumour are 0.66 and 0.57, respectively. These results are lower than the BRATS 2013 experiment. It should be noted the BRATS 2017 challenge dataset includes more LGG cases, and hence might affects the segmentation of tumour tissue subtypes.

Table 6-5 Segmentation results for the BRATS 2017 validation dataset, which was provided by CBICA portal. The results are the overall average and standard deviation of 46 patient cases.

| Case | Dice score | | | Sensitivity | | |
|---|---|---|---|---|---|---|
| | Complete | Core | Enhancing | Complete | Core | Enhancing |
| Mean | 0.86 | 0.78 | 0.66 | 0.83 | 0.72 | 0.57 |
| STD | 0.09 | 0.19 | 0.28 | 0.13 | 0.21 | 0.28 |

Figure 6-19 shows segmentation results of the proposed method on some cases of the BRATS 2017 validation and testing datasets. As can be seen in both Table 6-5 and Figure 6-19, the segmentation results for the whole tumour are very good, similar to the segmentation results on the BRATS 2013. By comparing the enhancing tumour segmentation results to the hyper-intense tumour area in T1-contast images, it can be seen that the FCN_Texton_RF method was successful in excluding the hyper-intense regions in the normal brain from the tumour.

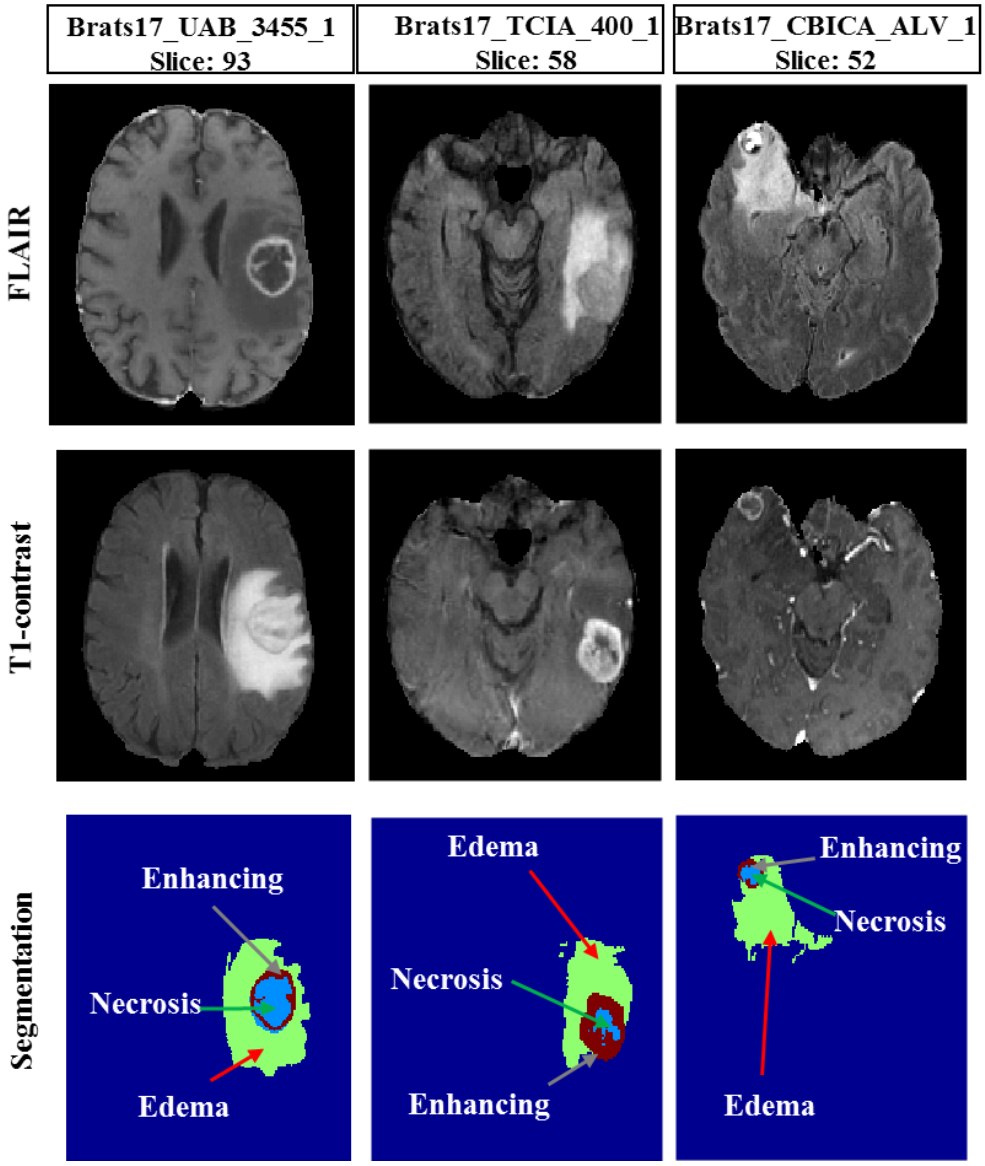



Figure 6-19 Segmentation masks for some validation datasets, using the proposed method. The case names and the slice number are mentioned on the top of the images. Upper row) FLAIR images, middle row) T1-contrast images, lower row) segmentation masks and labels using the proposed method.

## 6.4 Discussion

In this study, the publicly available BRATS 2013 dataset was used since it is a popular standard dataset in this field. The training dataset included the ground truth used for parameter optimisation and training the FCN and RF. The evaluation was used on the challenge (independent testing) dataset without having access to the ground-truth which make it a challenging competition. A fair comparison with other recent and related work which used the same dataset will be also feasible.

The experiments show that the segmentation method based on FCN was able to automatically locate the tumour areas and encompass them with a high sensitivity. The main problem with the method was that it could not accurately segment the tumour structures because of the coarse segmentation. To tackle this problem, local dependencies and the neighbourhood system of the pixel were taken into account into classification by using hand-designed texton features. The experimental results demonstrate that this modification to the FCN system increases the accuracy of segmentation while keeping balance between sensitivity and PPV.

The other advantage of the proposed method is that by using 3D textons the connectivity information will be considered in all directions in 3D space. This modification compensates the other limitation of FCN which works with 2D slices. Although modifying the FCN architecture may overcome this limitation, currently the pretrained models are available only for 2D FCN network.

The training stage in the proposed method is time consuming and is considered as a limitation. However, when the model is learned, the prediction process for new datasets is fast for both FCN and RF. Also, the FCN model can be saved and refined by adding any future training dataset to the previously trained model.

Table 6-6 provides the results of the methods which were discussed in Section 6.3 and applied on BRATS 2013 clinical challenge dataset. It also provides the results of the related top-ranked methods in BRATS 2013 Challenge. Table 6-6 also includes the methods which were proposed by Pereira *et al* (Pereira *et al.*, 2016), Havaei *et al.* (Havaei *et al.*, 2017), Davy *et al.* (Davy *et al.*, 2014), and Urban *et al.* (Urban *et al.*, 2014). Those methods are amongst the best publications presented on the website scoreboard which used the BRATS 2013 challenge dataset. In the third row of Table 6-6, the values mentioned in parentheses show the current ranking of each individual measure for the corresponding tumour part using FCN_Texton_RF in the VSD website. The screenshot of the VSD scoreboard for the top-rated methods which are evaluated on the challenge dataset based on blind testing is provided in Appendix 2. The overall rank of FCN_Texton_RF was 6$^{th}$ for the challenge dataset (Appendix 2).

Table 6-6 Segmentation results for BRATS 2013 challenge dataset which is evaluated by VSD website. Comparison with other works which used BRATS 2013 challenge dataset and are top ranked. The values which are presented in parentheses in the third row are the current ranking of the proposed method in each section, on VSD scoreboard (Appendix 2).

| **Method** | **Dice score** | | | **Positive Predictive Value** | | | **Sensitivity** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Complete** | **Core** | **Enhancing** | **Complete** | **Core** | **Enhancing** | **Complete** | **Core** | **Enhancing** |
| FCN | 0.79 | 0.69 | 0.62 | 0.77 | 0.68 | 0.58 | 0.83 | 0.82 | 0.75 |
| FCN _RF | 0.80 | **0.83** | 0.71 | 0.90 | **0.90** | 0.73 | 0.74 | 0.78 | 0.78 |
| FCN_Texton_RF | **0.88 (1)** | 0.80 (9) | 0.73 (21) | 0.88 (10) | 0.87 (6) | **0.80 (5)** | 0.89 (25) | 0.77 (24) | 0.70 (34) |
| Pereira (2016) (Pereira *et al.*, 2016) | **0.88** | **0.83** | 0.77 | 0.88 | 0.87 | 0.74 | 0.89 | 0.83 | 0.81 |
| Kwon (2014) (Kwon *et al.*, 2014) | **0.88** | **0.83** | 0.72 | **0.92** | **0.90** | 0.74 | 0.84 | 0.78 | 0.72 |
| Tustison (2014) (Tustison *et al.*, 2014) | 0.87 | 0.78 | 0.74 | 0.85 | 0.74 | 0.69 | 0.89 | **0.88** | **0.83** |
| Havaei (2017)(Havaei *et al.*, 2017) | **0.88** | 0.79 | 0.73 | 0.89 | 0.79 | 0.68 | 0.87 | 0.79 | 0.80 |
| Urban (2014) (Urban *et al.*, 2014) | 0.86 | 0.75 | 0.73 | 0.82 | 0.75 | 0.79 | **0.92** | 0.79 | 0.70 |
| Davy (2014) (Davy *et al.*, 2014) | 0.85 | 0.74 | 0.68 | 0.85 | 0.74 | 0.62 | 0.85 | 0.78 | 0.77 |
| Meier (2014)(Meier *et al.*, 2014b) | 0.82 | 0.73 | 0.69 | 0.76 | 0.78 | 0.71 | **0.92** | 0.72 | 0.73 |
| Reza (2013) (Reza and Iftekharuddin, 2013) | 0.83 | 0.72 | 0.82 | 0.82 | 0.81 | 0.70 | 0.86 | 0.69 | 0.76 |
| Zhao (2013) (Zhao *et al.*, 2013) | 0.84 | 0.70 | 0.80 | 0.80 | 0.67 | 0.65 | 0.89 | 0.79 | 0.70 |
| Cordier (2013) (Cordier *et al.*, 2013) | 0.84 | 0.68 | **0.88** | 0.88 | 0.63 | 0.68 | 0.81 | 0.82 | 0.66 |
| Festa (2013) (Festa *et al.*, 2013) | 0.72 | 0.66 | 0.77 | 0.77 | 0.77 | 0.70 | 0.72 | 0.60 | 0.70 |
| Doyle (2013) (Doyle *et al.*, 2013) | 0.71 | 0.46 | 0.66 | 0.66 | 0.38 | 0.58 | 0.87 | 0.70 | 0.55 |

Figure 6-20 shows comparison of the DSCs for FCN, FCN_RF, FCN_Texton_RF, and the top ranked methods in the BRATS 2013 challenge. The results are sorted based on the overall DSC to provide a better visualisation of the relative ranking of the methods. In terms of DSC for the complete tumour, the FCN_Texton_RF method is the highest score in the chart which is 0.88 and currently ranked 1st on the VSD scoreboard (Appendix 2). The results and rankings were acquired on 27th June 2017. FCN_RF provided DSC of 0.88 which is higher than 1st ranked DSC of 0.83, i.e. Pereira *et al.* (Pereira *et al.*, 2016) (Appendix 2). FCN_Texton_RF provided DSC of 0.80 which is very close to the top rated (i.e. 0.83). FCN_RF provides the best DSC for core. As explained in Chapter 2 (Section 2.6.2), the manual ground truth for the enhancing core was annotated using thresholding. Since thresholding works directly on the pixel values, and the FCN_RF method considers the normalised intensity (not the local dependencies) as the main feature, their segmentations provide more overlap.

