University of California
Santa Barbara

Integrating Brain Connectome and Lesion Data for Patient Outcome Prediction

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Electrical and Computer Engineering

by

Po-Yu Kao

Committee in charge:

Professor B.S. Manjunath, Chair
Professor Jefferson W. Chen, University of California, Irvine
Professor Shivkumar Chandrasekaran
Professor Kenneth Rose

December 2019

The Dissertation of Po-Yu Kao is approved.

Professor Jefferson W. Chen, University of California, Irvine

---

Professor Shivkumar Chandrasekaran

Professor Kenneth Rose

Professor B.S. Manjunath, Committee Chair

December, 2019

Integrating Brain Connectome and Lesion Data for Patient Outcome Prediction

by

Po-Yu Kao

Dedicated to my family and friends.

Acknowledgements

The completion of this dissertation would not have been possible without those who have been supportive and encouraging in the past seven years.

First and most importantly, I would like to express my deepest gratitude to Professor B.S. Manjunath, who has advised my research with unwavering support, patience, and inspiration, and guided me how to approach problems with logical thinking. I came to UCSB without digital image processing and computer vision background, but the courses I took here not only introduced me into the field of image processing and computer vision but gave me the fundamental knowledge needed in pursuit of my PhD degree.

Second, I am grateful to Professor Jefferson W. Chen, Professor Kenneth Rose, and Professor Shivkumar Chandrasekaran, for agreeing to be on my dissertation committee and offering insightful suggestions. I am also thankful to NIH funding # HD059217 and # 5R01NS103774 that supported my Ph.D. study.

It has been a pleasure to collaborate with Thuyen Ngo, Angela Zhang, Shailja, Tom Jiang, Eduardo Rojas, Michael Goebel, and Amil Khan. Especially I have learned a lot from Thuyen through the discussion and exploring many deep learning and computer vision ideas together. I enjoyed the time I spent interacting with Aruna Jammalamadaka, S. Karthikeyan, Diana Delibaltov, Santhoshkumar Sunderrajan, Niloufar Pourian, Lakshmanan Nataraj, Christopher Wheat, Hui Hu, Amir Mohaymen Rahimi, Carlos Torres, Dmitry Fedorov, Christian Lang, Archith John Bency, Kristian Kvilekval, Utkarsh Gaur, Rahul Vishwakarma, Wenhui Jiang, Xudong Lin, Zency Young, Tao Deng, Lingyun Song, Oytun Ulutan, Austin McEver, Satish Kumar, Devendra Jangid, A S M Iftekhar, Aditya Jonnalagadda, and Rohan Jain. My appreciation further goes to my good friends, Greg Kuo, Brian Hsu, Joseph Pan, Yi-Tai Chiu, Yen-Ju Lin, Shani Tang, Vivian Lin, Justin Wang, Ye Wang, Louis Lee, Jeremy Chiang, Colin Tseng, Ted Lee, Qian Xu, Wei-Ting

Lin, CY Xu, Ming-Yu Lin, Clint Greene, Kuo-Kai Hsieh, Chieh-Chi Kao, Fred Lin, Yuxiang Wang, Tina Wang, Stanley Hung, Martin Keh, Willy Liu, Cynthia Chung, Andy Chou, Chien-Hung Chen, Chong Huang, Chien-Yu Chang, Arthur Lee, Benson Lin, Chia-Chi Wu, Yu-Wei Liu, and Marco Lam, who have accompanied me and brought me joy all the time. I would like to express my gratitude to Rose Chang, Jasmine Hsu, and Stephanie Lee. Thank you for everything and the time we had together. My sincere thanks also goes to Dr. Cheng-Chun Chang, Dr. Tien-Ying Kuo, Dr. Yung-Chung Wang, and Mr. Jason Peng, who provided me their personal insights on the career development and wrote 18 recommendation letters for me. Without their supports, it would not be possible to attend UCSB.

Finally, special thanks to my dearest family, younger brother, grandmother, mother, and dad, who love me and support me unconditionally. Without them, I could not have gone this far.

Curriculum Vitæ

Po-Yu Kao

Education

December 2019	**Doctor of Philosophy** Electrical and Computer Engineering University of California, Santa Barbara, USA.
September 2015	**Master of Science** Electrical and Computer Engineering University of California, Santa Barbara, USA.
June 2011	**Bachelor of Science** Electrical Engineering National Taipei University of Technique, Taipei, Taiwan.

Field of Study

Medical Image Analysis, Computer Vision and Machine Learning

Honors & Awards

2018	6th out of 63 teams in BraTS 2018 challenge, International MICCAI Brainlesion Workshop
2018	Conference Travel Grant, Graduate Student Association, UCSB
2016	3rd place in mTOP 2016 challenge, International MICCAI Brainlesion Workshop
2016	UCSB Doctoral Student Travel Grant, Academic Senate, UCSB
2011	2nd Place in the senior capstone project, Electrical Engineering department, NTUT
2011	Chicony Electronics Scholarship, Chicony Electronics Company
2011	Alumni Association Scholarship, Electrical Engineering department, NTUT
2011	Academic Progress Award, NTUT
2011, 2008	President Award, Electrical Engineering department, NTUT
2011, 2009, 2008	Shin Kong Life Scholarship, Shin Kong Life Insurance Scholarship Foundation
2009	Outstanding Teaching Assistant Award, NTUT

Volunteer Experience

2019	Reviewer for MICCAI 2019 and ICIP 2019

2018	Reviewer for International MICCAI Brainlesion Workshop 2018
2016	Reviewer for International MICCAI Brainlesion Workshop 2016
2011	Youth Ambassador, Ministry of Foreign Affairs, Taiwan

Languages

English, Mandarin, Taiwanese

Organization

Mar 13 - Mar 14	President, Taiwanese Student Association, UCSB

Programming Languages

Python, Matlab, C++, C, R

Publications

Kao, Po-Yu, Shailja, Jiaxiang Jiang, Angela Zhang, Amil Khan, Jefferson W. Chen and B.S. Manjunath. "Improving Patch-Based Convolutional Neural Networks for MRI Brain Tumor Segmentation by Leveraging Location Information." (2019). (Submitted to Frontier in Neuroscience.)

Kao, Po-Yu, Jefferson W. Chen, and B.S. Manjunath. "Improving 3D U-Net for Brain Tumor Segmentation by Utilizing Lesion Prior." (2019).

Kao, Po-Yu, Jefferson W. Chen, and B.S. Manjunath. "Predicting Clinical Outcome of Stroke Patients with Tractographic Feature." (2019). (Accepted by BrainLesion 2019 as an oral presentation)

Kao, Po-Yu, Angela Zhang, Michael Goebel, Jefferson W. Chen, and B.S. Manjunath. "Predicting Fluid Intelligence of Children using T1-weighted MR Images and a StackNet." Adolescent Brain Cognitive Development Neurocognitive Prediction (ABCD-NP), vol 11791., Springer, Cham, 2019.

Jiang, Jiaxiang, **Po-Yu Kao**, Samuel A. Belteton, Daniel B. Szymanski and B.S. Manjunath. "Accurate 3D Cell Segmentation using Deep Feature and CRF Refinement." In 2019 IEEE International Conference on Image Processing (ICIP), pp. 1555-1559., IEEE, 2019.

Zhang, Angela, **Po-Yu Kao**, Ashutosh Shelat, Ronald Sahyouni, Jefferson Chen, and B.S. Manjunath. "Fully Automated Volumetric Classification in CT Scans for Diagnosis and Analysis of Normal Pressure Hydrocephalus." arXiv preprint arXiv:1901.09088 (2019).

Bakas, Spyridon, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara et al. "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge." arXiv preprint arXiv:1811.02629 (2018).

Kao, Po-Yu, Thuyen Ngo, Angela Zhang, Jefferson Chen, and B.S. Manjunath. "Brain tumor segmentation and tractographic feature extraction from structural mr images for overall survival prediction." In International MICCAI Brainlesion Workshop, pp. 128-141. Springer, Cham, 2018.

Kao, Po-Yu, Eduardo Rojas, Jefferson W. Chen, Angela Zhang, and B.S. Manjunath. "Unsupervised 3-D feature learning for mild traumatic brain injury." In International MICCAI Brainlesion Workshop, pp. 282-290. Springer, Cham, 2016.

Abstract

Integrating Brain Connectome and Lesion Data for Patient Outcome Prediction

by

Po-Yu Kao

This research focuses on introducing novel machine learning algorithms for predicting the outcome of patients with brain disorders using MR images. We first introduce the challenges in medical image analysis and overview the magnetic resonance imaging, datasets, and existing tools. A brain lesion segmentation method is then presented. The main novelty of this method is a new feature fusion method that integrates location information and the state-of-the-art patch-based neural networks for lesion segmentation. The proposed feature fusion method improves the segmentation performance of the patch-based neural networks. Thereafter, we focus on predicting the overall survival of brain tumor patients. We introduce a novel feature called the tractographic feature to capture the potentially damaged regions due to the presence of the lesion. The tractographic feature is built from the lesion and average connectome information from a group of normal subjects. It takes into account different functional regions that are affected by the lesion, thus complementing the commonly used lesion volume features. The tractographic feature is tested on the Multi-modal Brain Tumor Segmentation (BraTS) 2018 dataset and achieves a better overall survival prediction performance than other features and the gold standard that uses patient age. The proposed tractographic feature is also used to predict the clinical outcome of stroke patients. On the publicly available stroke benchmark, Ischemic Stroke Lesion Segmentation (ISLES) 2017, our proposed tractographic feature achieves higher accuracy than the state-of-the-art feature descriptors. Finally, we focus on predicting the outcome of mild traumatic brain injury (TBI) pa-

tients. The tractographic feature cannot be built in this case since the mTBI patients do not have brain lesions in the traditional MR images or CT scans. Here, we present an unsupervised 3D feature clustering algorithm, consisting of 3D dictionary learning, convolutional network, k-means clustering algorithm with group constraint, to gather mild TBI patients into three groups using their structural and diffusion MR images. The proposed method won the 3rd place in Mild Traumatic Brain Injury Outcome Prediction (mTOP) 2016 challenge.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cogito, ergo sum

René Descartes, 1637

The brain is the control center of the human body. It is a part of the nervous system including the spinal cord and a large network of nerves and neurons. Together, the nervous system controls everything from senses to the muscles throughout body [14].

When the brain is damaged, it affects many different functions, including memory, sensation, and personality. Brain disorders include any conditions or disabilities that affect the brain. This includes conditions that are caused by illness, genetics, or traumatic injury

In this dissertation, we focus on three different types of brain disorders: brain tumor, stroke and traumatic brain injury.

Brain Tumor A brain tumor is a collection, or mass, of abnormal cells in the brain. The skull, which encloses the brain, is very rigid. Any growth inside such a restricted space can cause problems. Brain tumors can be malignant or benign. When tumors grow, the pressure inside the skull is increased. This can cause brain damage, and it can

be life-threatening.

Brain tumors are categorized as primary or secondary. A primary brain tumor originates in the brain. Many primary brain tumors are benign. A secondary brain tumor, also known as a metastatic brain tumor, occurs when cancer cells spread to the brain from another organ, such as lung or breast [15].

Glioma is a common type of brain tumor in adults originating in the glial cells that support neurons and help them function. The World Health Organization (WHO) classification system categorizes gliomas from grade I (lowest grade) through grade IV (highest grade), based upon histopathologic characteristics which predict their behavior over time [16]. Low-grade gliomas (LGGs) consist of WHO-grade I tumors and WHO-grade II tumors, that tend to exhibit benign tendencies and indicate a better prognosis for the patient. WHO-grade III and IV tumors are included in high-grade gliomas (HGGs) that are malignant and more aggressive. In this dissertation, we focus on *high-grade giloma* and *low-grade giloma* patients.

Stroke Stroke is the top five cause of death and a leading cause of disability in the United States. A stroke occurs when a blood vessel that carries oxygen and nutrients to the brain is either blocked by a clot or bursts. When it happens, part of the brain cannot get the blood and oxygen, and the impacted brain cells die.

Stroke can be caused either by a clot obstructing the flow of blood to the brain (called an ischemic stroke) or by a blood vessel rupturing and preventing blood flow to the brain (called a hemorrhagic stroke). A transient ischemic attack (TIA), or mini-stroke, is caused by a temporary clot [17]. In this dissertation, we use publicly available data on the *ischemic* stroke patients.

Traumatic Brain Injury A traumatic brain injury (TBI) is defined as a blow to the head or a penetrating head injury that disrupts the normal function of the brain. TBI can happen when the head suddenly and violently hits an object or when an object pierces the skull and enters brain tissue. Symptoms of a TBI can be mild, moderate or severe, depending on the extent of damage to the brain. Mild cases may result in a brief change in mental state or consciousness, while severe cases may result in extended periods of unconsciousness, coma or even death [18]. In this dissertation, we focus on *mild traumatic brain injury* which is hard to be observed and diagnosed in the CT scan or structural MR image.

1.1 Challenges

Lack of appropriately annotated large-scale datasets is the biggest challenge in medical image analysis. Although every day there are millions of medical images generated worldwide, most of them lack proper annotations for training scalable machine learning methods. Unlike annotating natural images, annotating medical images requires medical knowledge from experts such as radiologists and physicians. Besides, manual annotation by clinical experts is both time-consuming and expensive. For these reasions, high-quality annotated medical imaging datasets are not easily available. For example, the well-known natural image dataset, ImageNet [19], has more than 14 million annotated images but one of the largest medical image datasets, the Adolescent Brain Cognitive Development Neurocognitive Prediction Challenge (ABCD-NP-Challenge 2019) [20–24], only provides data for 8556 subjects including 3739 training subjects, 415 validation subjects and 4402 testing subjects with subject-wised annotations. In addition, the Multi-modal Brain Tumor Segmentation (BraTS) 2016 and BraTS 2017 [25] only have 750 subjects with voxel-wised annotations.

1.2 Summary of Contributions

The main contributions of this dissertation are threefold: First, we propose a novel feature fusion method which integrates location information of brain and the state-of-the-art patch-based neural networks that improve the segmentation performance of these networks. Second, we propose a novel tractographic feature from the lesion region to predict the clinical outcome of stroke patients and the overall survival of brain tumor patients. Third, we propose a fully unsupervised methodology to learn the 3D features from the volumetric data and use these features to predict the outcome of patients with mild traumatic brain injury in an unsupervised manner.

1.3 Dissertation Organization

In Chapter 2, an overview of magnetic resonance imaging (MRI), datasets and tools used in this research are presented. We introduce (i) the structural and diffusion MRI, (ii) the datasets of brain tumor patients, stroke patients, mild traumatic brain injury patients, and normal subjects, and (iii) the tools for visualization, registration, and fiber tracking.