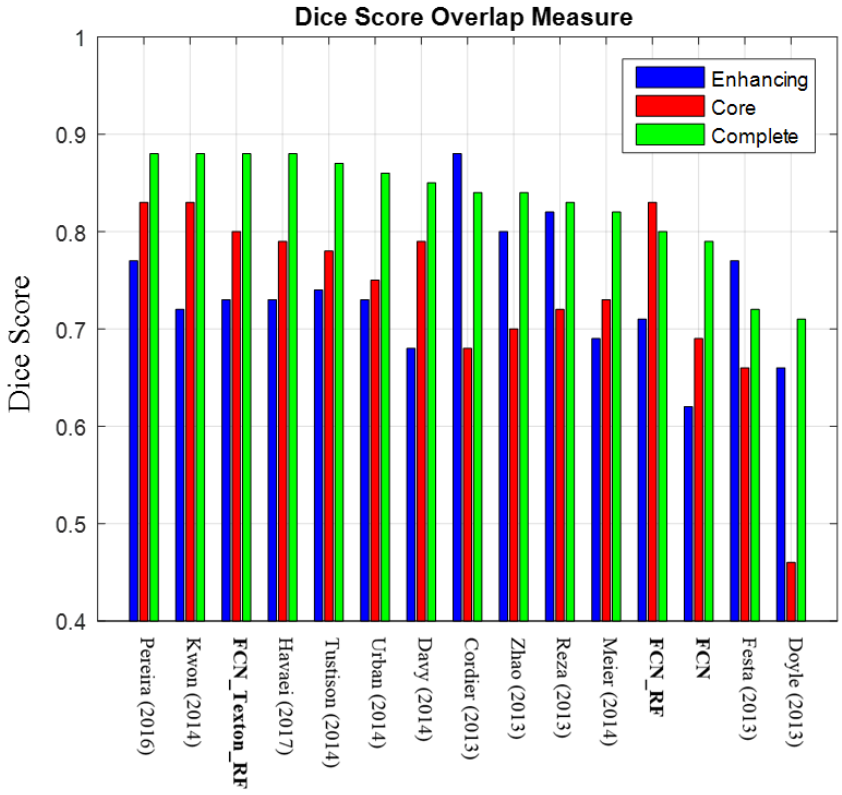



Figure 6-20 Comparison of DSC overlap measure with top ranked methods which used BRATS 2013 challenge dataset.

Figure 6-21 shows comparison of the PPV for FCN, FCN_RF, FCN_Texton_RF, and the top ranked methods in the BRATS 2013 challenge. The results are sorted the same as in Figure 6-20. In terms of PPV for the complete tumour and core, the FCN_RF method is amongst the highest scores in the chart which is 0.90 for both methods. For the enhancing tumour, FCN_Texton_RF provides PPV of 0.80 which is amongst the bests cores in the chart. For the complete and tumour core, FCN_Texton_RF has PPV of 0.88 and 0.87 respectively, which are similar to the score of the 1st ranked method, i.e. Pereira *et al.* (Pereira *et al.*, 2016), and very close to the top rated, i.e. Kwon *et al.* (Kwon *et al.*, 2014), which are 0.92 and 0.90, respectively.

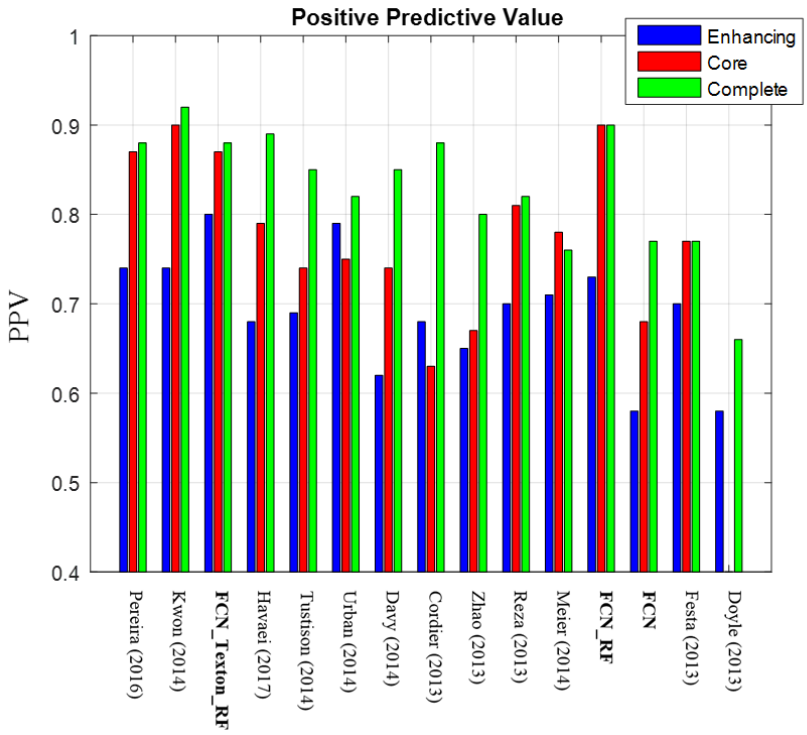



Figure 6-21 Comparison of PPV measure with top ranked methods which used BRATS 2013 challenge dataset.

Figure 6-22 shows comparison of the sensitivity for FCN, FCN_RF, FCN_Texton_RF, and the top ranked methods in the BRATS 2013 challenge. The results are sorted the same as in Figure 6-20. The sensitivity of FCN_Texton_RF for the complete tumour is 0.89, whereas it is very close to the 1st rank which is 0.92. The sensitivity of the FCN only method for tumour core is 0.82, whereas the best sensitivity score is 0.88 in the chart. For the tumour core and enhancing, FCN_Texton_RF has sensitivity of 0.77 and 0.70 respectively, which are lower than the top ranked method. The best corresponding sensitivity values were reported by Tustison *et al.* (Tustison *et al.*, 2014) which are 0.88 and 0.83, respectively (Appendix 2).

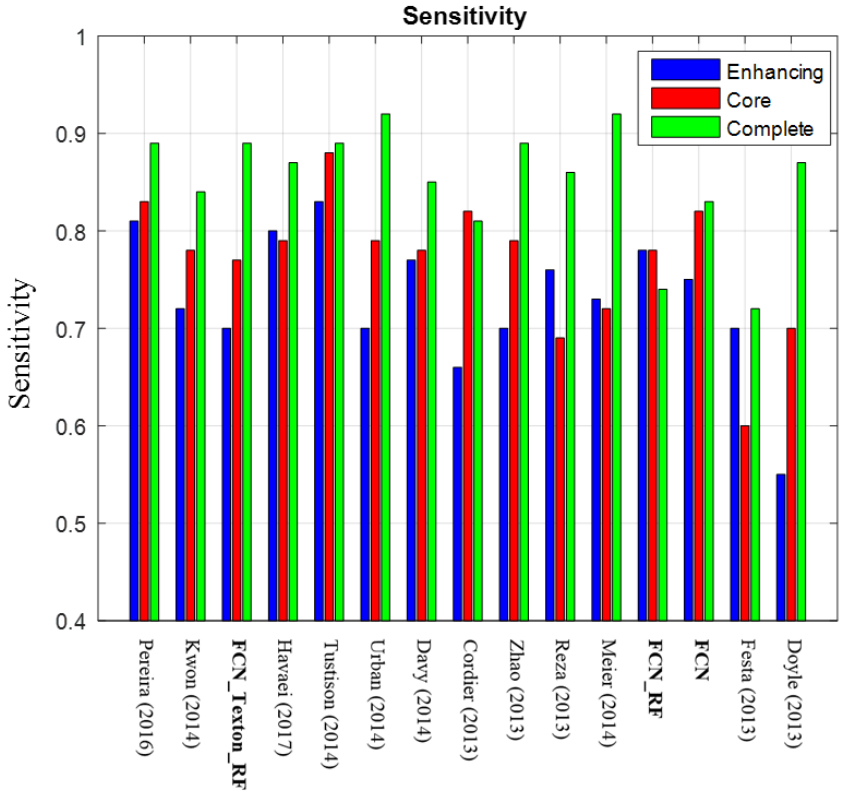



Figure 6-22 Comparison of sensitivity measure with top ranked methods which used BRATS 2013 challenge dataset.

Among the methods that are presented in the Table 6-6, Figure 6-20, Figure 6-21, and Figure 6-22, the methods which were proposed by Havaei *et al.* (Havaei *et al.*, 2017), Davy *et al.* (Davy *et al.*, 2014), Pereira *et al.* (Pereira *et al.*, 2016), and Urban *et al.* (Urban *et al.*, 2014) are based on CNN. The methods based on CNN have high rankings in Table 6-6 according to the VSD system, amongst which Pereira *et al.* (Pereira *et al.*, 2016) has the highest ranking. This shows the importance and proficiency of machine-learned based methods using deep learning. Regarding to the VSD results, none of the methods could achieve the best score for all the metrics. The method proposed by Pereira *et al.* (Pereira *et al.*, 2016) used a deep neural

network and also results in the best performance in the BRATS 2013 challenge dataset. They suggested that applying more nonlinearity on the data improves the segmentation of a brain tumour. They achieved the best rank in terms of DSC for all tumour structures. FCN_Texton_RF method has the same DSC, PPV and sensitivity for the complete tumour to theirs. Other methods that have the best DSC for the complete tumour are Havaei *et al.* (Havaei *et al.*, 2017) and Kwon *et al.* (Kwon *et al.*, 2014). The most difficult task in the brain tumour challenge according to Menze *et al.* (Menze *et al.*, 2015) is the segmentation of the enhancing region for HGG and the core tumour for LGG. There is a large difference in segmentation of the core region. Havaei *et al.* (Havaei *et al.*, 2017) used concatenation of two CNNs' pathways to include more context into training. Whilst FCN_Texton_RF method has one CNN pathway and meanwhile textons provide more context in terms of textural patterns. Therefore, FCN_Texton_RF method outperformed compared to the other methods in PPV.

The method proposed by Tustison *et al.* (Tustison *et al.*, 2014), which used RF and hand-designed features, was the winner of the on-site BRATS 2013 challenge. FCN_Texton_RF method outperformed (Tustison *et al.*, 2014) in terms of DSC for complete tumour and core and PPV value for all tumour tissue types.

The method proposed by Kwon *et al.* (Kwon *et al.*, 2014) which was an atlas based approach had the ranking 2$^{nd}$ at the time of their publication and their current ranking is 7$^{th}$ (Appendix 2). Their method has the same DSC and PPV for the complete tumour while FCN_Texton_RF sensitivity is higher. Their method is slightly better for core segmentation. However, FCN_Texton_RF outperforms theirs for segmentation of enhancing tumour. The method proposed by Zhao *et al.* (Zhao *et al.*, 2013) also used CNN. FCN_Texton_RF outperforms theirs significantly in terms of PPV for all the tumour structures. However, DSC of FCN_Texton_RF method is slightly better than theirs.