In Chapter 3, we present a brain lesion segmentation method that takes into account location information. Current segmentation methods for this problem include 3D U-Net [26] and DeepMedic [27, 28]. Due to the limitation of GPU memory in the current infrastructure, these state-of-the-art neural networks can only be trained by the sub-regions (patches) of the brain rather than the whole brain volume. Therefore, location information of these patches of the brain is missing for these networks. That is, these neural networks do not know where these patches come from the brain. Here, we present a novel feature fusion method which integrates location information and these state-of-

the-art patch-based neural networks for brain lesion segmentation. The proposed feature fusion method provides location information of sub-patches to the patch-based neural networks with a limited GPU memory constraint. An existing brain parcellation atlas is used as additional location information to the patch-based neural networks. The proposed location information fusion method improves the segmentation performance of these state-of-the-art patch-based neural networks. The experimental results are reported in a public multi-modal *brain tumor* segmentation (BraTS) 2018 dataset [2–4].

In Chapter 4, we focus on predicting the overall survival of *brain tumor* patients. We first use the brain tumor segmentation method proposed in Chapter 3 to segment the brain tumor from multi-modal structural MR images. We introduce a new feature, called the tractographic feature, that incorporates the lesion locations towards predicting the overall survival rates of tumor patients. From the lesion region, we also extract other features such as the volume of lesion, the shape characteristics of the lesion, and the distributions of the lesion in different brain regions to predict the OS of brain tumor patients.The proposed tractographic feature has a better OS prediction performance than other first-order features and patient's age which is shown to be the gold standard in this task [29–32]. We report experimental results on a public multi-modal brain tumor segmentation (BraTS) 2018 dataset [2–4]. We also compare the OS prediction performance of tractographic features extracted from the predicted lesion and the ground-truth lesion.

In Chapter 5, we focus on predicting the clinical outcome of *ischemic stroke* patients. Here, we use the ground-truth of the stroke lesion to predict the clinical outcome of patients to have a fair comparison with the-state-of-the-art features [9]. These features, computed from the ground truth lesion data, are used to train random forest based estimators. It is noted that one random forest regressor is only trained with one type of feature. The proposed tractographic feature again has the best clinical outcome predic-

tion performance compared to other features and similar performance compared to the state-of-the-art feature [9]. Besides, the proposed tractographic feature is more robust, and its dimensionality is much less than the state-of-the-art feature [9]. The experimental results are reported in a public ischemic stroke lesion segmentation (ISLES) 2017 [33, 34] dataset.

In Chapter 6, we focus on predicting the outcome of patients with *mild traumatic brain injury*. Since the lesion data is not available on this study, we can not use the proposed tractographic features. Instead, we propose an unsupervised 3D feature clustering method to cluster the subjects into 3 groups. We learn the 3D dictionaries from the brain MR-T1, diffusion tensor fractional anisotropy, and diffusion tensor mean diffusivity images provided by the mild traumatic brain injury outcome prediction challenge (mTOP) 2016 competition. These dictionaries are then used as the convolution kernels of a 3D network to extract the feature from each subject. Thereafter, we apply the k-means clustering algorithm on the extracted features to predict the outcomes of subjects with group constraint.

In Chapter 7, we present some future directions and conclude this dissertation.

Chapter 2

Brain Imaging, Datasets and Tools

> Man can alter his life by altering his thinking.
>
> *William James*

This chapter first introduces different types of Magnetic Resonance Imaging (MRI) including structural MRI (sMRI) and diffusion MRI (dMRI). We then introduce the datasets used in this research. These datasets contain brain tumor patients, stroke patients, mild traumatic injury patients, and normal subjects. We also briefly discuss the existing tools including ITK-SNAP [8], FSL [35–37], and DSI Studio [38] used in this research.

2.1 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is used in radiology to construct images of the anatomy and the physiological processes of the body. MRI scanners generate strong magnetic fields, magnetic field gradients, and radio waves to form images of the organs inside the body and the body tissue. Different from computed tomography (CT or CAT)

scans and positron emission tomography (PET) scans, MRI does not associate with X-rays or the use of ionizing radiation.

An MRI scan is still a better imaging choice than a CT scan because of the hazards of X-ray in CT. MRI is widely operated in hospitals and clinics for medical diagnosis, staging of disease and follow-up without exposing the body to ionizing radiation. An MRI scan yields different information compared with a CT scan. Risks and discomfort may be associated with MRI scans. Compared with CT scans, MRI scans typically take longer (approximately 40 minutes), and the examined subject is required to enter a narrow and confining tube. Besides, subjects with medical implants or non-removable metal inside their body are unsafe to undergo an MRI examination.

2.1.1 Structural Magnetic Resonance Imaging

Structural magnetic resonance imaging (sMRI) is a widely used medical imaging technique in research as well as in clinical practice [39]. sMRI translates the local differences in water content into different shades of gray that serve to outline the shapes and sizes of the brain's various sub-regions. An MRI scanner delivers a specific radio-frequency that excites hydrogen atoms in the water molecule, which returns some of this energy in the form of a characteristic nuclear magnetic resonance signal. Not all protons "resonate" in that way, but enough do such that the resulting computer-generated image constitutes a highly detailed map of the brain's tissues and structures. Therefore, this tool can be used to discover the presence of abnormal tissue through the changes in tissue density or composition. Radiologists examining an sMRI can readily distinguish between gray and white matter and other types of tissue, both normal, such as blood vessels, and abnormal, such as tumors, by their different shading and contrast compared with surrounding areas [40].

Different Types of Sequences

Many pulse sequences are available, emphasizing different aspects of normal and abnormal brain tissue. By modifying sequence parameters such as repetition time (TR) and echo time (TE), for example, anatomical images can emphasize contrast between gray and white matter (e.g., T1-weighted with short TR and short TE) or between brain tissue and cerebrospinal fluid (e.g., T2-weighted with long TR and long TE). Sequences vary in the information they provide and how long they take to acquire. Different image processing approaches often require specific types of sequences and may recommend specifically tuned sequences to provide the best results [41].

Four different types of MR images are used in this research. These MR images include T1-weighted image (MR-T1), contrast-enhanced T1-weighted image (MR-T1ce), T2-weighted image (MR-T2), and T2 Fluid Attenuated Inversion Recovery image (FLAIR). Figure 2.1 shows examples of MR-T1, MR-T2, and FLAIR images, and Figure 2.2 shows a comparison of MR-T1 and MR-T1ce images [42]. More details of these different types of MR images are described as follows.

T1-weighted image (MR-T1) T1-weighted image (MR-T1) with short TR and short TE provides good contrast between gray matter (dark gray) and white matter (lighter gray) tissues, while CSF is void of signal (black). Therefore, water, such as CSF, as well as dense bone and air, appear dark, and fat, such as lipids in the myelinated white matter, appears bright. Cortex (grey matter) appears dark, and the contrast between the grey matter and white matter is best. Besides, pathological processes, such as demyelination or inflammation, often increase water content in tissues, which decreases the signal on T1; white matter disease often shows up as darker areas in the lighter gray-colored white matter.

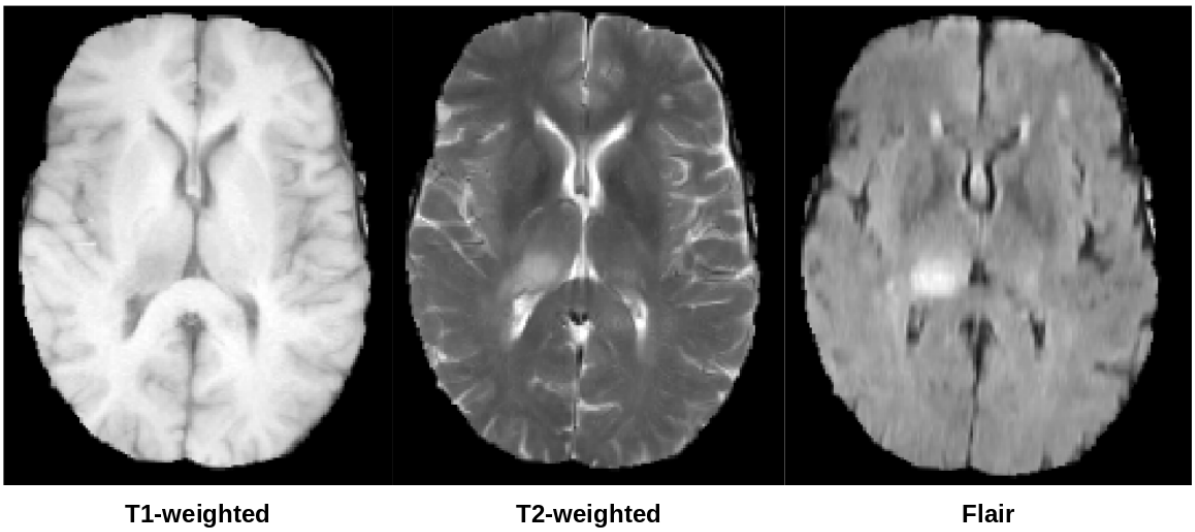


Figure 2.1: Comparison of MR-T1 (left), MR-T2 (middle), and FLAIR (right) images of the brain. These MR images are from MICCAI Brain Tumor Segmentation (BraTS) 2018 challenge [1–4].

Contrast-enhanced T1-weighted image (MR-T1ce) T1-weighted imaging can also be performed while infusing Gadolinium (Gad). Gad is a non-toxic paramagnetic contrast enhancement agent. When injected during scanning, Gad changes signal intensities by shortening T1. Thus, Gad is very bright on T1-weighted images. Gad enhanced images are especially useful in looking at vascular structures and breakdown in the blood-brain barrier, e.g., tumors or inflammation.

T2-weighted image (MR-T2) T2-weighted image with long TR and long TE provides good contrast between CSF (bright) and brain tissue (dark). Some T2 sequences demonstrate additional contrast between gray matter (lighter gray) and white matter (darker gray). Therefore, water, such as CSF, appears bright, while air appears dark, and fat, such as lipids in the white matter, appears dark. In addition, pathological processes, such as demyelination or inflammation, often increase water content in tissues, which increases signal on T2; white matter disease often shows up as brighter areas, which makes subtle changes easier to be detected.

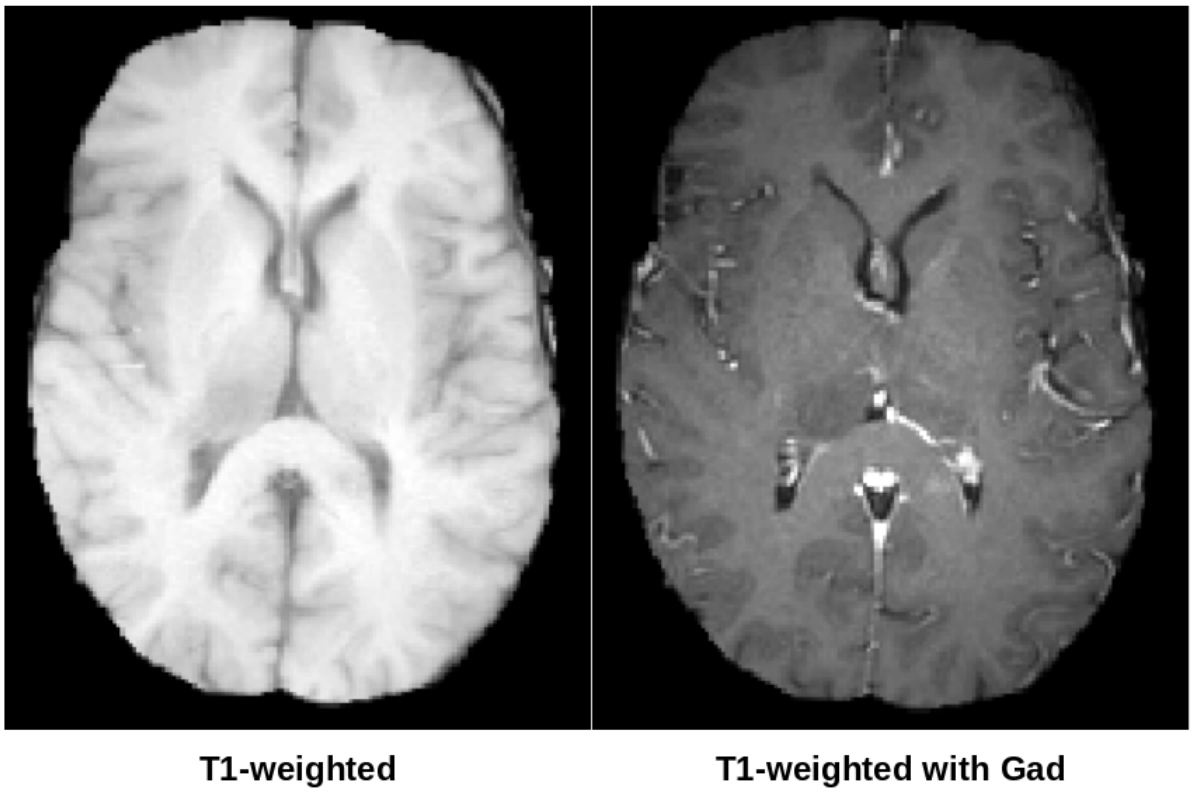


Figure 2.2: Comparison of MR-T1 (left), and MR-T1ce (right) images of the brain. These MR images are from MICCAI Brain Tumor Segmentation (BraTS) 2018 challenge [1–4].

T2 Fluid Attenuated Inversion Recovery image (FLAIR) FLAIR image with very long TR and very long TE is very sensitive to pathology and provides good contrast between CSF (dark) and abnormalities (bright).

2.1.2 Diffusion Magnetic Resonance Imaging

Diffusion-weighted imaging (DWI) is designed to detect the random movements of water protons. Water molecules diffuse relatively freely in the extracellular space. Their movement is significantly restricted in the intracellular space. Spontaneous movements, referred to as diffusion, rapidly become restricted in ischemic brain tissue. During ischemia, the sodium-potassium pump shuts down and sodium accumulates intra-cellularly. Water then shifts from the extracellular to the intracellular space due to the osmotic gradient. As water movement becomes restricted intra-cellularly, this results in an extremely bright signal on DWI. Thus, DWI is an extremely sensitive method for detecting an acute stroke. Figure 2.3 shows two example DWI images of the brain.

2.2 Datasets

The datasets used in this research include the brain MR images of (i) brain tumor patients, (ii) ischemic stroke patients, (iii) mild traumatic brain injury patients, and (iv) normal subjects.