The ranking provided by the VSD system was based on considering all the measures from different tissues. Although the FCN_Texton_RF method was ranked 6$^{th}$, the Dice score for segmentation of the whole tumour was ranked 1$^{st}$. It should be noted that the ranking of the FCN_Texton_RF method is tied, as other methods have the same Dice score and ranking for the whole tumour.

The structure of the whole tumour usually has a clear appearance in FLAIR and T2-weighted, except the infiltration of oedema in some cases. The tumour core tissue subtypes have more complex textures and appearances, which make them difficult to segment. Therefore, the proposed FCN_Texton_RF and other methods provide better performance for the whole tumour, compared to the tumour core. On the other hand, excluding T1-weighted from the experiment (due to the limitation of the FCN_Texton_RF network that accepts only three

protocols) decreases the accuracy of the tumour core segmentation, compared to some of the state-of-the-art methods.

As discussed in Chapter 3 (Section 3.5.2), U-Net (Ronneberger *et al.*, 2015) is a recent CNN architecture proposed to improve image segmentation accuracy that uses upsampling paths from both shallow and deep layers. The FCN_Texton_RF that was proposed in this thesis incorporates the local connectivity of pixels to compensate the loss of information occurred in the pooling layers. Therefore, U-Net can be compared to the proposed FCN_Texton_RF method. Dong *et al.* (Dong *et al.*, 2017) applied the U-net for brain tumour segmentation and evaluated it on the BRATS 2015 dataset. They did not report evaluation results on the BRATS 2013 dataset, hence they are not mentioned in Table 6-6 and Figure 6-20 to Figure 6-22. The evaluation results are provided only for the training set. They reported DSC measures of 0.86, 0.86 and 0.65 for the complete tumour, core and enhancing tumour, respectively. For the whole tumour, both U-Net and FCN_Texton_RF provided high accuracy. In the case of tumour core, U-Net provided more accurate segmentation, while the FCN_Texton_RF method outperformed U-Net for the enhancing tumour. It should be noted that the enhancing tumour appears with more jagged edges compared to oedema edges (that are usually smoother).

## 6.5 Limitations

The proposed FCN_Texton_RF method includes two trainable stages, which are designed and tuned separately. The RF and FCN stages are both computationally expensive. This is one of the drawbacks of the FCN_Texton_RF method. However, using parallel processing will reduce the computation time. In this thesis, the FCN was implemented on the GPU, while RF was trained on the CPU. The FCN architecture was comprised of a huge number of parameters that require a large amount of memory space.

Another limitation of using two stages in the FCN_Texton_RF method is that the results of RF depend on the results produced by the FCN. The effect of this dependency is not studied in this thesis. As an example, the RF parameters were set based on the experience and the outcome of the previous chapter (Chapter 5). The parameter optimisation was also conducted independently.

Another limitation of the FCN_Texton_RF, compared to the state-of the-art methods, is the sensitivity of segmentation for the enhancing tumour. Although the DSC for the enhancing tumour was comparable to the top-ranked method, few methods obtained DSC measures above 0.80 (Cordier *et al.*, 2013).

The proposed FCN_Texton_RF method was trained on a pre-trained model which was based on natural images. Starting from the parameters of a pre-trained model is usually better than initiating the model using random values for parameters. The idea is that more abstract characteristics of the images (e.g. edges) have been already trained and the new network learns the more semantic features for the specific task (in this thesis, brain tumours). However, this procedure does not ensure how much the segmentation accuracy is increased. The ideal case is training a network using the actual brain MRI images and use it as a pre-trained model for initialising a new network.

The reason for promising results for the complete tumour was the hyperintensity of the tumour tissues, especially in FLAIR and T2-weighted. However, for more complex structures (such as non-enhancing tumour that appears as hypointensity in T1-contrast), the algorithm had failed for some cases. Therefore, the non-enhancing tumour may be labelled mistakenly as oedema. This will result in poor segmentation of the tumour core. An example is the case HG-0307 in Table 6-4 that the DSC and PPV for core obtained 0.53 and 0.38, respectively, and shows the segmentation result for the case HG-0307 using FCN_Texton_RF method. As can be seen, the method failed to segment the upper part of necrosis which appears hypointense in both FLAIR and T1-contast image. Therefore, the segmentation results are poor for the tumour core. However, the segmentation accuracy is high for the complete tumour.

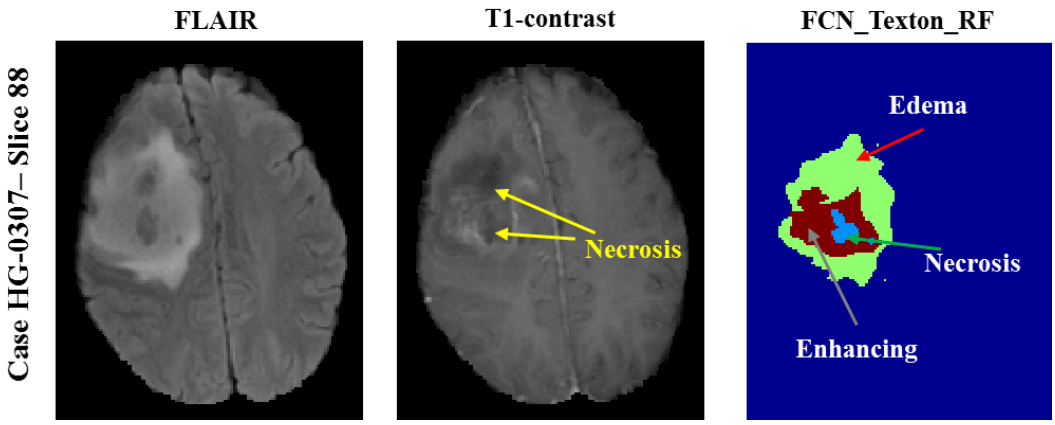



Figure 6-23 An example of failure of the proposed FCN_Texton_RF method for segmentation of tumour core in the case HG-0307 of BRATS 2013 challenge dataset.

## 6.6 Conclusion

In this chapter, a hybrid learning based automatic method was proposed for segmentation of brain tumour in MRI images. A new pipeline of the texton-based feature descriptor derived from the convolutional feature was proposed. The machine-learned features extracted using the FCN were used alongside with hand-designed texton-based features and applied to the RF classifier. For the CNN, it aimed to train a set of filter kernels (weights and bias) with a focus on the design of a network while for the texton, the focus was to design a specific filter to extract the representative features. The architecture of the FCN, containing VGG16 (Simonyan and Zisserman, 2014) and a deconvolutional layers (Long *et al.*, 2015), was proven to have strong efficiency for supervised learning. The score map with upsampling predictions for each pixel was used as a feature map and is learned from multimodal MRI training dataset using the FCN. The learned features were then applied to random forests to classify each MRI image pixel into normal brain tissues and different parts of tumour. The method was evaluated on BRATS 2013 (provided by VSD) and BRATS 2017 (provided by CBICA) datasets. The segmentation labels for the tumour were: necrosis, oedema, non-enhancing and enhancing tumour. FCN_Texton_RF was tested on an independent challenge dataset by uploading the results on VSD system and getting the evaluation results from the website. The results of this work have been uploaded on arXiv (Soltaninejad *et al.*, 2017). The results showed that the application of the random forest classifier to machine-learned features based on FCN and textons provides promising segmentations. The Dice scores for automatic brain tumour segmentation against ground truth for BRATS 2013 experiment were 0.88, 080 and 0.73 for complete tumour, tumour core and enhancing tumour, respectively. The corresponding results for BRATS 2017 experiment were 0.86, 0.78 and 0.66, respectively. Overall, the experimental results suggested that FCN_Texton_RF achieved promising results in the segmentation of brain tumour structures. Adding hand-designed texton features from different protocols to the system increased the classification accuracy of the pixels and results in more accurate final segmentations.

# Chapter 7

# Conclusions and Future Directions

## 7.1 Conclusions

Medical imaging plays an important role in clinical procedures related to cancer diagnosis and treatment selection. A variety of medical imaging modalities have been investigated for clinical tasks. MRI is one of the most popular acquisition modalities and is widely used in brain tumour analysis, since it is non-invasive and provides more information about different tissues. MRI also can be acquired with different procedures and parameters to provide specific information of tissue properties. The clinical background of brain tumour segmentation in MRI was provided in Chapter 2. The procedure of different conventional and DTI MRI modalities was explained. Chapter 2 also explained the clinical dataset that was used in this thesis, and the publicly available BRATS dataset used for the competition and comparison of the research in this field.

When different MR modalities are acquired with the high-resolution technologies, it produces a large number of data. Therefore, using computer-aided automatic methods becomes important and beneficial. The result of computer-based brain tumour image analysis will be used in clinical tasks, such as diagnosis, treatment planning and prognosis. One of the most important analyse of MR images is segmentation, which is a prerequisite for further clinical analysis. Much research has been conducted in recent decades in the field of brain tumour segmentation in MRI images and are partly reviewed in Chapter 3. This field is very broad so it is difficult to review all the methods. In this thesis, Chapter 3 provided a literature review and comparison of some state-of-the-art methods, with the focus on the methods involved in the MICCAI-BRATS grand challenge to obtain an insight on the research trend in the field of brain tumour segmentation.

The first objective of tumour segmentation was delineating the complete tumour. The SP_ERT method was investigated in Chapter 4 for detection and accurate segmentation of the complete tumour structure (which encompasses the oedema and tumour core) using a single modality (e.g. FLAIR). The method was mainly based on supervised ERT which is amongst the most powerful classifiers. However, later in chapter 5, RF was used which is faster to perform compared to ERT. To reduce the computational complexity, SP patches were used instead of sliding fixed-size windows. Different varieties of feature including statistical first order intensity-based, texton features, fractal, and curvature features were calculated for the SP. The

most important feature found out to be textons histograms of the superpixels, which were generated from Gabor filters. The single modal SP_ERT method demonstrated high segmentation performance with the DSC of 0.91 for clinical dataset and 0.88 for the BRATS dataset and is comparable to the top-ranked methods on the VSD scoreboard. The results also showed the importance of pre-processing, especially intensity normalisation and histogram matching for MR image processing.

The next main objective was using multiple modalities for detection and segmentation of detailed tumour tissue subtypes. For this task, a unified framework, i.e. SV_RF, was suggested in Chapter 5 to extend the single modality to multimodality, considering 3D connectivity of the voxels both in the SV patch volume calculations from all the modalities and in the feature extraction. The SV_RF method was successful in accurate segmentation of the core and oedema. Incorporating the advanced DTI protocols into the learning process improved the segmentation of the tumour core, e.g. increasing the DSC from 0.67 to 0.78.

Due to less accuracy of supervoxels for smaller tissues, the SV_RF method has limitations for further segmentation of the tumour core (i.e. necrosis, enhancing and non-enhancing tumour). On the other hand, owing to the advances in deep learning algorithms and their application in image segmentation, the CNN-based brain tumour segmentation methods outperformed other methods, very recently. This can be seen in the BRATS challenge scoreboard and comparison charts provided in Chapter 6. In 2016 the method based on a CNN with small kernels (Pereira *et al.*, 2016) outperformed all the methods and has achieved the best performance until now. Two other methods (Havaei *et al.*, 2017; Kamnitsas *et al.*, 2017) based on CNNs provided very competitive performance.

Deep FCN was developed in 2015 (Long *et al.*, 2015) for image segmentation and modified the standard CNN for end-to-end pixel-wise classification. The advantage of FCN is that it can handle images with any size and the information from different levels can be fused into the output image. In Chapter 6, the FCN method was adapted to multimodal MR images and applied to the BRATS 2013 and 2017 datasets. Although the results demonstrated good sensitivity of the method for detecting the general location of the tumour, the segmentations were coarse compared to the ground truth. Down-sampling data in the pooling layers results in losing spatial information so decreasing the local dependencies in the higher resolution levels. This causes coarser final segmentations, and is the most challenging aspect of the deep CNN-based segmentation algorithms.

The FCN architecture has the advantage of representing the feature maps with different resolutions extracted from corresponding layers of the network, and with the same size of the input image. One hypothesis of the thesis was these feature maps with the same size of the

input image can be used directly as machine-learned features for a supervised based segmentation such as random forest. Therefore, these feature maps were combined and fed into a RF classifier, which is the most important contribution of this thesis. The results interestingly show that using RF with those features extracted from different layers of the FCN network provides better DSC compared to the FCN alone. This means the feature maps include meaningful information from different hierarchy levels of the image resolutions. However, the coarse results show that the segmentations still were not accurate at full resolution. To address this limitation, the next hypothesis was using feature maps that consider more local voxel dependencies and its neighbourhood system. The experiments from Chapters 4 and 5 show that texton maps calculated based on Gabor filters provide useful descriptors of local dependencies. Therefore, the two types of feature vectors from texton histograms and FCN were combined. The results show that this combination benefits from the advantages of both types of features. The segmentation results show that the new integrated method provides the best performance in terms of DSC overlap measure while preserving balance between sensitivity and PPV. This provided the best results among all the algorithms in the thesis. The FCN_Texton_RF method was ranked 6$^{th}$ amongst the state-of-the-art methods in the VSD scoreboard table; whilst it obtains the 1$^{st}$ rank for the segmentation of the complete tumour in terms of DSC on 27$^{th}$ July 2017. It should be noted that the ranking of the FCN_Texton_RF method is tied, as other methods have the same Dice score and ranking for the whole tumour. The screenshot of the VSD scoreboard for the corresponding results are provided in Appendix 2.

The FCN_Texton_RF is a multi-object classifier that uses a small feature vector. The method makes use of the advantages of both machine-learned FCN-based, and hand-designed texton features to produce more accurate segmentation. The FCN features accurately detect and localise tumour regions with fewer false positives. Whilst, the hand-designed texton features accurately segment the tumour parts at the pixel level and provide fine segmentation outputs.

## 7.2 Contributions

The main contributions of the thesis based on the aims and objectives can be summarised as follows

- A fully automated method was proposed for detection and segmentation of the abnormal tissue associated with brain tumours as defined by the T2 hyperintensity from FLAIR MRI, which are routinely acquired as part of standard clinical diagnostic procedure related to brain tumours.

- The histogram of texton descriptors, calculated using a set of Gabor filters with different sizes and orientations, was proposed to provide improved performance for classification of brain tumour superpixels/voxels. Since superpixels/supervoxels are limited to clusters of similar intensities within each MRI modality, using the distribution of local textures inside each superpixels/supervoxel improves further classification. Texton has demonstrated its advantages of providing significant information to distinguish various patterns in both 2D and 3D spaces.
- A novel approach was proposed to form the supervoxel using multimodal MRI, including FLAIR, T1-weighted (with contrast), T2-weighted, *p* and *q* diffusion maps. Unlike existing methods (Wu *et al.*, 2014) in which a supervoxel is calculated from a single MRI protocol, in this thesis, information from multimodal images is combined to produce supervoxel boundaries across multiple image protocols.
- DTI modalities, e.g. the isotropic (*p*) and anisotropic (*q*) diffusion components were incorporated into learning process, to provide parameters that relate to the microscopic movement of molecules. Whilst, most of the existing studies on brain tumour segmentation have been performed on conventional MRI protocols only (i.e. FLAIR, T1-weighted (with contrast) and T2-weighted). Combining DTI and C-MRI provided quantitative features that increased the classification accuracy and improved segmentation results for the tumour core.
- A unified framework was proposed to address the limitations of the FCN (e.g. local dependencies are not sufficiently taken into account) by complementarily integrating hand-designed features with machine-learned features. The hand-designed features from shallow network (with designable filters) encode the prior-knowledge and context while the machine learned features from deep network (with trainable filters) learn the intrinsic features.
- The FCN_Texton_RF method was evaluated using the BRATS 2013 grand challenge dataset and demonstrated promising results that were competitive with the most powerful state-of-the-art methods. The method was ranked 6$^{th}$ in total and 1$^{st}$ for the complete tumour DSC on 27$^{th}$ July 2017.

## 7.3 Limitations

The methods based on superpixel or supervoxels are limited for the segmentation of small objects. For example, if a group of a few pixels are surrounded by a region with a different class, then they may be ignored and labelled incorrectly. This can be further highlighted in the supervoxel algorithm, as they involve larger groups of voxels.

Another limitation of the superpixel/supervoxel method is related to the regions with heterogeneous intensities that include pixels/voxels with several labels. Any classification method may confuse these superpixels/supervoxels and, even if the majority vote is allocated to that segment, there might still be pixels/voxels that will be classified incorrectly.

In this thesis, the labelling procedure for the superpixels/supervoxels was done using a threshold of 50% as the majority vote. Although this procedure seems to be a good criterion for binary classification (Chapter 4), it is not sufficient for multi-class problems (e.g. Chapter 5) The partitions in which the class pixel/voxel percentage is less than the threshold will be excluded from the training stage.

The FCN_Texton_RF provided promising results which were comparable with the state-of-the-art. However, the texton feature extraction phase requires parameter tuning and engineering for designing optimal features. This phase was designed to overcome the limitation of the FCN in decreasing considerably the high resolution dependencies in semantic segmentation. The state-of-the art methods such as U-Net can cope with this limitation by taking the information from the skip layers into account.

The RF phase of the FCN_Texton_RF method causes the whole algorithm to become two separate learning processes. This means two independent training processes are required. The results of the training stage for RF are also dependent on the first FCN stage.

## 7.4 Future Directions

The methods SP_ERT, SV_RF and FCN_Texton_RF have been evaluated on the publicly available BRATS 2013 and 2017 datasets, and achieved good performances compared to the state-of-the-art methods. The potential directions of the future work in both clinical and technical perspectives are summarised below.

### 7.4.1 Superpixel /supervoxel

The patch-based segmentation methods, SP_ERT and SV_RF, achieved good accuracy for the complete tumour and the tumour core segmentation, respectively. However, the supervoxel based segmentation accuracy decreases for tumours with smaller size. Using superpixel/supervoxel methods with more parameters gives more degrees of freedom such as multi-resolution superpixels may tackle this problem. It is suggested that for the heterogeneous regions, the superpixels are divided to smaller patches. The SP/SV algorithms in SP_ERT and SV_RF methods have been designed to be initialised from grids with similar window sizes. In

the case of multi-resolution SP/SV, the more heterogeneous regions can be gridded with the smaller initial size to produce smaller patches. However, this may increase the complexity of the model, as the distance equations and optimisations should be adopted for the new multi-resolution grid scheme.

Another solution for tackling the problem of small ROIs is utilising an additional stage for further segmentation of the regions that are partially included in the SP/SV. This can be achieved by splitting the supervoxels with non-homogenous intensities that are probably related to more than one classe, into sub-regions/sub-volumes and relabel them to obtain more accurate segmentation.

Instead of defining a threshold (e.g. 50%) for labelling a whole SP/SV, using a majority voting procedure might be more accurate. In future work, it is suggested that after counting the number of pixels/voxels in one partition, the assigned class should be the one with the highest number of pixels/voxels.

### 7.4.2 Feature Extraction and Parameter Optimisation

Feature extraction can be further investigated especially in the case of brain tumours as they have diverse appearance in MR images regarding to the tumour type and grade. In future work, it is recommended to take a closer look into the feature extraction and analyse features more deeply to determine their relevance and meaningfulness. Integrating those features to the machine-learned feature may provide more accurate segmentation. It is suggested that providing a comprehensive comparison between the features to make a better understanding of the hand-designed and machine-learned features may result in better optimisation of the feature extraction methods and lead to more accurate segmentation.

Some parameters of the SP/SV algorithms, e.g. the size of superpixel, and hand-designed features, e.g. Gabor filter kernel parameters were obtained through exhaustive parametric searching during the training stage on the selected data. Other optimisation algorithms such as Genetic Algorithm can be explored to effectively find the optimum parameters. The parameter selection may also depend on the dataset. Another future direction should be increasing the generalisation of this process so that the parameter selection will be independent from the acquisition parameters of the images.

### 7.4.3 Further Developing Deep Learning

A deep neural network provides promising results for medical image segmentation. However, there are limitations with application to dense pixel classification for image segmentation. The recent methods tried to tackle the limitations of CNN-based methods by integrating the CRF to increase the accuracy as suggested by Kamnitsas *et al.* (Kamnitsas *et al.*, 2017). Recently, Zheng *et al.* (Zheng *et al.*, 2015) formulated the CRF and merged it into the ResNet in a way that it can be trainable with regular gradient descent. End-to-end training of this CRF can be a future direction for pixel-wise semantic segmentation which can be integrated into coarse segmentation of FCN only method.

Another future direction can be looking into other state-of-the-art CNN approaches such as U-Nets and ResNets to investigate machine-learned data driven features. A large number of feature maps are produced via those deep networks. There may be lots of interpretations that can be obtained from the networks. Investigating and interpreting that information may provide understanding of how the networks works and why they fail and how they can be improved. An example can be implementing the algorithm on the multicentre dataset with different acquisition parameters and suggesting a general framework so that the CNN is capable of handling a large-scale multi-centre dataset.

The current deep learning methods applied for medical image analysis use a black box scheme trained based on the labelled images, regardless of the nature of those images. The method in this study used the pre-trained VGG16, which was trained based on natural images to optimise the weights and make the further learning process easier. Many traditional model-based segmentation techniques have utilised the clinical information from the medical images as a prior knowledge. It would be a good idea to use that clinical information of the tissue as the prior and feed them to the networks' feature space. Therefore, the optimisation of the network can be guided based on prior clinical information instead of using pretrained networks.

The current FCN model was trained using a pre-trained model based on natural images. Two approaches can be considered to improve the segmentation results from the FCN. The current FCN can be fine-tuned for the clinical or public datasets. The training process of the FCN can also be performed on a pre-trained model based on MRI brain images.

### 7.4.4 Future Clinical Directions

The clinical future directions are summarised below:

- **Clinical Prognosis**

The future MICCAI challenge (2017) intends to step forward and add the prognosis task to the segmentation challenge. This will be estimation of the survival rate of the patient with brain tumour cancer. One suggestion for future direction will be using the segmented regions of the tumour to provide the tumour related characteristics and information for prognosis. Another future direction could be using the most representative features alongside with other measures that are obtained from the segmented regions (e.g. tumour tissue subtypes volumes, texture, shape, texton histogram, etc.) to predict the grade of tumour and estimate the survival time. A research work (Soltaninejad *et al.*, 2014) was presented in MIUA conference on grading the tumour using the methods which were introduced in this thesis. This work can be expanded and developed by integrating state-of-the-art features to provide better classification of the tumour for prognosis tasks.