2.2.1 Multimodal Brain Tumor Segmentation (BraTS)

Multimodal Brain Tumor Segmentation (BraTS) 2018 [1–4, 43] dataset is designed for two tasks (i) brain tumor segmentation and (ii) overall survival (OS) prediction of the brain tumor patients. It provides pre-operative multimodal MRI scans of high-grade

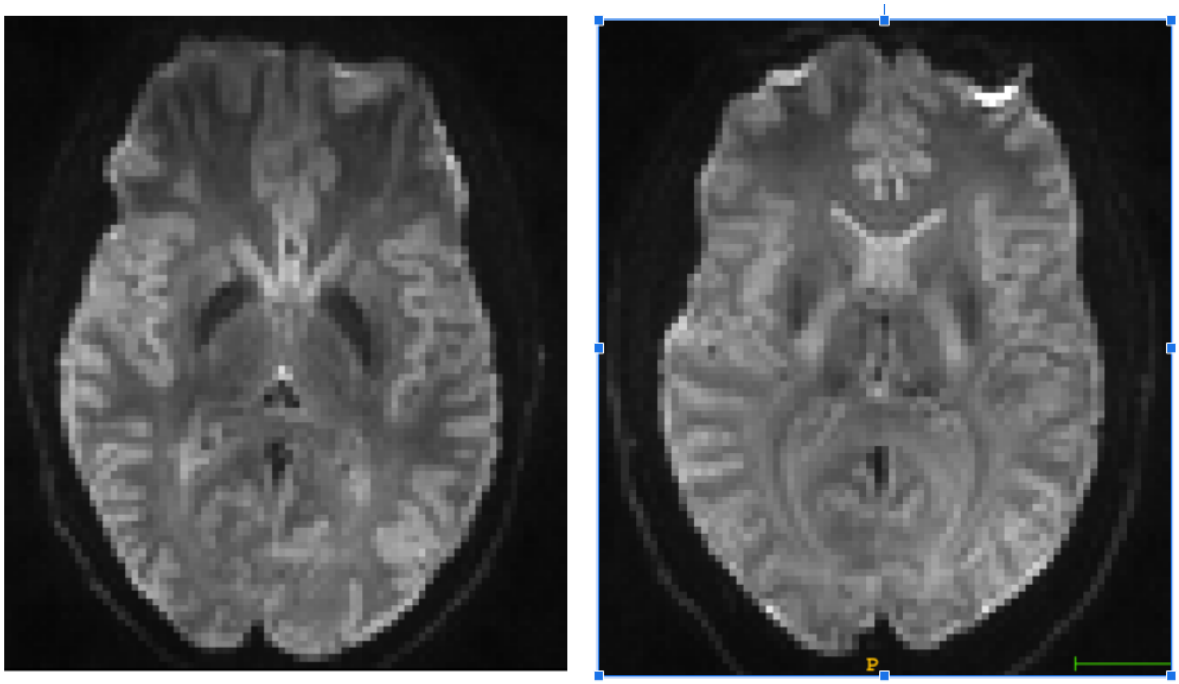


Figure 2.3: DWI images of the brain. These images are from MICCAI Mild Traumatic Brain Injury Outcome Prediction (mTOP) 2016 challenge.

glioma (HGG) patients and lower-grade glioma (LGG) patients, with a pathologically confirmed diagnosis and available overall survival (OS) as the training, validation and testing data.

BraTS 2018 dataset is public, and it can be accessed and downloaded through the official webpage (`https://www.med.upenn.edu/sbia/brats2018/registration.html`). It contains 285 training subjects (210 HGG and 75 LGG), 66 validation subjects, and 191 testing subjects. Each subject has four 3D MRI modalities (MR-T1, MR-T1ce, MR-T2, and MR-FLAIR) rigidly aligned, re-sampled to $1 \times 1 \times 1$ mm isotropic resolution, and skull-stripped. The image size is $240 \times 240 \times 155$ in the x-, y-, and z-directions, respectively. For brain tumor segmentation task, annotations include 3 tumor sub-regions: the enhancing tumor (ET), the edema (ED), and the necrotic & non-enhancing tumor (NCR/NET). The annotations are combined into three nested sub-regions: whole tumor (WT), tumor core (TC) and enhancing tumor (ET), as shown in Figure 2.4. The ground

truth labels are withheld for the validation and testing dataset. These datasets require participants to upload the segmentation masks to the official server for evaluations. The validation dataset (66 cases) allows multiple submissions and is designed for intermediate evaluations. The testing dataset (191 cases) allows only a single submission and is used to calculate the final BraTS segmentation challenge ranking.

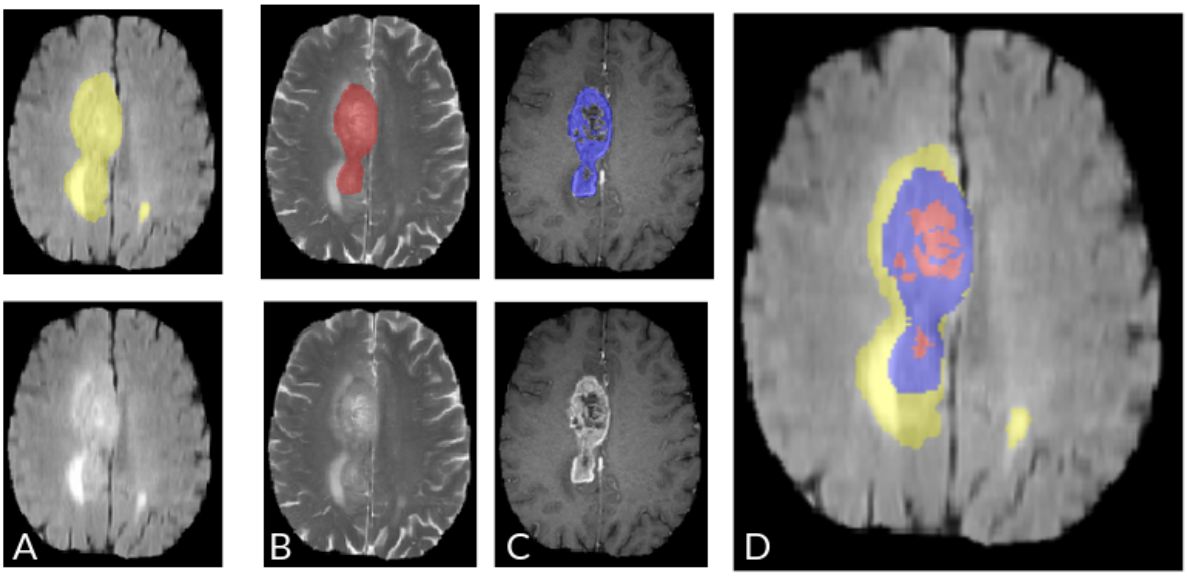


Figure 2.4: **Glioma sub-regions.** These multimodal MR image patches with annotations are from a BraTS 2018 training subject. The multimodal MR image patches show from left to right: the whole tumor (yellow) visible in FLAIR (A), the tumor core (red) visible in T2 (B), the enhancing tumor (blue) visible in T1ce (C). The annotations are integrated to generate the final labels of the tumor sub-regions (D): edema (yellow), non-enhancing tumor & necrosis (red), enhancing tumor (blue).

For the overall survival prediction task, the BraTS 2018 dataset also includes the age (in years), survival (in days) and resection status for each of 163 subjects in the training dataset, and 59 of them have the resection status of gross total resection (GTR). The validation dataset has 53 subjects with the age (in years) and resection status, and 28 of them have the resection status of GTR. The test dataset has 131 subjects with the age (in years) and resection status, and 77 of them have the resection status of GTR. For the OS prediction task, we only predict the overall survival (OS) for glioma patients

with resection status of GTR. That is, 59, 28, and 77 subjects are considered in the training, validation, and test set, respectively. The ground truth OS is withheld for the validation and testing dataset. These datasets required participants to upload the OS to the official server for evaluations. The validation dataset (28 cases) allows multiple submissions and is designed for intermediate evaluations. The testing dataset (77 cases) allows only a single submission and is used to calculate the final BraTS OS prediction challenge ranking.

2.2.2 Ischemic Stroke Lesion Segmentation (ISLES)

Ischemic Stroke Lesion Segmentation (ISLES) 2017 (`http://www.isles-challenge.org/ISLES2017/`) [33,34] is a public available dataset and contains diverse ischemic stroke cases. It provides two different tasks including lesion outcome prediction and clinical outcome prediction. ISLES 2017 is used for clinical outcome prediction in our research.

Figure 2.5 shows the timeline of ISLES 2017 data acquisition. The typical stroke treatment procedure involves the acquisition of brain images at a time-since-stroke (TSS) in minutes. Depending on these and other factors, a treatment decision is made and intervention occurs after a time-to-treatment (TTT) in minutes. The outcome of this intervention is assessed via the standardized thrombolysis in cerebral infarction (TICI) score (see Table 2.1). After the effects are stabilized, the final clinical outcome is determined with the modified Rankin Scale (mRS) (see Table 5.1) and the final lesion outcome by brain imaging.

The treating physician has to decide in each case individually, whether the risky intervention (be it thrombectomy or thrombolysis) is justified by the potential gain, based solely on the lesion appearance, the time-since-stroke (TSS) and her/his experience. A method to reliably predict the actual lesion and clinical outcome from the acute scans

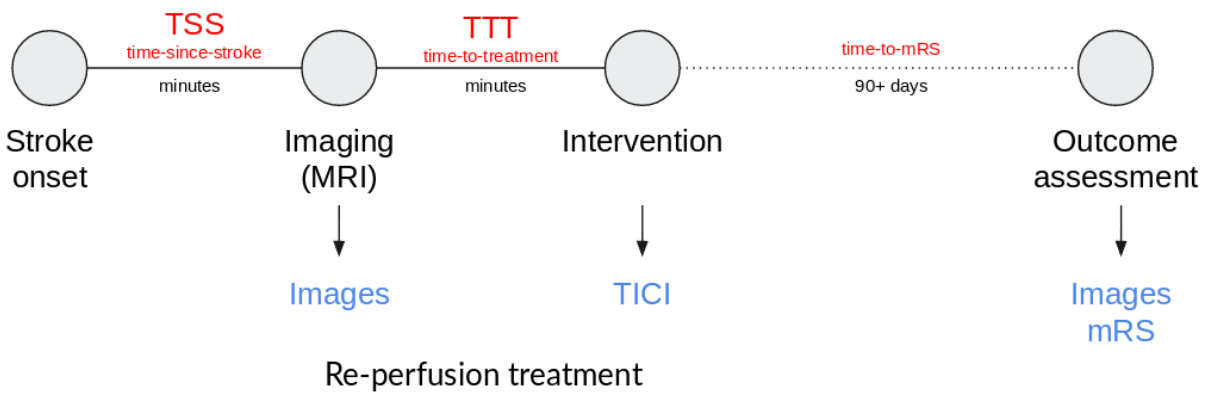


Figure 2.5: Timeline of ISLES 2017 data. Figure adjusted from the official ISLES 2016 website (`http://www.isles-challenge.org/ISLES2016/`).

alone would provide invaluable evidence for the clinical decision-making process.

Image Format & Pre-processing The provided images are in uncompressed Neuroimaging Informatics Technology Initiative (NIfTI) format. All MR images have been skull stripped and co-registered to each individual subject space.

Acute Imaging Data Details The training set contains 43 ischemic stroke patients. Two diffusion maps are provided for each subject. The diffusion maps contain the DWI trace images (DWI maps) and the apparent diffusion coefficient (ADC) maps. The perfusion maps contain the cerebral blood volume (CBV), cerebral blood flow (CBF) and mean transit time (MTT), defined as the ratio of volume to flow of cerebral blood (CBV/CBF). Furthermore, the time to peak (TTP) concentration of the CA and the time need at which the residue function reaches its maximum value (Tmax).

Follow-up Stroke Imaging After approximate 90 days of the intervention, an anatomical sequence (MR-T2 or MR-FLAIR) was obtained when the stroke lesion had stabilized to assess the final lesion outcome. The provided ground-truth stroke lesion maps of ISLES 2017 were manually delineated on those images.

Clinical Parameter Details The ISLES 2017 also provides clinical parameters including time-since-stroke (TSS), time-to-treatment (TTT), thrombolysis in cerebral infarction (TICI) scale, and modified Rankin scale (mRS) for each training subject.

Time-since-stroke (TSS) in minutes is the main factor as only a fast re-establishing of the cerebral perfusion can lead to benefits for the ischemic stroke patient. The less time passed since stroke onset, the more likely a re-perfusion can rescue brain tissue.

Time-To-Treatment (TTT) in minutes denotes the time passed between imaging the patient and the intervention, e.g., re-perfusion treatment. Similar to the TSS, its value determines the amount of recoverable brain tissue.

The TICI scale shown in Table 2.1, proposed by Higashida et al. [11], provides a standardized method to evaluate intra-cranial perfusion assessed in cerebral angiography. The TICI scale is used to assess the achieve re-perfusion after a flow-restoration intervention such as thrombectomy.

Days to the modified Rankin scale (mRS) are also provided for each ISLES 2017 training subject. The 90-days mRS is a commonly used and standardized scale to assess the degree of disability 90 days after a stroke incidence and can be considered as the clinical outcome [13]. In this research, we only focus on the subjects who acquire the mRS on 90 days. The purpose of our research is to predict the clinical outcome of stroke patients based on their stroke lesion mask. Table 2.2 shows the details of mRS grades.

2.2.3 Mild Traumatic Brain Injury Outcome Prediction (mTOP)

Mild traumatic brain injury outcome prediction (mTOP) 2016 dataset (`https://tbichallenge.wordpress.com`) is a private dataset. It can only be used in participating mTOP 2016 challenge. No copyright transfer of any kind will take place. mTOP 2016 dataset consists of 27 subjects belonging to 3 different categories (healthy, patient cat-

Table 2.1: Thrombolysis in Cerebral Infarction (TICI) scores [11, 12]

Grade 0	No Perfusion. No antegrade flow beyond the point of occlusion.
Grade 1	Penetration With Minimal Perfusion. The contrast material passes beyond the area of obstruction but fails to opacify the entire cerebral bed distal to the obstruction for the duration of the angiographic run.
Grade 2	Partial Perfusion. The contrast material passes beyond the obstruction and opacifies the arterial bed distal to the obstruction. However, the rate of entry of contrast into the vessel distal to the obstruction and/or its rate of clearance from the distal bed are perceptibly slower than its entry into and/or clearance from comparable areas not perfused by the previously occluded vessel, eg, the opposite cerebral artery or the arterial bed proximal to the obstruction.
Grade 2a	Only partial filling (2/3) of the entire vascular territory is visualized.
Grade 2b	Complete filling of all of the expected vascular territory is visualized, but the filling is slower than normal.
Grade 3	Complete Perfusion. Antegrade flow into the bed distal to the obstruction occurs as promptly as into the obstruction and clearance of contrast material from the involved bed is as rapid as from an uninvolved other bed of the same vessel or the opposite cerebral artery.

Table 2.2: Modified Rankin Scale (mRS) scores [13]

Grade 0	No Symptoms at all.
Grade 1	No significant disability despite symptoms; able to carry out all usual duties and activities.
Grade 2	Slight disability; unable to carry out all previous activities, but able to look after own affairs without assistance.
Grade 3	Moderate disability; requiring some help, but able to walk without assistance.
Grade 4	Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance.
Grade 5	Severe disability; bedridden, incontinent and requiring constant nursing care and attention.
Grade 6	Deceased.

egory I or patient category II) each consisting of 9 subjects. Patients were categorized in one of the two groups according to their outcome after mild TBI. The testing set contains 15 subjects belonging to 3 different categories (healthy, patient category I or

patient category II) each consisting of 5 subjects.

Imaging Data Details The dataset is available in both raw and pre-processed format. The raw dataset consists of MR T1-weighted images (MR-T1), diffusion-weighted images (DWI) and corresponding meta-data including b-vectors (bvecs) and b-values (bvals).

The pre-processed data include MR T1-weighted images (MR-T1), fractional anisotropy (DT-FA) and mean diffusivity (DT-MD) maps from DWI, as well as probability maps for grey matter (GM) and white matter (WM) shown in Figure 2.6.

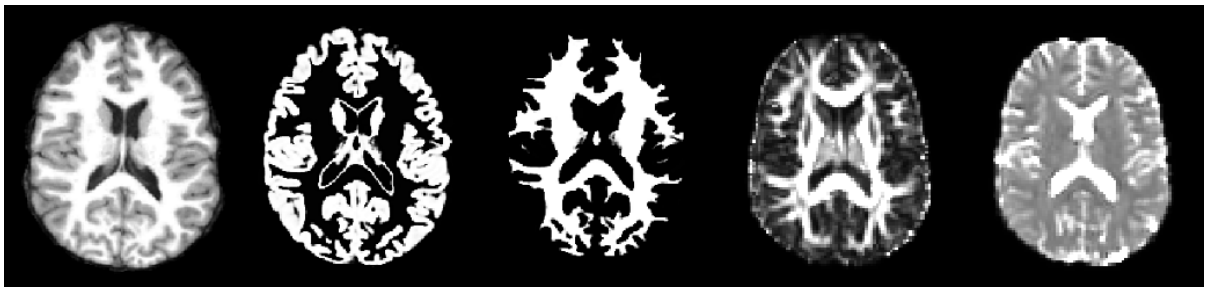


Figure 2.6: Pre-processed images of one mild traumatic brain injury outcome prediction (mTOP) 2016 subject. MR-T1, grey and white matter probability maps, fractional anisotropy and mean diffusivity (from left to right). Figure taken from the official mTOP 2016 website.

The dimension of the pre-processed MR-T1 image is $182 \times 218 \times 182$ in the x-, y-, and z-directions, respectively, with $1mm \times 1mm \times 1mm$ voxel resolution. The dimension of DT-FA and DT-MD image is $91 \times 109 \times 91$ in the x-, y-, and z-directions, respectively, with $2mm \times 2mm \times 2mm$ voxel resolution. All provided MR images are in the Neuroimaging Informatics Technology Initiative (NIfTI) format.

Image Pre-Processing MR-T1 images were first bias field corrected and skull-stripped. Then, they were rigidly registered to the MNI 152 [44] 1mm space. Grey and white matter probability maps were extracted from the registered images in the MNI 152 1mm space. Diffusion images were first de-noised, corrected for head motion and eddy current

artifacts as well as bias-field corrected. Then, extracted DT-FA and DT-MD maps were mapped to the MNI 2mm space by rigid registration to the corresponding MR-T1 image (inter-subject registration).

2.2.4 Human Connectome Project (HCP)

The Human Connectome Project (HCP) was started in July 2009 [45] as the first of three Grand Challenges of the NIH's Blueprint for Neuroscience Research [46]. The goals are to construct a connectome that will elucidate the anatomical and functional connectivity within the *healthy human brain*, and to create a great number of data that will facilitate research into brain disorders [47,48]. On September 15, 2010, the NIH awarded grants totaling 40 million US dollars to multiple research institutions led by Washington University, St. Louis, the University of Minnesota, Twin Cities, Massachusetts General Hospital (MGH)/Harvard University, Boston, and the University of California Los Angeles (UCLA) [49].

HCP is a public dataset and can be accessed through their official website (`https://db.humanconnectome.org/app/template/Login.vm`). HCP S1200 [50] is used in this research to construct the average diffusion information of a group of normal subjects. The 1200 Subjects Release (S1200) includes behavioral and 3T MR imaging data from 1206 healthy young adult participants (1113 with structural MR scans) collected in 2012-2015. In addition to 3T MR scans, 184 subjects have multi-modal 7T MRI scan data and 95 subjects also have some resting-state MEG (rMEG) and/or task MEG (tMEG) data available.

2.3 Tools

In this section, we introduce the existing tools used in this research. These tools are used for visualization, registration and fiber tracking.

2.3.1 ITK-SNAP

ITK-SNAP [8] is a free, open-source, and multi-platform software application used to visualize and segment structures in 3D medical images. It provides manual and semi-automatic segmentation tools using active contour methods. It also offers image navigation for 3D medical images. The main advantages of ITK-SNAP include (1) support for multiple 3D image formats, including NIfTI and DICOM, (2) support for multi-channel and time-varying images, and (3) manual segmentation in the sagittal, coronal, and transverse planes at once. Figure 2.7 shows an example of using ITK-SNAP for visualizing the multi-modal MR images with the ground-truth tumor lesions labels of a BraTS 2018 training subject. ITK-SNAP is used for visualization purposes in our research.

2.3.2 FMRIB Software Library (FSL)

FSL [35–37] provides extensive and comprehensive tools for analyzing structural MRI, functional MRI, and diffusion tensor imaging (DTI) brain imaging data. It runs on both Linux and Windows via a virtual machine. Most tools can be run through both the command line and GUI. The overview of FSL tools is available on the official website (`https://fsl.fmrib.ox.ac.uk/fsl/fslwiki`).

FMRIB's Linear Image Registration Tool (FLIRT) We only use the brain image registration tool called FLIRT [51–53] (FMRIB's Linear Image Registration Tool) in our research. FLIRT (`https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT/UserGuide`) is

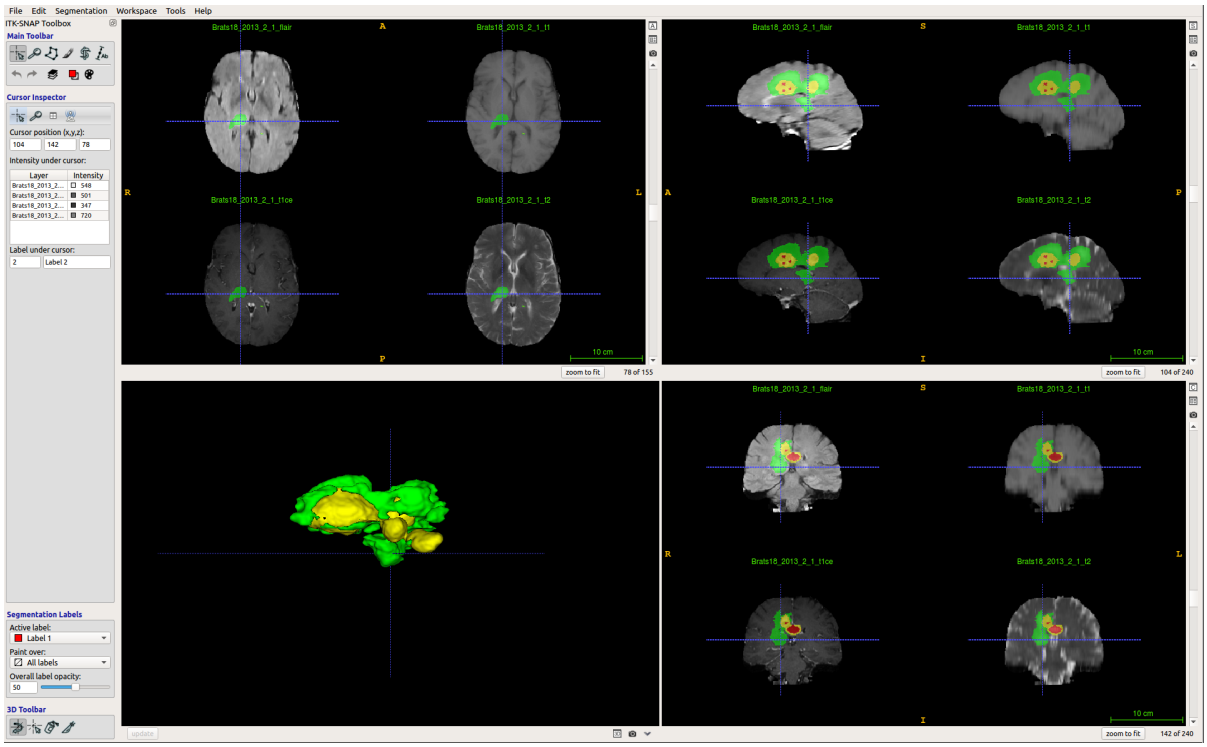


Figure 2.7: Visualizing multi-modal MR images with the ground-truth tumor lesions labels of a BraTS 2018 training subject in ITK-SNAP. Top left, top right, and bottom left windows show the axial, sagittal, and coronal planes of multi-modal MR images, respectively. Bottom left window shows the 3D reconstruction from the labels.

an automated affine registration tool for the brain image. It registers the brain image in the subject to the normalization space, e.g., the MNI 152 space. Both linearly and non-linearly generated MNI 152 template images are stored in *fsl/data/standard* (see Figure 2.8).

Figure 2.9 shows the GUI of FLIRT in FSL. Here, we try to register a BraTS 2018 training subject ($240 \times 240 \times 155$) to the MNI 152 1mm space ($182 \times 218 \times 182$) using its MR-T1 image. One needs to specify the reference image and the input image, and FLIRT will generate the registered image. Figure 2.10 shows an example of the original MR-T1 images of the BraTS 2018 training subject and the resulting registered MR-T1 images of the same subject in the MNI 152 space.

2.3.3 DSI Studio

DSI Studio [38] is a tractography software tool that pinpoints brain connectome. It provides several analysis methods for diffusion MRI (`http://dsi-studio.labsolver.org/Manual`). It has both GUIs and command-line tools. In our research, we use a deterministic fiber tracking method which is provided by DSI Studio to reconstruct the average fiber tracts of a group of normal subjects and the data analysis tool to build the connectivity matrix of the fiber tracts.

DSI Studio provides a deterministic fiber tracking algorithm that uses quantitative anisotropy to improve accuracy [54]. Deterministic fiber tracking tends to yield the best estimation of fiber tracts since it is more similar to a *maximum likelihood estimation* of the fiber tracks. Deterministic fiber tracking is suitable for our research purpose since we would like to find the most likely route of a fiber pathway from a group-averaged connectome. Figure 2.11 shows the whole-brain fiber tracts of a group average diffusion information, and these tracts are reconstructed by using DSI Studio. Please refer to their

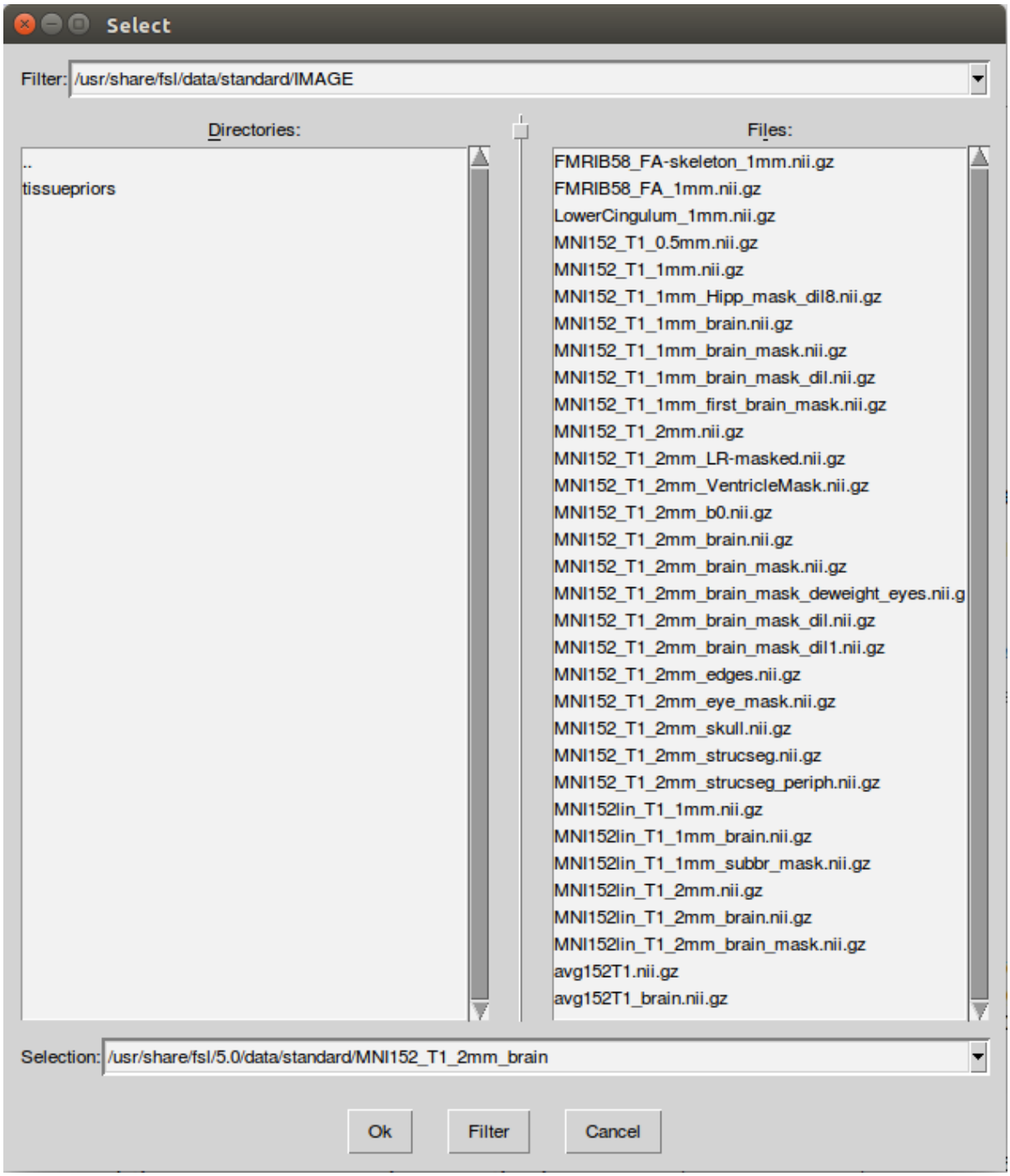


Figure 2.8: A list of linearly and non-linearly generated MNI 152 template images in *fsl/data/standard*.