- **Evaluation of FCN-based Method on Clinical Dataset**

Very few completely automatic segmentation algorithms have been adopted in the clinic. Recently only one automated tool has been clinically evaluated (Meier *et al.*, 2016). They adopted the method which is called BraTumIA (Brain Tumour Image Analysis) (Meier *et al.*, 2014b; Porz *et al.*, 2014) in the clinic by investigating it for longitudinal brain tumour volumetry using their own clinical dataset and ground truth. In this thesis, the works in Chapters 4 and 5 were evaluated on the clinical datasets. However, due to insufficient clinical multimodal dataset for training the FCN, the method in chapter 6 was evaluated only on the publicly available BRATS dataset. By acquiring more clinical MRI dataset, the method can be clinically evaluated in the future.

- **Extending the Clinical Cases**

The FCN based method requires a large number of data for efficient training. The reason that DTI modalities were excluded from FCN_Texton_RF experiments in Chapter 6 was the lack of sufficient dataset. However, the results in Chapter 5 suggested that using DTI *p*- and *q*-maps improve the segmentation of the tumour core with the SV_RF method which means better feature descriptors for the tumour core were provided. Providing more datasets with DTI modalities may improve the segmentation of the tumour parts, such as enhancing, which still has the least accuracy in the current BRATS competitions.

Furthermore, the current clinical dataset mainly contains general cases, such as different tumour grades from a wide range of patient ages (patient ages at the time of scanning ranged from 22 to 73). In the future, more complicated cases such as calcification, intratumoural bleeding, or elderly patients with white matter diseases, which are clinically important to distinguish against, need to be further investigated.

# Appendix 1

# List of Publications

Journal Papers

- M. Soltaninejad, G. Yang, N. Allinson, T. Lambrou, 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," *International Journal of Computer Assisted Radiology and Surgery*, Vol.12 pp. 183-203, Feb 2017.
- M. Soltaninejad, G. Yang, N. Allinson, T. Lambrou, T. L. Jones, T. R. Barrick, F. A. Howe, X. Ye "Supervised Learning based Multimodal MRI Brain Tumour Segmentation using Texture Features from Supervoxels," submitted to *Journal of Computers in Biology and Medicine*, 2017, (under revision).

Magazine

- M. Soltaninejad, X. Ye, G. Yang, N. Allinson, T. Lambrou, "An Image Analysis Approach to MRI Brain Tumour Grading," *Oncology News*, Vol. 9 issue 6, pp. 204-207, Jan 2015.

Conference Papers

- M. Soltaninejad, X. Ye, G. Yang, N. Allinson, T. Lambrou, "Brain Tumour Grading in Different MRI Protocols using SVM on Statistical Features," *Medical Image Understanding and Analysis*, 9$^{th}$ – 11$^{th}$ July 2014, London, UK.
- M. Soltaninejad, T. Lambrou, A. Qureshi, N. Allinson, X. Ye, "A Hybrid Method for Haemorrhage Segmentation in Trauma Brain CT," *Medical Image Understanding and Analysis*, 9$^{th}$ – 11$^{th}$ July 2014, London, UK.

# Appendix 2

## VSD Scoreboard Screenshot

### Evaluation Results: Challenge

#### Patient

| Position | User | Dice | | | Positive Predictive Value | | | Sensitivity | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | complete | core | enhancing | complete | core | enhancing | complete | core | enhancing |
| 1 | peres1 | 0.88 (4) | 0.83 (1) | 0.77 (2) | 0.88 (12) | 0.87 (5) | 0.74 (22) | 0.89 (26) | 0.83 (9) | 0.81 (7) |
| 2 | xiaoz2 | 0.88 (8) | 0.83 (3) | 0.77 (1) | 0.88 (8) | 0.83 (17) | 0.77 (15) | 0.88 (29) | 0.84 (8) | 0.79 (10) |
| 3 | shenh1 | 0.88 (2) | 0.83 (4) | 0.75 (5) | 0.86 (16) | 0.85 (12) | 0.70 (34) | 0.91 (16) | 0.85 (7) | 0.83 (4) |
| 4 | kuant1 | 0.87 (10) | 0.80 (8) | 0.76 (4) | 0.91 (5) | 0.88 (3) | 0.77 (13) | 0.83 (42) | 0.77 (25) | 0.76 (19) |
| 5 | jessa1 | 0.87 (12) | 0.82 (5) | 0.76 (3) | 0.82 (32) | 0.81 (23) | 0.67 (41) | 0.93 (11) | 0.86 (4) | 0.90 (1) |
| 6 | soltm1 | 0.88 (1) | 0.80 (9) | 0.73 (21) | 0.88 (10) | 0.87 (6) | 0.80 (5) | 0.89 (25) | 0.77 (24) | 0.70 (34) |
| 7 | kwond1 | 0.88 (6) | 0.83 (2) | 0.72 (23) | 0.92 (4) | 0.90 (2) | 0.74 (23) | 0.84 (38) | 0.78 (19) | 0.72 (30) |
| 8 | alexv1 | 0.86 (19) | 0.78 (15) | 0.74 (7) | 0.81 (36) | 0.77 (32) | 0.79 (9) | 0.93 (9) | 0.82 (12) | 0.73 (25) |
| 9 | kayab1 | 0.87 (9) | 0.78 (11) | 0.71 (28) | 0.82 (33) | 0.80 (25) | 0.83 (1) | 0.95 (2) | 0.80 (14) | 0.67 (41) |
| 10 | tustn1 | 0.87 (16) | 0.78 (14) | 0.74 (8) | 0.85 (21) | 0.74 (41) | 0.69 (36) | 0.89 (23) | 0.88 (2) | 0.83 (5) |
| 11 | fidol1 | 0.88 (7) | 0.77 (19) | 0.72 (22) | 0.83 (28) | 0.77 (34) | 0.81 (3) | 0.93 (4) | 0.82 (11) | 0.69 (39) |
| 12 | havam2 | 0.88 (5) | 0.75 (26) | 0.74 (9) | 0.88 (11) | 0.77 (33) | 0.78 (10) | 0.88 (27) | 0.78 (21) | 0.72 (28) |

Figure Appendix 2-4 Screenshot of the overall ranking on 27$^{h}$ Jun 2017. The corresponding row for the FCN_Texton_RF is highlighted by VSD. The username "**soltm1**" is assigned automatically by VSD, abbreviated from the name of the author, i.e. M. Soltaninejad. The ranking is determined by the VSD system based on taking into account all the evaluation measurements. It should be noted that the ranking of the FCN_Texton_RF method is tied, as other methods have the same Dice score and ranking for the whole tumour.

# Reference List


3D Slicer [WWW Document], 2017. URL https://www.slicer.org/ (accessed 5.10.17).

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

Amiri, S., Rekik, I., Mahjoub, M.A., 2016. Deep random forest-based learning transfer to SVM for brain tumor segmentation, in: 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). Presented at the 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), pp. 297–302. doi:10.1109/ATSIP.2016.7523095

Angelini, E.D., Clatz, O., Mandonnet, E., Konukoglu, E., Capelle, L., Duffau, H., 2007. Glioma Dynamics and Computational Models: A Review of Segmentation, Registration, and In Silico Growth Algorithms and their Clinical Applications. Current Medical Imaging Reviews 3, 262–276.

Arbelaez, P., Maire, M., Fowlkes, C., Malik, J., 2011. Contour Detection and Hierarchical Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 898–916. doi:10.1109/TPAMI.2010.161

Arridge, S.R., n.d. A note on image and curvature [WWW Document]. URL http://www0.cs.ucl.ac.uk/staff/S.Arridge/teaching/ndsp/curvature.pdf

Aslian, H., Sadeghi, M., Mahdavi, S.R., Babapour Mofrad, F., Astarakee, M., Khaledi, N., Fadavi, P., 2013. Magnetic resonance imaging-based target volume delineation in radiation therapy treatment planning for brain tumors using localized region-based active contour. Int. J. Radiat. Oncol. Biol. Phys. 87, 195–201. doi:10.1016/j.ijrobp.2013.04.049

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C., 2017a. Segmentation Labels for the Pre-operative Scans of the TCGA-GBM collection. Segmentation Labels for the Pre-operative Scans of the TCGA-GBM collection. doi:https://doi.org/10.7937/k9/tcia.2017.klxwjj1q

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C., 2017b. Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. doi:https://doi.org/10.7937/k9/tcia.2017.klxwjj1q

Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Rozycki M, Freymann J, Farahani K, Davatzikos C, 2017c. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features.

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

Bauer, S., Fejes, T., Slotboom, J., Wiest, R., Nolte, L.-P., Reyes, M., 2012. Segmentation of brain tumor images based on integrated hierarchical classification and regularization, in: MICCAI BraTS Workshop. Nice: Miccai Society.

Bauer, S., Nolte, L.-P., Reyes, M., 2011. Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011. Presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Berlin, Heidelberg, pp. 354–361. doi:10.1007/978-3-642-23626-6_44

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

Bengio, Y., Courville, A., Vincent, P., 2013. Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 1798–1828. doi:10.1109/TPAMI.2013.50

Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences 10, 191–203. doi:10.1016/0098-3004(84)90020-7

BRATS :: The Virtual Skeleton Database Project [WWW Document], n.d. URL https://www.smir.ch/BRATS/Start2012 (accessed 7.3.16).

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. doi:10.1023/A:1010933404324

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees, 1 edition. ed. Chapman and Hall/CRC, Boca Raton.

Cai, H., Verma, R., Ou, Y., Lee, S. k, Melhem, E.R., Davatzikos, C., 2007. PROBABILISTIC SEGMENTATION OF BRAIN TUMORS BASED ON MULTI-MODALITY MAGNETIC RESONANCE IMAGES, in: 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Presented at the 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 600–603. doi:10.1109/ISBI.2007.356923

Cancer registration statistics, England Statistical bulletins - Office for National Statistics [WWW Document], n.d. URL https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/previousReleases (accessed 5.10.17).

Cancer Research UK [WWW Document], n.d. . Cancer Research UK. URL https://www.cancerresearchuk.org/home (accessed 5.10.17).

Canny, J., 1986. A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8, 679–698. doi:10.1109/TPAMI.1986.4767851

Capelle, A.S., Alata, O., Fernandez, C., Lefevre, S., Ferrie, J.C., 2000. Unsupervised segmentation for automatic detection of brain tumors in MRI, in: Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101). Presented at the Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101), pp. 613–616 vol.1. doi:10.1109/ICIP.2000.901033

Cavaliere, R., Lopes, M.B.S., Schiff, D., 2005. Low-grade gliomas: an update on pathology and therapy. The Lancet Neurology 4, 760–770. doi:10.1016/S1474-4422(05)70222-2

Cawley, G.C., Talbot, N.L.C., 2006. Gene selection in cancer classification using sparse logistic regression with Bayesian regularization. Bioinformatics 22, 2348–2355. doi:10.1093/bioinformatics/btl386

Chinot, O.L., Macdonald, D.R., Abrey, L.E., Zahlmann, G., Kerloëguen, Y., Cloughesy, T.F., 2013. Response Assessment Criteria for Glioblastoma: Practical Adaptation and Implementation in Clinical Trials of Antiangiogenic Therapy. Curr. Neurol. Neurosci. Rep. 13. doi:10.1007/s11910-013-0347-2

Cho, K.H., Choi, J.H., Kim, J.Y., Lee, S.H., Yoo, H., Shin, K.H., Kim, T.H., Moon, S.H., Lee, S.H., Park, H.C., 2012. Volumetric response evaluation after intensity modulated radiotherapy in patients with supratentorial gliomas. Technol. Cancer Res. Treat. 11, 41–48.