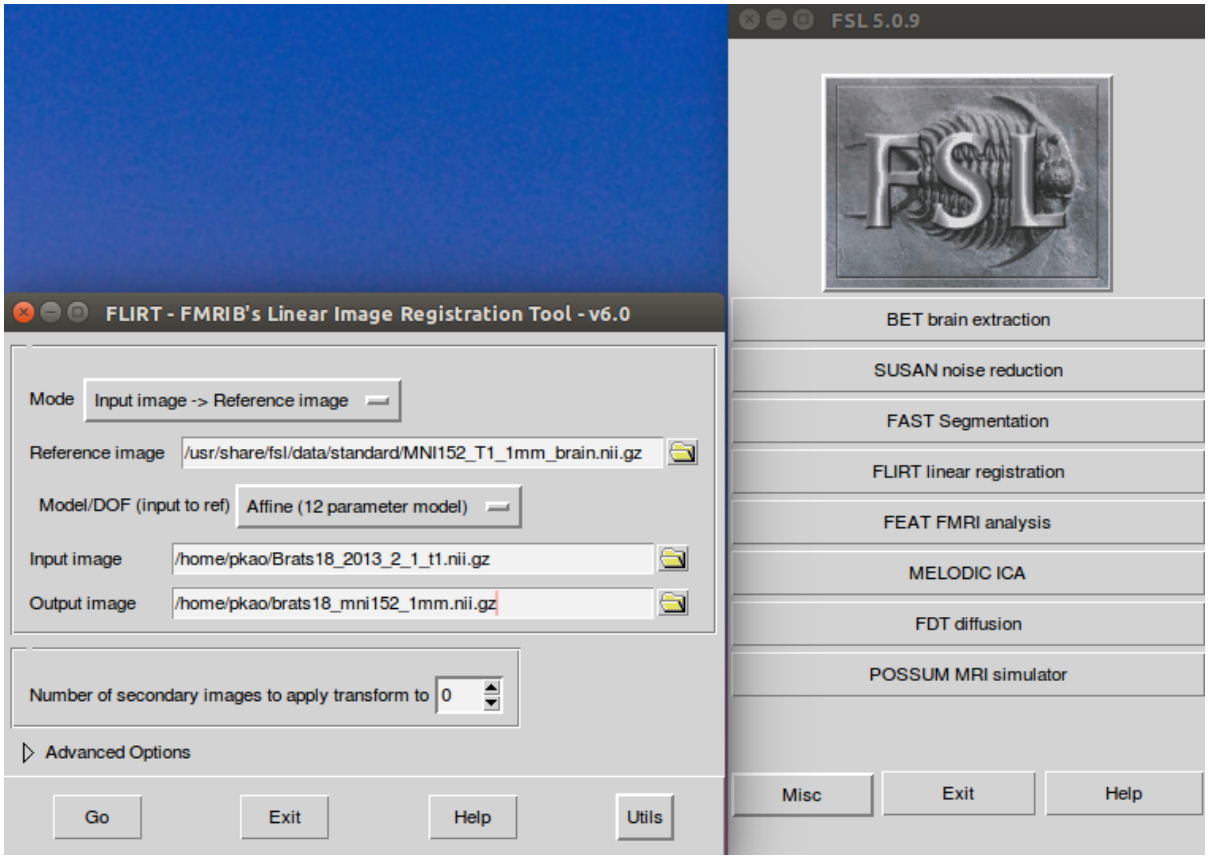


Figure 2.9: GUI of FLIRT in FSL. Here, we try to register a BraTS 2018 training subject ($240 \times 240 \times 155$) to the MNI 152 1mm space ($182 \times 218 \times 182$) using its MR-T1 image. You need to specify the reference image and the input image, and FLIRT will generate the registered image.

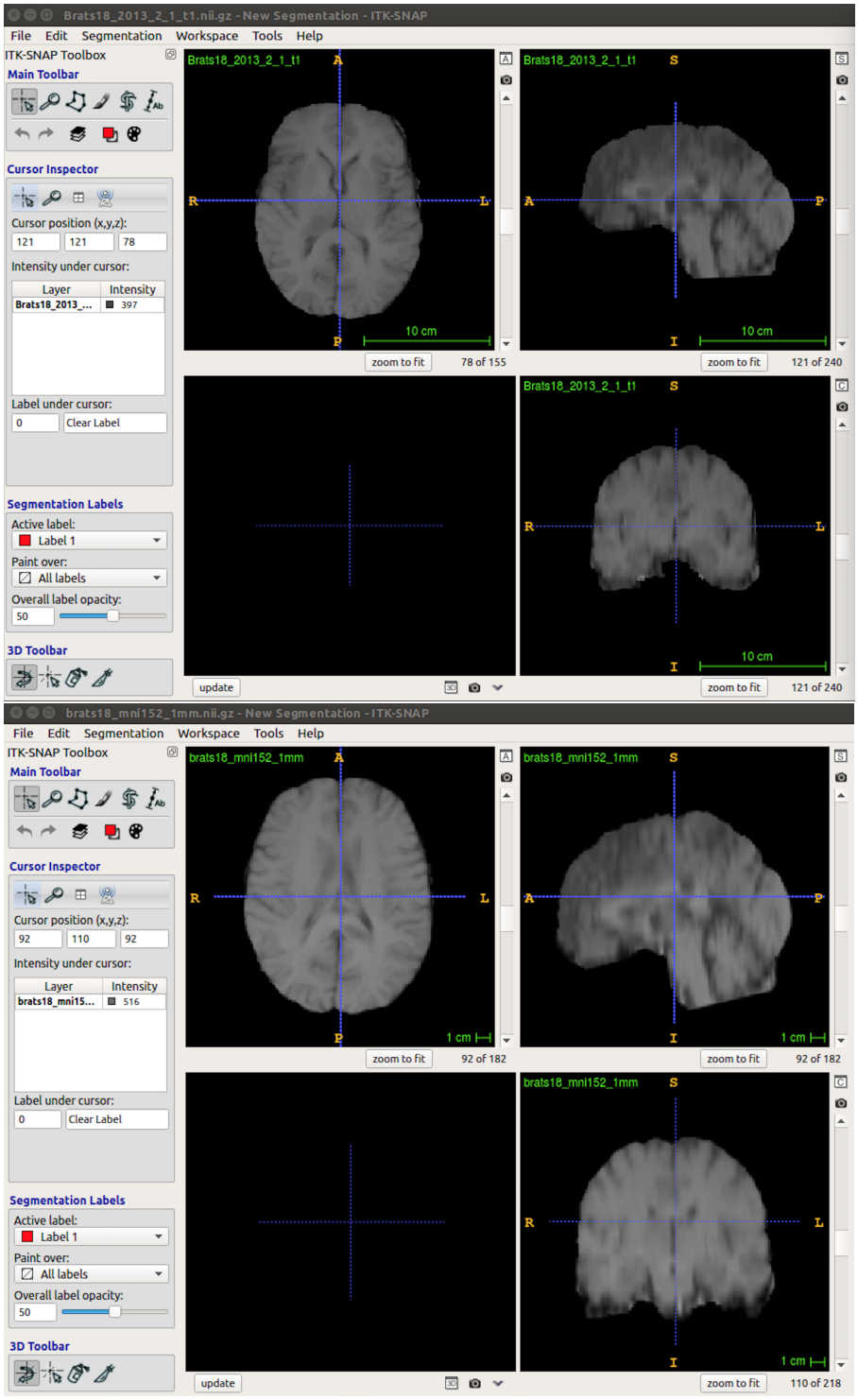


Figure 2.10: Top: The original MR-T1 images of BraTS 2018 training subject ($240 \times 240 \times 155$). Bottom: The resulting registered MR-T1 images of the same subject in the MNI 152 space ($182 \times 218 \times 182$).

official web-page for more details of fiber tracking (`http://dsi-studio.labsolver.org/Manual/Fiber-Tracking#TOC-Step-T3:-Fiber-tracking`).

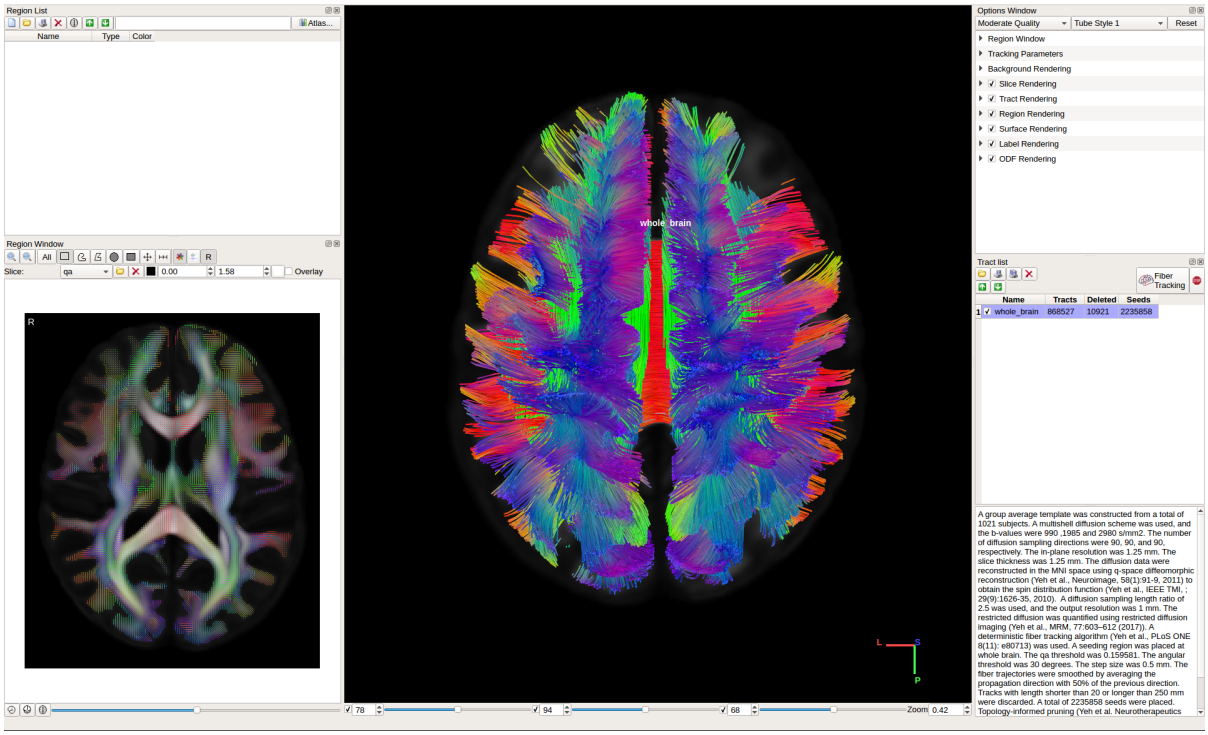


Figure 2.11: The whole-brain fiber tracts of a group average diffusion information, and these tracts are reconstructed by using DSI Studio. The left bottom figure shows the diffusion tensor imaging (DTI), and the middle figure shows the constructed fiber tracts. Different colors of fiber tracts represent different directions.

DSI Studio also provides data analysis tools for the constructed fiber tracts. It can convert the fiber tracts into the connectivity matrix which describes the numbers of fibers between different brain parcellation regions. Figure 2.12 shows the connectivity matrix of the whole-brain fiber tracts in Figure 2.11 built by DSI Studio with the automated anatomical labeling [5] (AAL) atlas.

The network measures such as density, transitivity, global efficiency, the diameter of the graph, etc. can be extracted from the connectivity matrix using DSI Studio. Figure 2.13 shows the network measures of the connectivity matrix in Figure 2.12 by using DSI Studio. Please refer to their office web-page for more details of track-

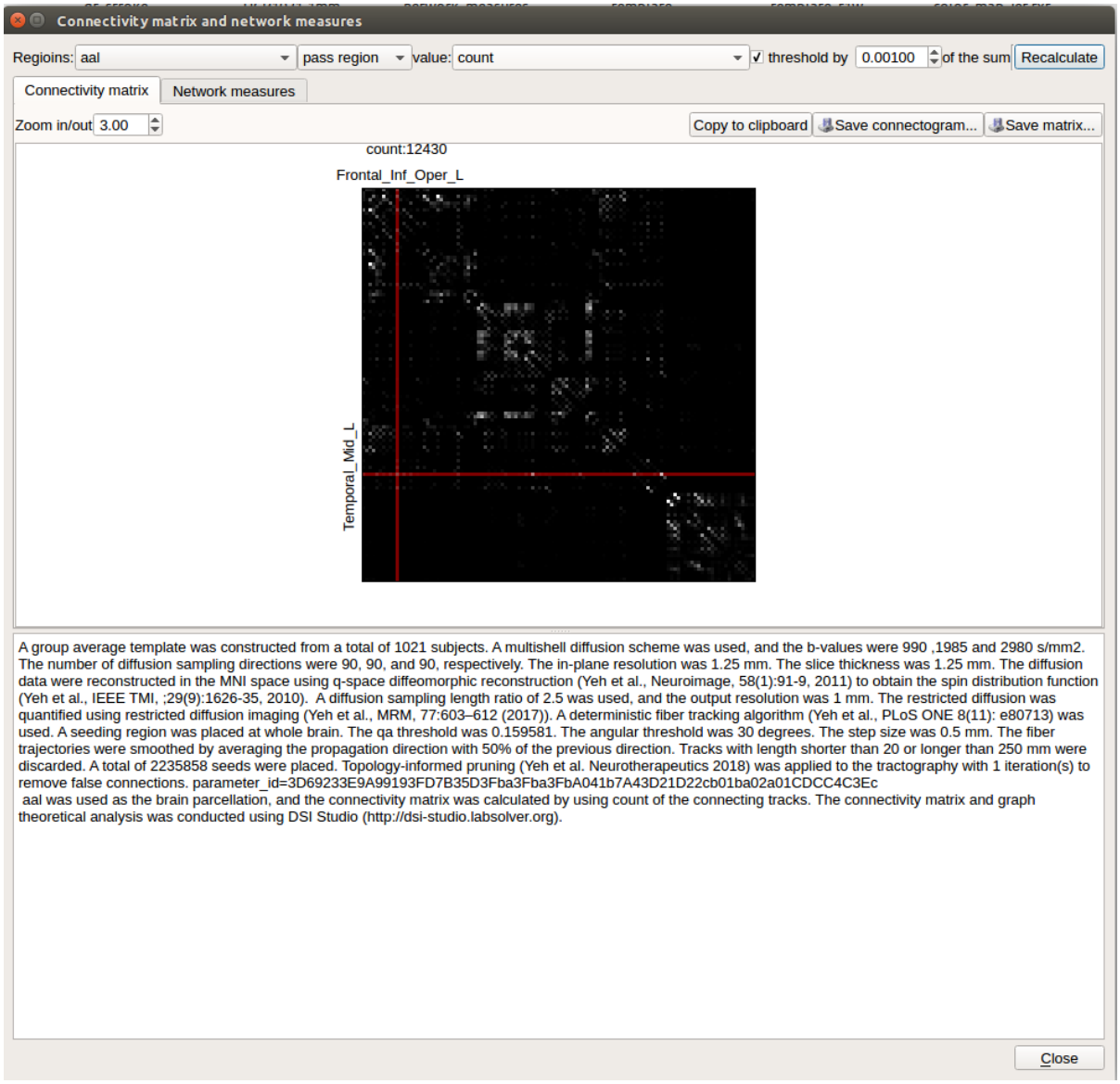


Figure 2.12: The connectivity matrix of the whole-brain fiber tracts in Figure 2.11 built by DSI Studio with the automated anatomical labeling [5] (AAL) atlas. The dimension of this connectivity matrix is 116×116, and the element value of this connectivity matrix represents the number of tracts passing through two AAL regions. The brighter element values represents the higher numbers of tracts passing through two regions.

specific analysis and network measures (`http://dsi-studio.labsolver.org/Manual/tract-specific-analysis`).

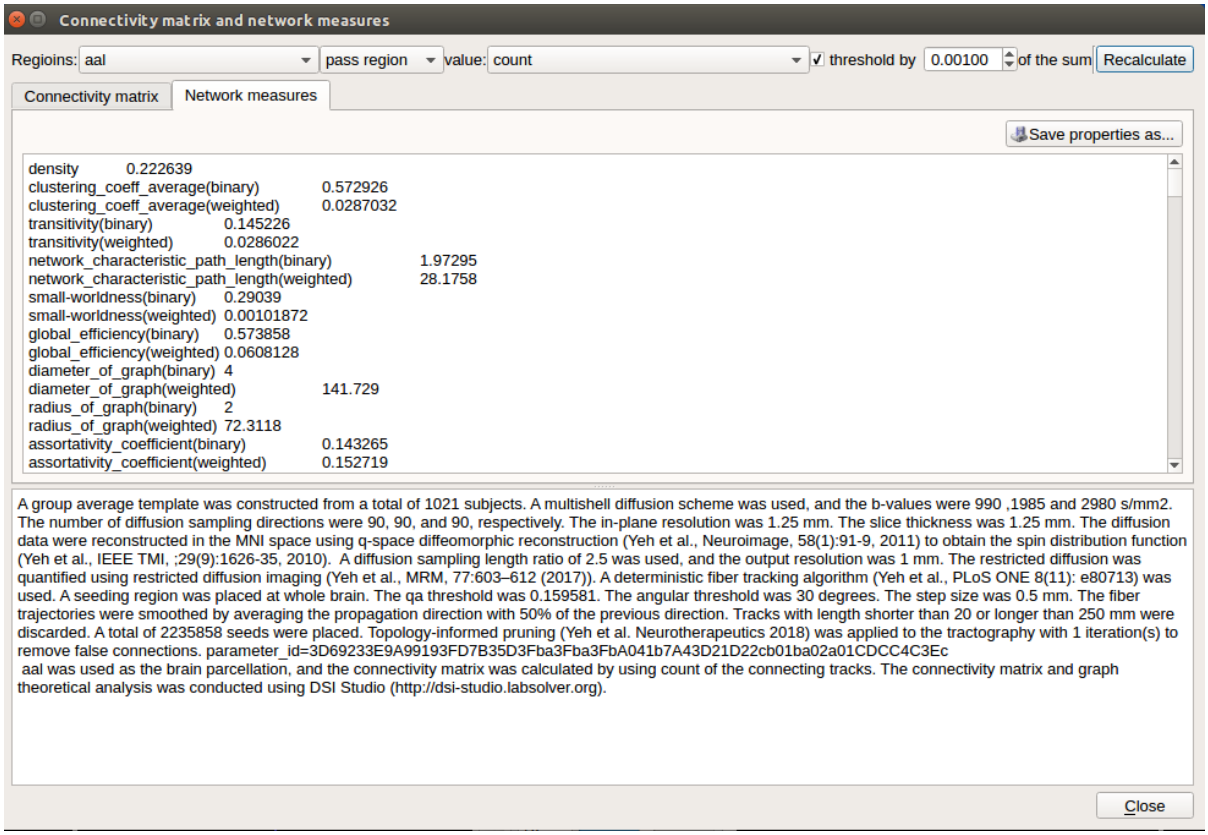


Figure 2.13: The network measures of the connectivity matrix in Figure 2.12 by using DSI Studio.

Chapter 3

MRI Brain Tumor Segmentation by Leveraging Location Information

> The world as we have created it is a process of our thinking. It cannot be changed without changing our thinking.
>
> *Albert Einstein*

An automated and accurate brain lesion segmentation tool is needed to automatically predict the outcome of patients with brain lesions. In this chapter, we proposed a feature fusion method which integrates location information and multimodal MR images into the state-of-the-art patch-based neural networks for brain tumor segmentation. This is motivated by the observation that lesions are not uniformly distributed across different brain regions and that a locality-sensitive segmentation is likely to obtain better segmentation accuracy. Towards this, we use an existing brain parcellation atlas in the Montreal Neurological Institute (MNI) space and map this atlas to the individual subject data. This mapped atlas in the subject data space is integrated with structural

Magnetic Resonance (sMR) imaging data and patch-based neural networks, including 3D U-Net [26] and DeepMedic [28], are trained to classify the different brain lesions. Multiple state-of-the-art neural networks are trained and integrated with XGBoost fusion in the proposed two-level ensemble method. The first level reduces the uncertainty of the same type of models with different seed initializations, and the second level leverages the advantages of different types of neural network models. The two main contributions of this work are: First, we propose a location information fusion method which improves the segmentation performance of state-of-the-art networks; Second, we proposes a novel two-level ensemble method which reduces the uncertainty of prediction and leverages the advantages of different neural networks.

3.1 Introduction

Glioma is a common type of brain tumor in adults originating in the glial cells that support neurons and help them function. The World Health Organization (WHO) classification system categorizes gliomas from grade I (lowest grade) through grade IV (highest grade), based upon histopathologic characteristics which predict their behavior over time [16]. Low-grade gliomas (LGGs) consist of WHO-grade I tumors and WHO-grade II tumors, that tend to exhibit benign tendencies and indicate a better prognosis for the patient. WHO-grade III and IV tumors are included in high-grade gliomas (HGG) that are malignant and more aggressive. Patients with HGG had median survival time (MST) 18 months, and the MST of patients with Grade III and IV glioma were 26 and 13 months, respectively [55]. Gliomas are further divided into four types of lesions, namely edema, non-enhancing core, necrotic core, and non-enhancing core based on the acuteness of the tumor cells. However, segmenting the different lesions of gliomas is a daunting task because of the intrinsic heterogeneity which affects their visual appearance as well

as shape. Clinically, MR images help a doctor to evaluate the tumor and plan the treatment. Moreover, the treatment depends on the type, size, shape, grade, and location of the tumor, which varies widely. Consequently, this observation leads to the importance of an automated and accurate brain tumor segmentation for better diagnosis of brain tumors. Also, the manual annotation process is time and resource consuming, therefore, an automated and accurate brain tumor segmentation method is in great demand.

Deep neural networks (DNNs) have achieved state-of-the-art segmentation performance on the recent Multimodal Brain Tumor Segmentation (BraTS) Challenges [1]. Kamnitsas et al. [56] conducted the comparative study on performance and concluded that deep learning along with ensemble learning-based methods outperform the others as they leverage the advantage of each deep learning model. Wang et al. [57] analyzed three different binary segmentations task rather than a single multi-class segmentation task, and three different binary segmentations task has a better performance than a single multi-class segmentation task. Similarly, Isensee et al. [58] proposed to integrate segmentation layers at different levels of optimized 3D U-Net-like architectures followed by element-wise summation. [59] implemented a modified decoder and encoder structure of CNN to generate dense segmentation. Likewise, Isensee et al. [60] demonstrated that an original U-Net architecture trained with additional institution dataset improved the dice score of enhancing tumor. Mckinley et al. [61] also proposed a U-Net-like network and introduce a new loss function, a generalization of binary cross-entropy, to account for label uncertainty. Furthermore, Zhou et al. [62] explored the ensemble of different networks including multi-scale context information and also segmented three tumor subregions in cascade with an additional attention block.

Our recent work [63] utilizes an existing parcellation to bring location information of brain into patch-based neural networks that improve the brain tumor segmentation performance. Outputs from 26 models are averaged, including 19 different types of

DeepMedics [28] and seven different types of 3D U-Nets [26], to get the final tumor predictions. The proposed ensemble containing six models including three DeepMedics and three 3D U-Nets with different seed initializations takes less than 1 minutes in the inference time. We also propose a novel two-level ensemble method which reduces the uncertainty of predictions in the first level and takes advantage of different types of models in the second level. Our experimental results show that the proposed location fusion methods improve the segmentation performance of the single state-of-the-art patch-based network and an ensemble of multiple state-of-the-art patch-based networks. The proposed ensemble has better segmentation performance compared to the state-of-the-art networks in BraTS 2017 dataset and competitive performance to the state-of-the-art networks in BraTS 2018 dataset.

3.2 Dataset

The Multimodal Brain Tumor Segmentation Challenges (BraTS) 2017 dataset and BraTS 2018 dataset [2–4, 43] comprise clinically-acquired pre-operative multimodal MRI scans of glioblastoma (GBM/HGG) and lower-grade glioma (LGG) as training, validation and test data. There are 285 subjects in the training set and 46 and 66 subjects in the validation set of BraTS 2017 and BraTS 2018, respectively. The lesion ground-truth labels are available for the training subjects but withheld for both the validation and test subjects. MRI scans were available as native (T1), contrast-enhanced T1-weighted (T1ce), T2-weighted (T2) and T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes. These scans were distributed after being skull-stripped, pre-processed, re-sampled and interpolated into 1 mm isotropic resolution with an image size of $240 \times 240 \times 155$ in the x-, y-, and z-direction. Tumor segmentation labels were produced manually by a trained team of radiologists and radiographers. The edema was segmented primarily

from T2 images, non-enhancing and enhancing core of the tumor from T1ce together with the lesions visible in T1 and necrotic core from T1ce. We used the annotated and co-registered imaging datasets including the Gd-enhancing tumor, the peri-tumoral edema and the necrotic and non-enhancing tumor core for our training and test procedure.

3.3 Data Pre-processing

Different modalities used for mapping tumor-induced tissue changes include MR-T1, MR-T1ce, MR-T2, and MR-FLAIR, which leads to varying intensity ranges. We first normalize each modality to a standard range of values. Each MR image is pre-processed by first clipping it at (0.2 percentile, 99.8 percentile) of non-zero voxels to remove the outliers. Subsequently, each modality is normalized individually using $\bar{x}_i = (x_i - \mu)/\sigma$ where i is the index of voxel inside the brain, $\bar{x}_i$ is the normalized voxel, x_i is the corresponding raw voxel, and μ and σ are the mean and standard deviation of the raw voxels inside the brain, respectively.

3.4 Network Architectures

Two different network architectures adapted from DeepMedic [28] and 3D U-Net [26] are examined in this study. DeepMedic was initially designed for brain lesion segmentation, e.g., stroke lesions [64] and brain tumor lesions [27], and 3D U-Net which is the 3D version of U-Net [65] is widely used for the volumetric image segmentation tasks [66–68]. More details of network architectures are described below.

3.4.1 Modified DeepMedic

The first network architecture shown in Figure 3.1 is modified from DeepMedic [28]. The number of convolutional kernels is indicated within the white box. Batch normalization [69] is used. Residual connection [70] is used in the normal resolution path, and trilinear interpolation is used in the upsampling layer of the downsampled resolution path. The size of the receptive field of normal resolution path is $25 \times 25 \times 25$, and the size of the receptive field of the downsampled resolution path is $19 \times 19 \times 19$. The receptive field of downsampled resolution path is downsampled from an image patch of size $55 \times 55 \times 55$ by a factor of 3 in the same center as the receptive field of normal resolution path. The modified DeepMedic predicts the central $9 \times 9 \times 9$ voxels of the receptive field of normal resolution path.

Training and Test Procedure

The modified DeepMedic is only trained with patches that have approximately 50% foreground (lesion) and 50% background to solve the class imbalance problem, and it is trained with batch size 50. In every epoch, 20 patches are extracted from each subject. The network is trained for a total of 500 epochs. The weights of the network are updated by Adam algorithm [71] with an initial learning rate of $l_0 = 10^{-3}$ following the schedule of $l_0 \times 0.1^{\text{epoch}}$, L2 penalty weight decay of 10^{-4}, and AMSGrad [72]. A standard multi-class cross-entropy loss is used. Randomly flipping in x-, y-, and z-axis with a probability of 50%, and random noise are applied in the data augmentation of the training procedure. At the test time, a sliding window scheme of step size 9 is used to get the tumor lesion prediction of the test subject. Training takes approximately 6 hours, and a test for each subject takes approximately 24 seconds on an Nvidia 1080 Ti GPU and an Intel Xeon CPU E5-2696 v4 @ 2.20 GHz.

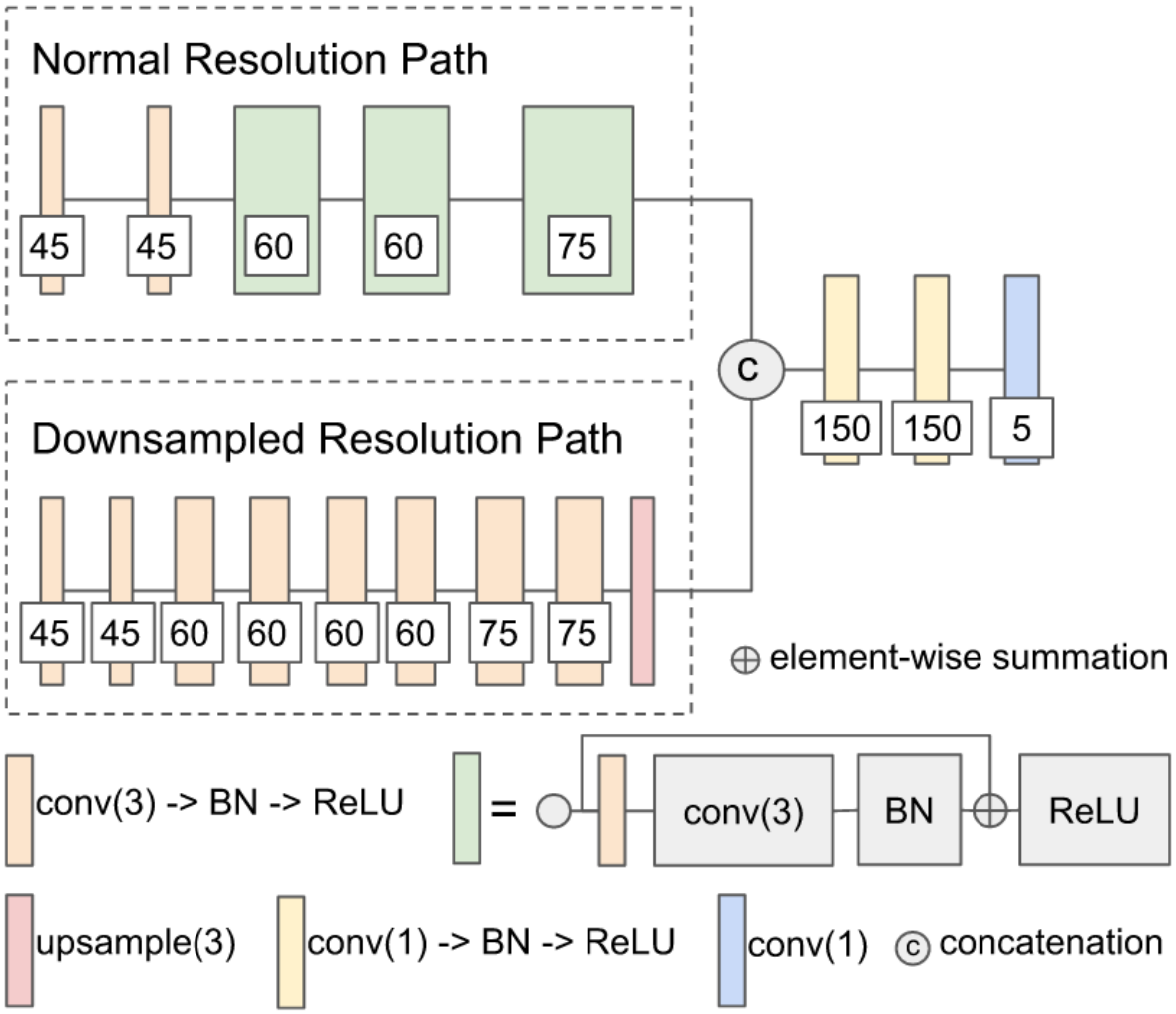


Figure 3.1: The network architecture of modified DeepMedic. conv(3): $3 \times 3 \times 3$ convolutional layer, BN: batch normalization, upsample(3): trilinear interpolation by a factor of 3, and conv(1): $1 \times 1 \times 1$ convolutional layer.