Choubey, M., Agrawal, S., 2012. A fully automatic approach to detect brain cancer using random walk algorithm. Int J Comput Technol Appl 3, 265–268.

Clark, M.C., Hall, L.O., Goldgof, D.B., Velthuizen, R., Murtagh, F.R., Silbiger, M.S., 1998. Automatic tumor segmentation using knowledge-based techniques. IEEE Transactions on Medical Imaging 17, 187–201. doi:10.1109/42.700731

Cobzas, D., Birkbeck, N., Schmidt, M., Jagersand, M., Murtha, A., 2007. 3D Variational Brain Tumor Segmentation using a High Dimensional Feature Set, in: 2007 IEEE 11th

International Conference on Computer Vision. Presented at the 2007 IEEE 11th International Conference on Computer Vision, pp. 1–8. doi:10.1109/ICCV.2007.4409130

Colah, C., 2017. Neural Networks, Manifolds, and Topology -- colah's blog [WWW Document]. URL http://colah.github.io/posts/2014-03-NN-Manifolds-Topology/ (accessed 5.4.17).

Cordier, N., Menze, B., Delingette, H., Ayache, N., 2013. Patch-based Segmentation of Brain Tissues, in: Menze, B., Reyes, M., Jakab, A., Gerstner, E., Kirby, J., Farahani, K. (Eds.), MICCAI Challenge on Multimodal Brain Tumor Segmentation, Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) 2013. IEEE, Nagoya, Japan, pp. 6–17.

Costa, A.F., Humpire-Mamani, G., Traina, A.J.M., 2012. An Efficient Algorithm for Fractal Analysis of Textures, in: 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). Presented at the 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 39–46. doi:10.1109/SIBGRAPI.2012.15

Criminisi, A., Shotton, J., Konukoglu, E., 2012. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. Found. Trends. Comput. Graph. Vis. 7, 81–227. doi:10.1561/0600000035

CS231n Convolutional Neural Networks for Visual Recognition [WWW Document], 2017. URL http://cs231n.github.io/convolutional-networks/ (accessed 5.4.17).

Davy, A., Havaei, M., Warde-Farley, D., Biard, A., Tran, L., Jodoin, P.-M., Courville, A., Larochelle, H., Pal, C., Bengio, Y., 2014. Brain tumor segmentation with deep neural networks, in: Proceeding of BRATS-MICCAI. Presented at the Proceeding of BRATS-MICCAI.

Dice, L.R., 1945. Measures of the Amount of Ecologic Association Between Species. Ecology 26, 297–302. doi:10.2307/1932409

Dong, H., Yang, G., Liu, F., Mo, Y., Guo, Y., 2017. Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks, in: Medical Image Understanding and Analysis, Communications in Computer and Information Science. Presented at the Annual Conference on Medical Image Understanding and Analysis, Springer, Cham, pp. 506–517. doi:10.1007/978-3-319-60964-5_44

Doyle, S., Vasseur, F., Dojat, M., Forbes, F., 2013. Fully automatic brain tumor segmentation from multiple mr sequences using hidden markov fields and variational EM, in: Proceedings of NCI-MICCAI BRATS 2013. pp. 18–22.

Dvořák, P., Menze, B., 2015. Local Structure Prediction with Convolutional Neural Networks for Multimodal Brain Tumor Segmentation, in: Medical Computer Vision: Algorithms for Big Data. Presented at the International MICCAI Workshop on Medical Computer Vision, Springer, Cham, pp. 59–71. doi:10.1007/978-3-319-42016-5_6

Egger, J., Kapur, T., Fedorov, A., Pieper, S., Miller, J.V., Veeraraghavan, H., Freisleben, B., Golby, A.J., Nimsky, C., Kikinis, R., 2013. GBM Volumetry using the 3D Slicer Medical Image Computing Platform. Sci. Rep. 3. doi:10.1038/srep01364

Eisele, S.C., Wen, P.Y., Lee, E.Q., 2016. Assessment of Brain Tumor Response: RANO and Its Offspring. Curr Treat Options Oncol 17, 35. doi:10.1007/s11864-016-0413-5

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

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

Fletcher-Heath, L.M., Hall, L.O., Goldgof, D.B., Murtagh, F.R., 2001. Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. Artificial Intelligence in Medicine, Fuzzy Theory in Medicine 21, 43–63. doi:10.1016/S0933-3657(00)00073-7

FSL [WWW Document], n.d. URL https://fsl.fmrib.ox.ac.uk/fsl/fslwiki (accessed 2.28.17).

Fulkerson, B., Vedaldi, A., Soatto, S., 2009. Class segmentation and object localization with superpixel neighborhoods, in: 2009 IEEE 12th International Conference on Computer Vision. Presented at the 2009 IEEE 12th International Conference on Computer Vision, pp. 670–677. doi:10.1109/ICCV.2009.5459175

Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M., Haussler, D., 2000. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 16, 906–914. doi:10.1093/bioinformatics/16.10.906

Geremia, E., Menze, B.H., Ayache, N., 2013. Spatially Adaptive Random Forests, in: 2013 IEEE 10th International Symposium on Biomedical Imaging. Presented at the 2013 IEEE 10th International Symposium on Biomedical Imaging, pp. 1344–1347. doi:10.1109/ISBI.2013.6556781

Geremia, E., Menze, B.H., Prastawa, M., Weber, M.-A., Criminisi, A., Ayache, N., 2012. Brain Tumor Cell Density Estimation from Multi-modal MR Images Based on a Synthetic Tumor Growth Model, in: Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging. Presented at the International MICCAI Workshop on Medical Computer Vision, Springer, Berlin, Heidelberg, pp. 273–282. doi:10.1007/978-3-642-36620-8_27

Gering, D.T., Grimson, W.E.L., Kikinis, R., 2002. Recognizing Deviations from Normalcy for Brain Tumor Segmentation, in: Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002. Presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Berlin, Heidelberg, pp. 388–395. doi:10.1007/3-540-45786-0_48

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach Learn 63, 3–42. doi:10.1007/s10994-006-6226-1

Ginat, D.T., Meyers, S.P., 2012. Intracranial Lesions with High Signal Intensity on T1-weighted MR Images: Differential Diagnosis. RadioGraphics 32, 499–516. doi:10.1148/rg.322105761

Gonzalez, R.C., Woods, R.E., 2007. Digital Image Processing, 3 edition. ed. Pearson, Upper Saddle River, N.J.

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

Gotz, M., Weber, C., Blocher, J., Stieltjes, B., Meinzer, H., Maier-Hein, K., 2014. Extremely randomized trees based brain tumor segmentation, in: Proceeding of BRATS Challenge-MICCAI. Presented at the Proceeding of BRATS Challenge-MICCAI, pp. 006-011.

Grigorescu, S.E., Petkov, N., Kruizinga, P., 2002. Comparison of texture features based on Gabor filters. IEEE Transactions on Image Processing 11, 1160–1167. doi:10.1109/TIP.2002.804262

Hamamci, A., Kucuk, N., Karaman, K., Engin, K., Unal, G., 2012. Tumor-Cut: Segmentation of Brain Tumors on Contrast Enhanced MR Images for Radiosurgery Applications. IEEE Transactions on Medical Imaging 31, 790–804. doi:10.1109/TMI.2011.2181857

Hamon, J., Dhaenens, C., Even, G., Jacques, J., 2013. Feature selection in high dimensional regression problems for genomic, in: Tenth International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics. Nice, France.

Hariharan, B., Arbeláez, P., Girshick, R., Malik, J., 2014. Simultaneous Detection and Segmentation, in: Computer Vision – ECCV 2014. Presented at the European

Conference on Computer Vision, Springer, Cham, pp. 297–312. doi:10.1007/978-3-319-10584-0_20
Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28, 100–108. doi:10.2307/2346830
Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H., 2017. Brain tumor segmentation with Deep Neural Networks. Medical Image Analysis 35, 18–31. doi:10.1016/j.media.2016.05.004
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. doi:10.1109/CVPR.2016.90
Helen, R., Kamaraj, N., 2015. CAD scheme to detect brain tumour in MR images using active contour models and tree classifiers. Journal of Electrical Engineering and Technology 10, 670–675. doi:10.5370/JEET.2015.10.2.670
Henriksen, J.J., 2007. 3D surface tracking and approximation using Gabor filters (Tech. Rep.). South Denmark University.
Hsieh, T.M., Liu, Y.-M., Liao, C.-C., Xiao, F., Chiang, I.-J., Wong, J.-M., 2011. Automatic segmentation of meningioma from non-contrasted brain MRI integrating fuzzy clustering and region growing. BMC Medical Informatics and Decision Making 11, 54. doi:10.1186/1472-6947-11-54
Huisman, T.A.G.M., 2010. Diffusion-weighted and diffusion tensor imaging of the brain, made easy. Cancer Imaging 10, S163–S171. doi:10.1102/1470-7330.2010.9023
Inversion recovery [WWW Document], n.d. . Questions and Answers in MRI. URL http://mriquestions.com/what-is-ir.html (accessed 9.4.17).
Itakura, H., Achrol, A.S., Mitchell, L.A., Loya, J.J., Liu, T., Westbroek, E.M., Feroze, A.H., Rodriguez, S., Echegaray, S., Azad, T.D., Yeom, K.W., Napel, S., Rubin, D.L., Chang, S.D., Harsh, G.R., Gevaert, O., 2015. Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. Sci Transl Med 7, 303ra138. doi:10.1126/scitranslmed.aaa7582
Jain, A.K., 1989. Fundamentals of Digital Image Processing. Prentice Hall.
Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage 62, 782–790. doi:10.1016/j.neuroimage.2011.09.015
Jenkinson, M.D., Du Plessis, D.G., Walker, C., Smith, T.S., 2007. Advanced MRI in the management of adult gliomas. Br J Neurosurg 21, 550–561. doi:10.1080/02688690701642020
Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional Architecture for Fast Feature Embedding, in: Proceedings of the 22Nd ACM International Conference on Multimedia, MM '14. ACM, New York, NY, USA, pp. 675–678. doi:10.1145/2647868.2654889
Jones, T., Bell, B., Barrick, T., 2012. 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). p. 188.
Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis 36, 61–78. doi:10.1016/j.media.2016.10.004
Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., Büchler, P., 2013. The virtual skeleton database: an open access repository for biomedical research and collaboration. J. Med. Internet Res. 15, e245. doi:10.2196/jmir.2930
Kong, J., Wang, J., Lu, Y., Zhang, J., Li, Y., Zhang, B., 2006. A novel approach for segmentation of MRI brain images. IEEE Mediterranean Electrotechnical Conference 525–528.

Kononenko, I., 1994. Estimating attributes: Analysis and extensions of RELIEF, in: Bergadano, F., Raedt, L.D. (Eds.), Machine Learning: ECML-94, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 171–182.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. Presented at the Advances in Neural Information Processing Systems, pp. 1097–1105.