3.4.2 Modified 3D U-Net

The second network architecture shown in Figure 3.2 is modified from 3D U-Nets [26]. Different colors of blocks represent different types of layers. The number of convolutional kernels is indicated within the white box. Group normalization [73] is used, and the number of groups is set to 4. Residual connection [70] is used in the encoding path, and trilinear interpolation is used in the upsampling layer.

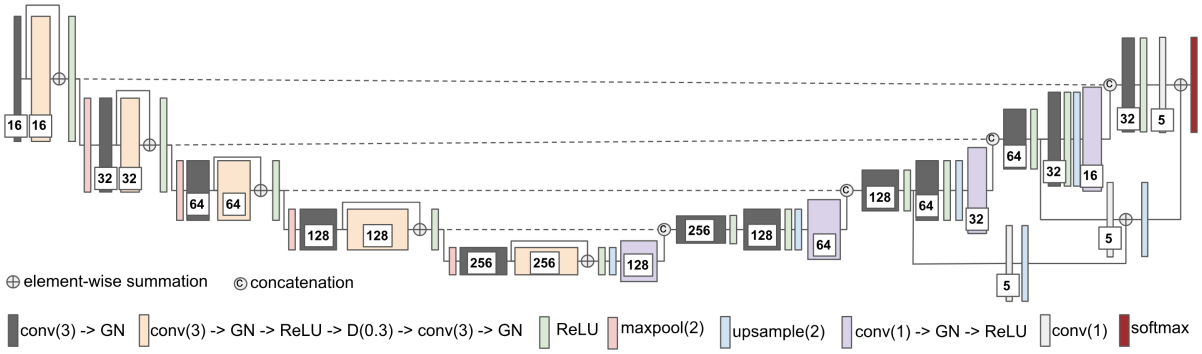


Figure 3.2: The network architecture of modified 3D U-Net. conv(3): $3 \times 3 \times 3$ convolutional layer, GN: group normalization, D(0.3): dropout layer with 0.3 dropout rate, maxpool(2): $2 \times 2 \times 2$ max pooling layer, and conv(1): $1 \times 1 \times 1$ convolutional layer.

Training and Test Procedure

The modified 3D U-Net is trained with randomly cropped patches of size $128 \times 128 \times 128$ voxels and batch size 2. In every epoch, a cropped patch is randomly extracted from each subject. The network is trained for a total of 300 epochs. The weights of the network are updated by Adam algorithm [71] with an initial learning rate $l_0 = 10^{-3}$ following the schedule of $l_0 \times 0.1^{\text{epoch}}$, L2 penalty weight decay of 10^{-4}, and AMSGrad [72]. For the loss function, the standard multi-class cross-entropy loss with the hard negative mining is used to solve the class imbalance problem of the dataset. We only back-propagate the negative (background) voxels with the largest losses (hard negative) and the positive

(lesions) voxels to the gradients. In our implementation, the number of selected negative voxels is at most three times more than the number of positive voxels. Besides, data augmentation is not used for both training and testing. At the test time, we input the entire image of size $240 \times 240 \times 155$ voxels into the trained 3D U-Net for each test patient to get the predicted lesion mask. Training takes approximately 12.5 hours, and the test takes approximately 1.5 seconds per subject on an Nvidia 1080 Ti GPU and an Intel Xeon CPU E5-2696 v4 @ 2.20 GHz.

3.5 Incorporating Location Information

The heatmaps (see Figure 3.3) of different brain tumor lesions reveal that different lesions have different probability occurring in different locations. However, the patch-based convolutional neural networks (CNNs), e.g., DeepMedic or 3D U-Net, do not consider location information for brain tumor segmentation.

In this study, an existing brain parcellation atlas, Harvard-Oxford Subcortical atlas [6] (see Figure 3.3), is used as location information for the patch-based CNNs. The details of Harvard-Oxford Subcortical parcellation regions are described in Table. 3.1. There are two main reasons for choosing this atlas: (1) this atlas covers more than 90% of a brain region, and (2) lesion information and location information are converted into this atlas. The distribution in Figure 3.3 is calculated by dividing the total volume of the lesions from 285 training subjects by the total volume of the corresponding brain parcellation in the MNI 152 space.

Our proposed location information fusion method which is shown in Figure 3.4 explicitly includes location information as input into a patch-based CNN. First, the Harvard-Oxford subcortical atlas is registered to the individual subject space from MNI 152 1mm space [44] using FMRIB's Linear Image Registration Tool (FLIRT) [51] from FSL. The

registered atlas is then split into 21 binary masks and concatenated with the multimodal MR images as input to a patch-based CNN for both training and test. As a result, the fused input has 25 channels. The first four channels provide the image information, and the last 21 channels contain the location information.

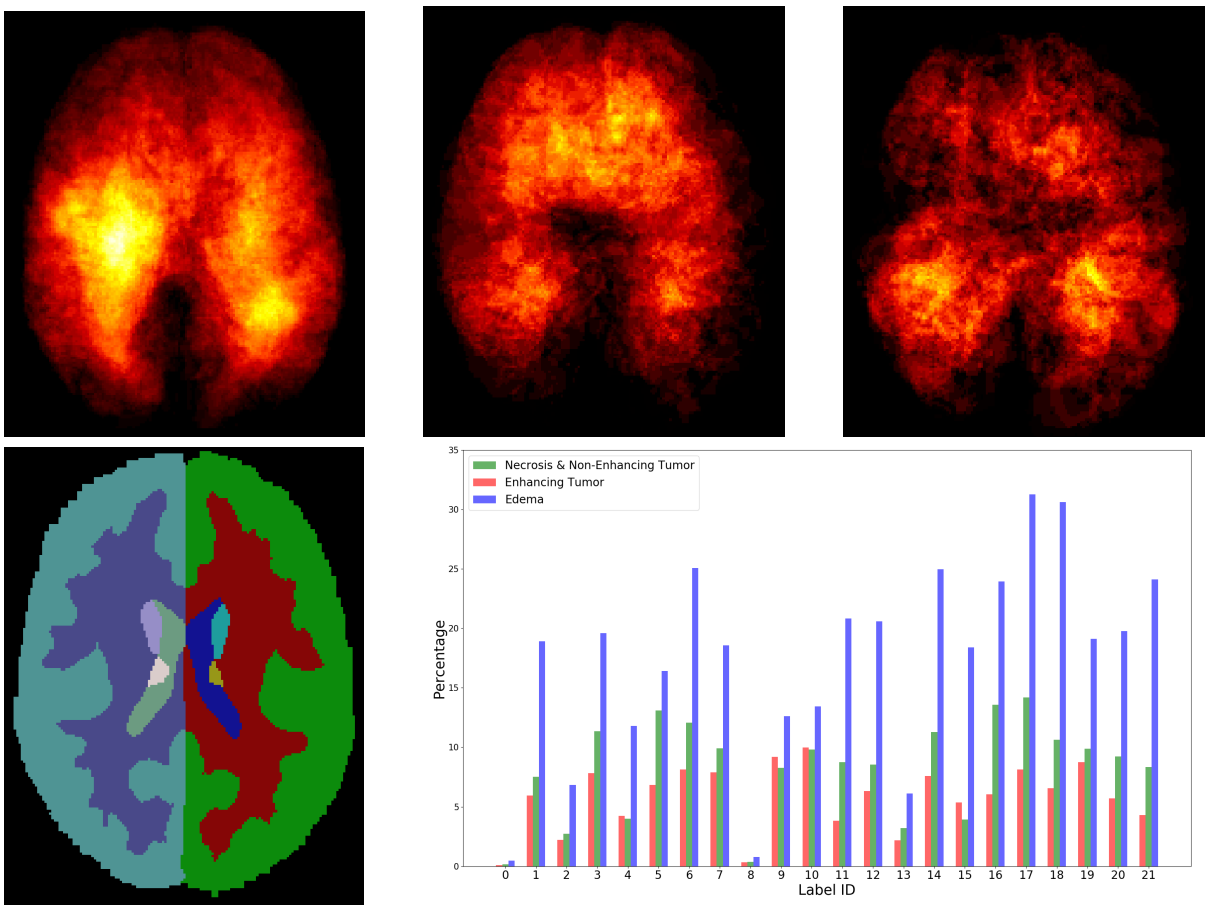


Figure 3.3: Top row shows the heatmaps of different lesions (from left to right: edema, necrosis & non-enhancing tumor, and enhancing tumor) from 285 training subjects of BraTS 2018. The brighter (yellow) voxel represents higher value. Bottom row shows (left) Harvard-Oxford subcortical structural atlas [6], and (right) the percentage of brain lesion types observed in different parcellation regions of the Harvard-Oxford subcortical atlas from 285 training subjects of BraTS 2018. The x-axis indicates the brain parcellation label ID. Regions not covered by the Harvard-Oxford subcortical atlas are in label 0.

Table 3.1: The label ID and corresponding brain region of Harvard-Oxford Subcortical Atlas.

Label ID	Brain region
1	Left Cerebral White Matter
2	Left Cerebral Cortex
3	Left Lateral Ventrical
4	Left Thalamus
5	Left Caudate
6	Left Putamen
7	Left Pallidum
8	Brain-Stem
9	Left Hippocampus
10	Left Amygdala
11	Left Accumbens
12	Right Cerebral White Matter
13	Right Cerebral Cortex
14	Right Lateral Ventricle
15	Right Thalamus
16	Right Caudate
17	Right Putamen
18	Right Pallidum
19	Right Hippocampus
20	Right Amygdala
21	Right Accumbens

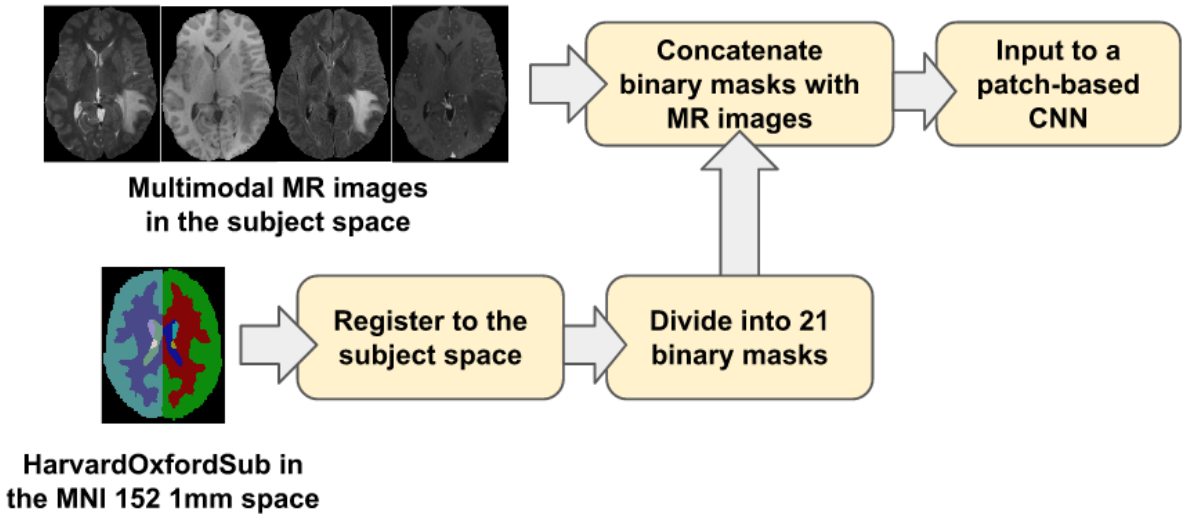


Figure 3.4: The proposed location information fusion method for brain tumor segmentation using a patch-based convolutional neural network.

3.6 Ensemble Methods

Ensemble methods aim at improving the predictive performance of a given statistical learning or model fitting technique. The general principle of ensemble methods is to construct a linear combination of some model fitting methods, instead of using a single fit of the method [74]. Ensembles have been proven to have better performance than any single model [75]. In the following, we propose a two-level ensemble approach, including the arithmetic mean and boosting, and more details of these methods are described below.

3.6.1 Arithmetic Mean

The arithmetic mean, $\bar{x}$, is the average of n values $x_1, x_2, ...x_n$, i.e., $\bar{x} = (x_1 + x_2 + ... + x_n)/n$. If we have n models in our ensemble, then the arithmetic mean P is defined by the formula:

$$P = \frac{1}{n}\sum_{i=1}^{n} p_i = \frac{p_1 + p_2 + ... + p_n}{n} \tag{3.1}$$

where p_i is the probability map of model i. The arithmetic mean ensemble method reduces the uncertainties of different models.

3.6.2 XGBoost

Boosting algorithms are widely used in machine learning to achieve state-of-art performance. It improves the prediction of the models by training the base learners sequentially to improve their predecessor. There are different boosting algorithms such as AdaBoost [76,77], short for Adaptive Boosting, and Gradient Boosting [78,79]. AdaBoost tunes the weights for every incorrect classified observation at every iteration while Gra-

dient Boosting tries to fit the new predictor to the residual errors made by the previous predictor. Both of the boosting algorithms are generally very slow in implementation and not very scalable. Chen et al. [80] described a scalable tree boosting system called XGBoost which is an implementation of gradient boosted decision trees that are efficient in run-time and space complexity. It also supports parallelization of tree construction, distributed computing for training very large models, out-of-core computing for very large datasets that do not fit into memory and cache optimization to make the best use of hardware. These features make XGBoost ideal for our purpose of study in brain tumor segmentation, therefore, it is used in our study.