Kwon, D., Shinohara, R.T., Akbari, H., Davatzikos, C., 2014. Combining Generative Models for Multifocal Glioma Segmentation and Registration. Med Image Comput Comput Assist Interv 17, 763–770.

Lafferty, J.D., McCallum, A., Pereira, F.C.N., 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, in: Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 282–289.

Law, M., Young, R.J., Babb, J.S., Peccerelli, N., Chheang, S., Gruber, M.L., Miller, D.C., Golfinos, J.G., Zagzag, D., Johnson, G., 2008. Gliomas: predicting time to progression or survival with cerebral blood volume measurements at dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging. Radiology 247, 490–498. doi:10.1148/radiol.2472070898

Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D., 1997. Face recognition: a convolutional neural-network approach. IEEE Transactions on Neural Networks 8, 98–113. doi:10.1109/72.554195

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. doi:10.1038/nature14539

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86, 2278–2324. doi:10.1109/5.726791

Lee, C.-H., Schmidt, M., Murtha, A., Bistritz, A., Sander, J., Greiner, R., 2005. Segmenting Brain Tumors with Conditional Random Fields and Support Vector Machines, in: Liu, Y., Jiang, T., Zhang, C. (Eds.), Computer Vision for Biomedical Image Applications, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 469–478. doi:10.1007/11569541_47

Lee, C.-H., Wang, S., Murtha, A., Brown, M.R.G., Greiner, R., 2008. Segmenting brain tumors using pseudo-conditional random fields. Med Image Comput Comput Assist Interv 11, 359–366.

Leung, T., Malik, J., 2001. Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons. International Journal of Computer Vision 43, 29–44. doi:10.1023/A:1011126920638

Li, B., Wang, Q., Hu, J., 2011. Feature subset selection: a correlation-based SVM filter approach. IEEJ Trans Elec Electron Eng 6, 173–179. doi:10.1002/tee.20641

Liao, P.-S., Chen, T.-S., Chung, P.-C., 2001. A fast algorithm for multilevel thresholding. J. Inf. Sci. Eng. 17, 713–727.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2, 18–22.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440.

Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., Scheithauer, B.W., Kleihues, P., 2007. The 2007 WHO Classification of Tumours of the Central Nervous System. Acta Neuropathol 114, 97–109. doi:10.1007/s00401-007-0243-4

Louis, D.N., Perry, A., Reifenberger, G., Deimling, A. von, Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., Ellison, D.W., 2016. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol 131, 803–820. doi:10.1007/s00401-016-1545-1

Luan, S., Zhang, B., Chen, C., Cao, X., Han, J., Liu, J., 2017. Gabor Convolutional Networks. arXiv:1705.01450 [cs].

Lyksborg, M., Puonti, O., Agn, M., Larsen, R., 2015. An Ensemble of 2D Convolutional Neural Networks for Tumor Segmentation, in: Image Analysis. Presented at the Scandinavian Conference on Image Analysis, Springer, Cham, pp. 201–211. doi:10.1007/978-3-319-19665-7_17

Madabhushi, A., Udupa, J.K., 2005. Interplay between intensity standardization and inhomogeneity correction in MR image processing. IEEE Transactions on Medical Imaging 24, 561–576. doi:10.1109/TMI.2004.843256

Matan, O., Burges, C.J.C., Cun, Y.L., Denker, J.S., 1992. Multi-Digit Recognition Using A Space Displacement Neural Network, in: Neural Information Processing Systems. Morgan Kaufmann, pp. 488–495.

McRobbie, D.W., Moore, E.A., Graves, M.J., Prince, M.R., 2006. MRI from Picture to Proton. Cambridge university press.

Meier, R., Bauer, S., Slotboom, J., Wiest, R., Reyes, M., 2014a. Patient-specific semi-supervised learning for postoperative brain tumor segmentation. Med Image Comput Comput Assist Interv 17, 714–721.

Meier, R., Bauer, S., Slotboom, J., Wiest, R., Reyes, M., 2014b. Appearance-and context-sensitive features for brain tumor segmentation, in: Proceedings of MICCAI BRATS Challenge. Presented at the Proceedings of MICCAI BRATS Challenge, pp. 020–026.

Meier, R., Knecht, U., Loosli, T., Bauer, S., Slotboom, J., Wiest, R., Reyes, M., 2016. Clinical Evaluation of a Fully-automatic Segmentation Method for Longitudinal Brain Tumor Volumetry. Sci Rep 6, 23376. doi:10.1038/srep23376

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

MICCAI BRATS - The Multimodal Brain Tumor Segmentation Challenge [WWW Document], n.d. URL http://braintumorsegmentation.org/ (accessed 5.10.17).

MICCAI-BraTS 2017 Leaderboard [WWW Document], 2017. URL https://www.cbica.upenn.edu/BraTS17/ (accessed 10.22.17).

MNIST Demos on Yann LeCun's website [WWW Document], n.d. URL http://yann.lecun.com/exdb/lenet/ (accessed 5.10.17).

MRI Signal weighting (T1, T2, PD) and sequences parameters : TR, TE [WWW Document], n.d. . IMAIOS. URL https://www.imaios.com/en/e-Courses/e-MRI/MRI-signal-contrast/Signal-weighting (accessed 9.4.17).

Murphy, K., 2012. Machine Learning: A Probabilistic Perspective. MIT Press, Cambridge, MA.

Ning, F., Delhomme, D., LeCun, Y., Piano, F., Bottou, L., Barbano, P.E., 2005. Toward automatic phenotyping of developing embryos from videos. IEEE Transactions on Image Processing 14, 1360–1371. doi:10.1109/TIP.2005.852470

Niyazi, M., Brada, M., Chalmers, A.J., Combs, S.E., Erridge, S.C., Fiorentino, A., Grosu, A.L., Lagerwaard, F.J., Minniti, G., Mirimanoff, R.-O., Ricardi, U., Short, S.C., Weber,

D.C., Belka, C., 2016. ESTRO-ACROP guideline “target delineation of glioblastomas.” Radiother Oncol 118, 35–42. doi:10.1016/j.radonc.2015.12.003

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

Noh, H., Hong, S., Han, B., 2015. Learning Deconvolution Network for Semantic Segmentation. Presented at the Proceedings of the IEEE International Conference on Computer Vision, pp. 1520–1528.

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

Odland, A., Server, A., Saxhaug, C., Breivik, B., Groote, R., Vardal, J., Larsson, C., Bjørnerud, A., 2015. Volumetric glioma quantification: comparison of manual and semi-automatic tumor segmentation for the quantification of tumor growth. Acta Radiol 56, 1396–1403. doi:10.1177/0284185114554822

Papadopoulos, M.C., Saadoun, S., Binder, D.K., Manley, G.T., Krishna, S., Verkman, A.S., 2004. Molecular mechanisms of brain tumor edema. Neuroscience 129, 1011–1020. doi:10.1016/j.neuroscience.2004.05.044

Patel, M.R., Tse, V., 2004. Diagnosis and staging of brain tumors. Semin Roentgenol 39, 347–360.

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

Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 1226–1238. doi:10.1109/TPAMI.2005.159

Penn Imaging – Home [WWW Document], 2017. URL https://ipp.cbica.upenn.edu/ (accessed 10.22.17).

Pereira, S., Pinto, A., Alves, V., Silva, C.A., 2016. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Transactions on Medical Imaging 35, 1240–1251. doi:10.1109/TMI.2016.2538465

Perry, A., 2003. Pathology of low-grade gliomas: an update of emerging concepts. Neuro Oncol 5, 168–178. doi:10.1215/S1152-8517-02-00044-3

Phillips, W.E., Velthuizen, R.P., Phuphanich, S., Hall, L.O., Clarke, L.P., Silbiger, M.L., 1995. Application of fuzzy c-means segmentation technique for tissue differentiation in MR images of a hemorrhagic glioblastoma multiforme. Magnetic Resonance Imaging 13, 277–290. doi:10.1016/0730-725X(94)00093-I

Pierpaoli, C., Jezzard, P., Basser, P.J., Barnett, A., Di Chiro, G., 1996. Diffusion tensor MR imaging of the human brain. Radiology 201, 637–648. doi:10.1148/radiology.201.3.8939209

Pinto, A., Pereira, S., Correia, H., Oliveira, J., Rasteiro, D.M.L.D., Silva, C.A., 2015. 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). Presented at the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3037–3040. doi:10.1109/EMBC.2015.7319032

Popuri, K., Cobzas, D., Murtha, A., Jägersand, M., 2012. 3D variational brain tumor segmentation using Dirichlet priors on a clustered feature set. Int J CARS 7, 493–506. doi:10.1007/s11548-011-0649-2

Porz, N., Bauer, S., Pica, A., Schucht, P., Beck, J., Verma, R.K., Slotboom, J., Reyes, M., Wiest, R., 2014. Multi-Modal Glioblastoma Segmentation: Man versus Machine. PLOS ONE 9, e96873. doi:10.1371/journal.pone.0096873

Price, S.J., Jena, R., Burnet, N.G., Carpenter, T.A., Pickard, J.D., Gillard, J.H., 2007. Predicting patterns of glioma recurrence using diffusion tensor imaging. Eur Radiol 17, 1675–1684. doi:10.1007/s00330-006-0561-2

Price, S.J., Peña, A., Burnet, N.G., Jena, R., Green, H.A.L., Carpenter, T.A., Pickard, J.D., Gillard, J.H., 2004. Tissue signature characterisation of diffusion tensor abnormalities in cerebral gliomas. Eur Radiol 14, 1909–1917. doi:10.1007/s00330-004-2381-6

Rajendran, A., Dhanasekaran, R., 2012. Fuzzy Clustering and Deformable Model for Tumor Segmentation on MRI Brain Image: A Combined Approach. Procedia Engineering, International Conference on Communication Technology and System Design 2011 30, 327–333. doi:10.1016/j.proeng.2012.01.868

Rao, V., Sharifi Sarabi, M., Jaiswal, A., 2015. Brain Tumor Segmentation with Deep Learning, in: MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS). Presented at the MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS), pp. 56–59.

Rees, J., Watt, H., Jäger, H.R., Benton, C., Tozer, D., Tofts, P., Waldman, A., 2009. Volumes and growth rates of untreated adult low-grade gliomas indicate risk of early malignant transformation. Eur J Radiol 72, 54–64. doi:10.1016/j.ejrad.2008.06.013

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

Roberts, T.P.L., Mikulis, D., 2007. Neuro MR: Principles. J. Magn. Reson. Imaging 26, 823–837. doi:10.1002/jmri.21029

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, pp. 234–241. doi:10.1007/978-3-319-24574-4_28

Rosario, S.F., Thangadurai, K., 2015. RELIEF: Feature Selection Approach. International Journal of Innovative Research and Development 4.

Ruan, S., Lebonvallet, S., Merabet, A., Constans, J. m, 2007. Tumor segmentation from a multispectral MRI images by using support vector machine classification, in: 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Presented at the 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1236–1239. doi:10.1109/ISBI.2007.357082

Ruan, S., Zhang, N., Liao, Q., Zhu, Y., 2011. Image fusion for following-up brain tumor evolution, in: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Presented at the 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 281–284. doi:10.1109/ISBI.2011.5872406

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis 115, 211–252. doi:10.1007/s11263-015-0816-y

Saeys, Y., Inza, I., Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. Bioinformatics 23, 2507–2517. doi:10.1093/bioinformatics/btm344

Sánchez-Maroño, N., Alonso-Betanzos, A., Tombilla-Sanromán, M., 2007. Filter Methods for Feature Selection – A Comparative Study, in: Intelligent Data Engineering and Automated Learning - IDEAL 2007. Presented at the International Conference on Intelligent Data Engineering and Automated Learning, Springer, Berlin, Heidelberg, pp. 178–187. doi:10.1007/978-3-540-77226-2_19

Sauwen, N., Sima, D.M., Van Cauter, S., Veraart, J., Leemans, A., Maes, F., Himmelreich, U., Van Huffel, S., 2015. Hierarchical non-negative matrix factorization to characterize brain tumor heterogeneity using multi-parametric MRI. NMR Biomed 28, 1599–1624. doi:10.1002/nbm.3413

Schölkopf, B., Smola, A.J., 2002. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press.
Schroeder, M.R., 2009. Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise. Dover Publications Inc., Mineola, N.Y.
Scott, J.N., Brasher, P.M.A., Sevick, R.J., Rewcastle, N.B., Forsyth, P.A., 2002. How often are nonenhancing supratentorial gliomas malignant? A population study. Neurology 59, 947–949.
Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., 2013. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. arXiv:1312.6229 [cs].
Shanthi, K.J., Kumar, M.S., 2007. Skull stripping and automatic segmentation of brain MRI using seed growth and threshold techniques, in: 2007 International Conference on Intelligent and Advanced Systems. Presented at the 2007 International Conference on Intelligent and Advanced Systems, pp. 422–426. doi:10.1109/ICIAS.2007.4658421
Shaw, E., Arusell, R., Scheithauer, B., O'Fallon, J., O'Neill, B., Dinapoli, R., Nelson, D., Earle, J., Jones, C., Cascino, T., Nichols, D., Ivnik, R., Hellman, R., Curran, W., Abrams, R., 2002. Prospective randomized trial of low- versus high-dose radiation therapy in adults with supratentorial low-grade glioma: initial report of a North Central Cancer Treatment Group/Radiation Therapy Oncology Group/Eastern Cooperative Oncology Group study. J. Clin. Oncol. 20, 2267–2276. doi:10.1200/JCO.2002.09.126
Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge, UK ; New York.
Shevade, S.K., Keerthi, S.S., 2003. A simple and efficient algorithm for gene selection using sparse logistic regression. Bioinformatics 19, 2246–2253. doi:10.1093/bioinformatics/btg308
Simard, P.Y., Steinkraus, D., Platt, J.C., others, 2003. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis., in: Proceedings of Document Analysis and Recognition. Presented at the Seventh International Conference on Document Analysis and Recognition, Citeseer, pp. 958–963.
Simha, A., Irodi, A., David, S., 2012. Magnetic resonance imaging for the ophthalmologist: A primer. Indian J Ophthalmol 60, 301–310. doi:10.4103/0301-4738.98711
Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs].
Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T.L., Barrick, T.R., Howe, F.A., Ye, X., 2016. Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. Int J CARS 1–21. doi:10.1007/s11548-016-1483-3
Soltaninejad, M., Ye, X., Yang, G., Allinson, N., Lambrou, T., 2014. Brain tumour grading in different MRI protocols using SVM on statistical features, in: Medical Image Understanding and Analysis. British Machine Vision Association.
Soltaninejad, M., Zhang, L., Lambrou, T., Allinson, N., Ye, X., 2017. Multimodal MRI brain tumor segmentation using random forests with features learned from fully convolutional neural network. arXiv:1704.08134 [cs].
SPM - Statistical Parametric Mapping [WWW Document], n.d. URL http://www.fil.ion.ucl.ac.uk/spm/ (accessed 2.28.17).
Sprawls, P., 2000. Magnetic Resonance Imaging: Principles, Methods, and Techniques. Medical Physics Pub Corp, Estados Unidos.
Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M., 2014. Striving for Simplicity: The All Convolutional Net. arXiv:1412.6806 [cs].
Stall, B., Zach, L., Ning, H., Ondos, J., Arora, B., Shankavaram, U., Miller, R.W., Citrin, D., Camphausen, K., 2010. Comparison of T2 and FLAIR imaging for target delineation in high grade gliomas. Radiat Oncol 5, 5. doi:10.1186/1748-717X-5-5

Su, P., Yang, J., Li, H., Chi, L., Xue, Z., Wong, S.T., 2013. Superpixel-Based Segmentation of Glioblastoma Multiforme from Multimodal MR Images, in: Shen, L., Liu, T., Yap, P.-T., Huang, H., Shen, D., Westin, C.-F. (Eds.), Multimodal Brain Image Analysis, Lecture Notes in Computer Science. Springer International Publishing, pp. 74–83.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9. doi:10.1109/CVPR.2015.7298594

Taigman, Y., Yang, M., Ranzato, M., Wolf, L., 2014. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708.

Taormina, R., n.d. MATLAB_ExtraTrees - File Exchange - MATLAB Central [WWW Document]. URL http://uk.mathworks.com/matlabcentral/fileexchange/47372-rtaormina-matlab-extratrees (accessed 2.16.16).

Thiruvenkadam, K., Perumal, N., 2016. Fully automatic method for segmentation of brain tumor from multimodal magnetic resonance images using wavelet transformation and clustering technique. International Journal of Imaging Systems and Technology 26, 305–314. doi:10.1002/ima.22202

Tompson, J., Goroshin, R., Jain, A., LeCun, Y., Bregler, C., 2015. Efficient object localization using Convolutional Networks, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 648–656. doi:10.1109/CVPR.2015.7298664

Tran, T.N., Wehrens, R., Buydens, L.M.C., 2005. Clustering multispectral images: a tutorial. Chemometrics and Intelligent Laboratory Systems, FESTSCHRIFT HONOURING PROFESSOR D.L. MASSARTMassart S.I. 77, 3–17. doi:10.1016/j.chemolab.2004.07.011

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

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

Urban, G., Bendszus, M., Hamprecht, F., Kleesiek, J., 2014. Multi-modal brain tumor segmentation using deep convolutional neural networks, in: Proceedings of NCI-MICCAI BRATS. pp. 31–35.

Vaillant, R., Monrocq, C., Cun, Y.L., 1994. Original approach for the localisation of objects in images. IEE Proceedings - Vision, Image and Signal Processing 141, 245–250. doi:10.1049/ip-vis:19941301

Vedaldi, A., Lenc, K., 2015. MatConvNet: Convolutional Neural Networks for MATLAB, in: Proceedings of the 23rd ACM International Conference on Multimedia, MM '15. ACM, New York, NY, USA, pp. 689–692. doi:10.1145/2733373.2807412

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

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., Lang, K.J., 1989. Phoneme recognition using time-delay neural networks. IEEE Transactions on Acoustics, Speech, and Signal Processing 37, 328–339. doi:10.1109/29.21701

Wang, W., Steward, C.E., Desmond, P.M., 2009. Diffusion tensor imaging in glioblastoma multiforme and brain metastases: the role of p, q, L, and fractional anisotropy. AJNR Am J Neuroradiol 30, 203–208. doi:10.3174/ajnr.A1303

Wei Wu, A.Y.C.C., 2013. 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. doi:10.1007/s11548-013-0922-7

Wen, P.Y., Macdonald, D.R., Reardon, D.A., Cloughesy, T.F., Sorensen, A.G., Galanis, E., DeGroot, J., Wick, W., Gilbert, M.R., Lassman, A.B., Tsien, C., Mikkelsen, T., Wong, E.T., Chamberlain, M.C., Stupp, R., Lamborn, K.R., Vogelbaum, M.A., Bent, M.J. van den, Chang, S.M., 2010. Updated Response Assessment Criteria for High-Grade Gliomas: Response Assessment in Neuro-Oncology Working Group. JCO 28, 1963–1972. doi:10.1200/JCO.2009.26.3541

Winston, G.P., 2012. The physical and biological basis of quantitative parameters derived from diffusion MRI. Quant Imaging Med Surg 2, 254–265. doi:10.3978/j.issn.2223-4292.2012.12.05

Wolf, R., Platt, J.C., 1993. Postal Address Block Location Using a Convolutional Locator Network, in: Proceedings of the 6th International Conference on Neural Information Processing Systems, NIPS'93. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 745–752.

Wu, W., Chen, A.Y.C., Zhao, L., Corso, J.J., 2014. 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, 241–253. doi:10.1007/s11548-013-0922-7

Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.-H., Steinbach, M., Hand, D.J., Steinberg, D., 2008. Top 10 algorithms in data mining. Knowl Inf Syst 14, 1–37. doi:10.1007/s10115-007-0114-2

Yao, H., Chuyi, L., Dan, H., Weiyu, Y., 2016. Gabor Feature Based Convolutional Neural Network for Object Recognition in Natural Scene, in: 2016 3rd International Conference on Information Science and Control Engineering (ICISCE). Presented at the 2016 3rd International Conference on Information Science and Control Engineering (ICISCE), pp. 386–390. doi:10.1109/ICISCE.2016.91

Yi, J., Su, F., 2014. Histogram of Log-Gabor Magnitude Patterns for face recognition, in: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented at the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 519–523. doi:10.1109/ICASSP.2014.6853650

Young, G.S., 2007. Advanced MRI of adult brain tumors. Neurol Clin 25, 947–973, viii. doi:10.1016/j.ncl.2007.07.010

Yu, L., Liu, H., 2003. Feature selection for high-dimensional data: A fast correlation-based filter solution, in: ICML. pp. 856–863.

Yu, Z., Li, A., Au, O.C., Xu, C., 2012. Bag of textons for image segmentation via soft clustering and convex shift, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition. Presented at the 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 781–788. doi:10.1109/CVPR.2012.6247749

Zeiler, M.D., Fergus, R., 2014. Visualizing and Understanding Convolutional Networks, in: Computer Vision – ECCV 2014. Presented at the European Conference on Computer Vision, Springer, Cham, pp. 818–833. doi:10.1007/978-3-319-10590-1_53

Zeiler, M.D., Krishnan, D., Taylor, G.W., Fergus, R., 2010. Deconvolutional networks, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Presented at the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2528–2535. doi:10.1109/CVPR.2010.5539957

Zhang, L., Ye, X., Lambrou, T., Duan, W., Allinson, N., Dudley, N.J., 2016. A supervised texton based approach for automatic segmentation and measurement of the fetal head

and femur in 2D ultrasound images. Phys Med Biol 61, 1095–1115. doi:10.1088/0031-9155/61/3/1095

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

Zhao, Z., Liu, H., 2007. Spectral feature selection for supervised and unsupervised learning, in: Proceedings of the 24th International Conference on Machine Learning. ACM, pp. 1151–1157.

Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.S., 2015. Conditional Random Fields as Recurrent Neural Networks, in: 2015 IEEE International Conference on Computer Vision (ICCV). Presented at the 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1529–1537. doi:10.1109/ICCV.2015.179

Zhuang, A.H., Valentino, D.J., Toga, A.W., 2006. Skull-stripping magnetic resonance brain images using a model-based level set. NeuroImage 32, 79–92. doi:10.1016/j.neuroimage.2006.03.019

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

Zikic, D., Ioannou, Y., Brown, M., Criminisi, A., 2014. Segmentation of brain tumor tissues with convolutional neural networks. Proceedings MICCAI-BRATS 36–39.