3.6.3 Two-Level Ensemble Approach: Arithmetic Mean and XGBoost

An ensemble of multiple identical network architectures with different seed initializations has been proven to reduce the uncertainty of models and improve the segmentation performance [81]. Moreover, Dietterich [75] demonstrated that the boosting algorithm has the best performance compared to bagging and randomized trees. Inspired by their works, we propose a two-level ensemble approach shown in Figure 3.5 that averages the probability maps from the same type of models in the first level and then boosts the averaged probability maps from different models by using the XGBoost algorithm in the second level. We have examined three different classification strategies in the second level, and these classification strategies are based on multi-class classification and binary class classification.

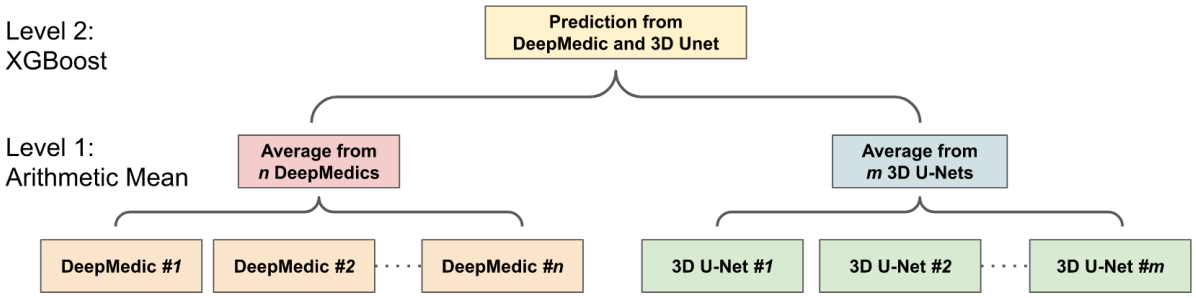


Figure 3.5: The workflow of two-level ensemble approach.

Multi-Class Classification

The multi-class classification problem refers to classifying voxels into one of the four classes. It produces segmentation labels of the background and different glioma lesions that include: 1) the "enhancing tumor", 2) the "edema", and 3) the "necrosis & non-enhancing tumor". Since XGBoost is known to produce better results in different machine learning problems [82], XGBoost is used in our multi-class classification problem with the softmax function objective. The softmax function σ is defined by

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad \text{for} \quad i = 1, \ldots, K \quad \text{and} \quad \mathbf{z} = (z_1, \ldots, z_K) \in \mathbb{R}^K.$$

where K is the number of classes in the classification problem. Using the softmax objective function, we get a neural network that models the probability of a class z_i as multinominal distribution.

Binary Classification

The multi-class classification problem can be reduced to several binary classification problems where each binary classifier is trained to classify voxels into two classes.

There are two different approaches, one-versus-all and one-versus-one, to perform such a transformation. For a k-class problem, one-versus-all method trains k different binary classifiers where the two-class classifier C_i learns to distinguish the class i from all the other $k - i$ classes.

$$C_+ = C_i \quad \text{and} \quad C_- = \{C_j | j = 1, \cdots, K,\ j \neq i\}$$

One-versus-one approach is based on training $k \times (k-1)/2$ classifiers, where each classifier learns to distinguish 2 classes only.

$$C_+ = C_i \quad \text{and} \quad C_- = \{C_j | j \neq i\}$$

where C_+ and C_- are the two classes of the binary class classification problem.

3.7 Evaluation Metrics

Two evaluation metrics, dice similarity score (DSC) and Hausdorff distance, are commonly used in the brain tumor segmentation problem. DSC is used to measure the similarity of the predicted lesions and ground-truth lesions, and Hausdorff distance is used to measure how far the predicted lesions are from the ground-truth lesions. More details of these two evaluation metrics are described in the following sections.

3.7.1 Dice Similarity Score

Dice similarity score (DSC) is a statistic used to measure the similarity of two sets. It is defined as

$$DSC = \frac{2|G \cap P|}{|G| + |P|} \tag{3.2}$$

where $|G|$ and $|P|$ are the number of voxels in the ground-truth and prediction, respectively. DSC ranges between 0 to 1 (1 means perfect matching).

3.7.2 Hausdorff Distance

Hausdorff distance $d_H(X, Y)$ measures how far two subsets $\{X, Y\}$ of a metric space are from each other. It is defined as

$$\mathrm{d_H}(X, Y) = \max\{\sup_{x \in X} \inf_{y \in Y} \mathrm{d}(x, y), \sup_{y \in Y} \inf_{x \in X} \mathrm{d}(x, y)\} \tag{3.3}$$

where d is the Euclidean distance, sup is the supremum, and inf is the infimum. Hausdorff distance ranges from 0 to infinity (0 means perfect matching). In this study, 95 percentile of Hausdorff distance (HD95) is used to disregard the outliers.

3.8 Experiments and Results

In this section, we demonstrate the advantage of the proposed location information fusion method and the proposed two-level ensemble learning method. In Experiment 1, we first examine the segmentation performance of the proposed location information fusion method on a single model. In Experiment 2, we examine the performance of the proposed location information fusion method on an ensemble of the same type of models.

In Experiment 3, we examine different ensemble methods which predict the final brain tumor lesions based on the output probability maps from DeepMedics and 3D U-Nets. In Experiment 4, we compare the segmentation performance of the proposed method with state-of-the-art methods.

3.8.1 Experiment 1: Location information fusion method on a single model

In the first experiment, we examine the performance of the proposed location information fusion method on a single patch-based neural network. We first train a DeepMedic and a 3D U-Net using only multimodal MR images. Thereafter, we train another identical DeepMedic and another identical 3D U-Net with multimodal MR images and binary brain parcellation masks. BraTS 2018 training set is used to train the models with 5-fold cross-validation, and BraTS 2018 validation set is used as the test set. The experimental results are shown in Table 3.2. Here, the proposed location fusion method improves the brain tumor segmentation performance of both single state-of-the-art models.

Table 3.2: Results of the first experiment on BraTS 2018 validation set. The result are reported as mean. Bold numbers highlight the improved results with additional brain parcellation masks within the same type of model.

	DSC			HD95		
Model Description	ET	WT	TC	ET	WT	TC
DeepMedic	78.1	89.5	81.4	4.21	10.60	9.90
DeepMedic + BP	**79.0**	**89.6**	81.3	**3.78**	**8.87**	**6.55**
3D U-Net	74.9	89.7	76.6	5.85	4.88	10.46
3D U-Net + BP	**76.4**	**90.1**	**76.9**	**5.48**	**4.87**	**10.07**

3.8.2 Experiment 2: Location information fusion method on an ensemble

In the second experiment, we examine the performance of the proposed location information fusion on the ensemble of DeepMedics and the ensemble of 3D U-Nets. Each ensemble has identical network architectures with different seed initializations, and the output of the ensemble is the arithmetic mean from networks. We first train ensembles of DeepMedics without additional brain parcellation masks. Thereafter, we train ensembles of 3D U-Nets without additional brain parcellation masks. In the end, we train another identical ensemble of DeepMedics and another identical ensemble of 3D U-Nets with additional brain parcellation masks. BraTS 2018 training set is used to train the models with 5-fold cross-validation, and BraTS 2018 validation set is used as the test set. The experimental results are shown in Table 3.3. Here, the proposed location fusion method improves the ensemble of multiple same types of state-of-the-art models

Table 3.3: Results of the second experiment on BraTS 2018 validation set. The result are reported as mean. Bold numbers highlight the improved results with additional brain parcellation masks within the same type of ensemble.

	DSC			HD95		
Ensemble Description	ET	WT	TC	ET	WT	TC
DeepMedic	79.7	90.0	81.4	3.94	7.44	8.88
DeepMedic + BP	78.4	**90.2**	**81.8**	**3.37**	**5.64**	**7.01**
3D U-Net	77.6	90.0	78.0	5.01	4.39	9.77
3D U-Net + BP	77.4	**90.4**	**79.3**	**4.25**	4.59	**9.66**

3.8.3 Experiment 3: Different ensemble methods

In the third experiment, we examine the performance of different ensemble methods including arithmetic mean and two-level ensemble approaches described in Section. 3.6. We first train three identical DeepMedics with additional brain parcellation channels and

different seed initializations. We also train three identical 3D U-Nets with additional brain parcellation channels and different seed initializations. Then, we apply different ensemble methods on the probability maps from these models to generate the final tumor segmentation mask. More details of different ensemble methods are described below.

Experiment 3.1: Arithmetic Mean

In this experiment, the final tumor segmentation mask is directly generated by averaging the probability maps from three DeepMedics and three 3D U-Nets. BraTS 2018 training set is used to train the models with 5-fold cross-validation, and BraTS 2018 validation set is used as the test set. The experimental results are shown in Table 3.4.

Experiment 3.2: Two-Level Ensemble: Multi-Class Classification

In this experiment, we directly apply an XGBoost classifier on the probability maps from three DeepMedics and three 3D U-Nets. The input vector of XGBoost classifier has 10 dimensions (5-class probability maps from 2 ensembles of the same type of models). The XGBoost classifier outputs the 5-class labels which contain a background (label 0), enhancing tumor (label 1), edema (label 2), and necrosis & non-enhancing tumor (label 4). BraTS 2018 training set is used to train the models with 5-fold cross-validation, and BraTS 2018 validation set is used as the test set. The experimental results are shown in Table 3.4 as TLMC.

Experiment 3.3: Two-Level Ensemble: Binary Classification

In this experiment, we train three XGBoost binary classifiers on the resulting probability maps generated from three DeepMedics and three 3D U-Nets in the first level. During training, each classifier uses a one-versus-one approach to distinguish between

two binary classes. We trained three different models namely, model_WT (whole tumor), model_TC (tumor core) and model_ET (enhancing tumor) as shown in Figure 3.6.

$$\begin{aligned}
&\text{For model_WT:} \quad C_+ = C_{WT} \quad \text{and} \quad C_- = C_{background} \\
&\text{For model_TC:} \quad C_+ = C_{TC} \quad \text{and} \quad C_- = C_{WT} \\
&\text{For model_ET:} \quad C_+ = C_{ET} \quad \text{and} \quad C_- = C_{TC}
\end{aligned}$$

The tumor core class is a subset class of the whole tumor, and the enhancing tumor class is a subset of the tumor core class. For prediction, we feed the average probability maps from three DeepMedics and three 3D U-Nets to the three models. The input vector has 10 dimensions (5-class probability maps from 2 ensembles of the same type of models). The model_WT classifies the voxels into the whole tumor and background. For model_TC, we feed the probability maps of such voxels that are classified as the whole tumor from the experiment in Section. 3.8.3. For model_ET, we feed the probability maps of such voxels that are classified as tumor core from the previous prediction in the experiment. BraTS 2018 training set is used to train the models with 5-fold cross-validation, and BraTS 2018 validation set is used as the test set. The experimental results are shown in Table 3.4 as TLBC.

Experiment 3.4: Two-Level Ensemble: Fusion Classifications

This is the final experiment to integrate the methods from the previous experiments. We observe that while the experiment in Section. 3.8.3 performs best for classifying voxels into the background, whole tumor and tumor core, the experiment in Section. 3.8.3 has the best performance on necrosis & non-enhancing tumor. We use model_WT, model_TC and multi-class classifier model for the fusion model. For prediction, we feed the average

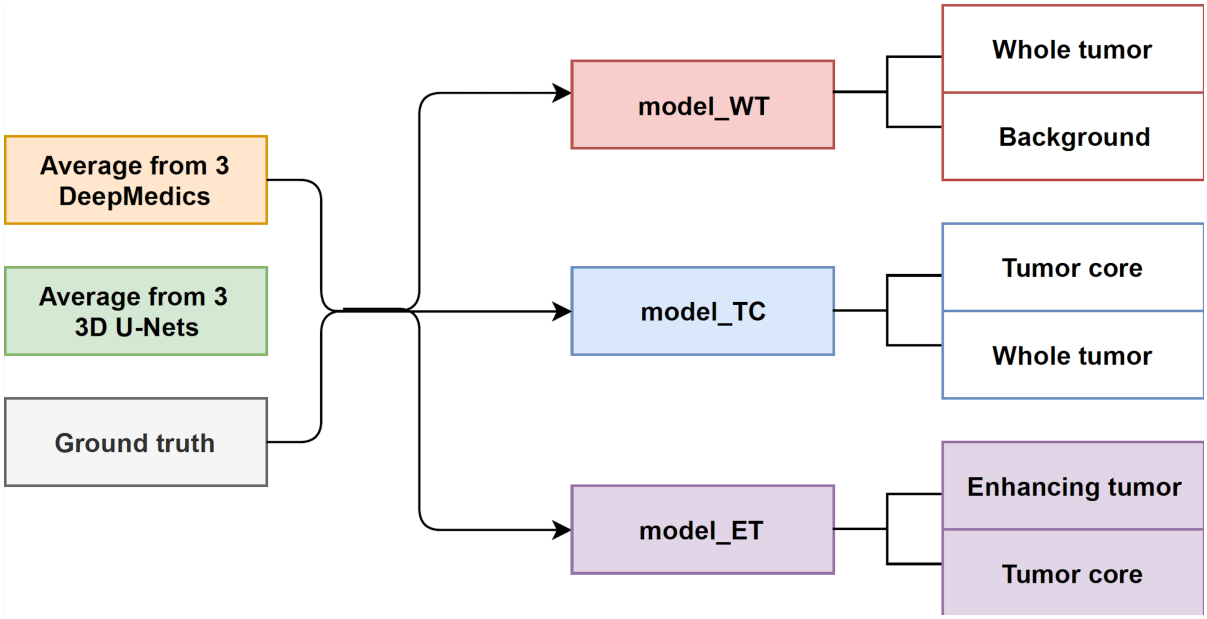


Figure 3.6: The training workflow of two-level binary classification approach.

probability maps from three DeepMedics and three 3D U-Nets to the three models. The input vector has 10 dimensions (5-class probability maps from 2 ensembles of the same type of models). The model_WT classifies the voxels into the whole tumor and background. For model_TC that is trained to classify voxels into the whole tumor and tumor core, we feed the probability maps of such voxels that are classified as the whole tumor from the experiment in Section. 3.8.3. For necrosis & non-enhancing tumor class, we feed the probability maps to the multi-class classifier as in Section. 3.8.3. To merge the three different predicted results, we classify the voxels into four classes with background, whole tumor, tumor core, and necrosis & non-enhancing tumor in decreasing order of priority. For example, if a voxel is classified into both whole tumor and tumor core, we give the final label as that of tumor core according to the preference mentioned before. Therefore integrating these two gives the effective scores as shown in Table 3.4 as TLFC. BraTS 2018 training set is used to train the models with 5-fold cross-validation, and BraTS 2018 validation set is used as the test set. The workflow of fusion classification is

shown in Figure 3.7. Here, the proposed ensemble method, two-level fusion classification (TLFC) method, has the best performance compared to other ensemble methods including arithmetic mean, two-level multi-class classification (TLMC), and two-level binary classification (TLBC).

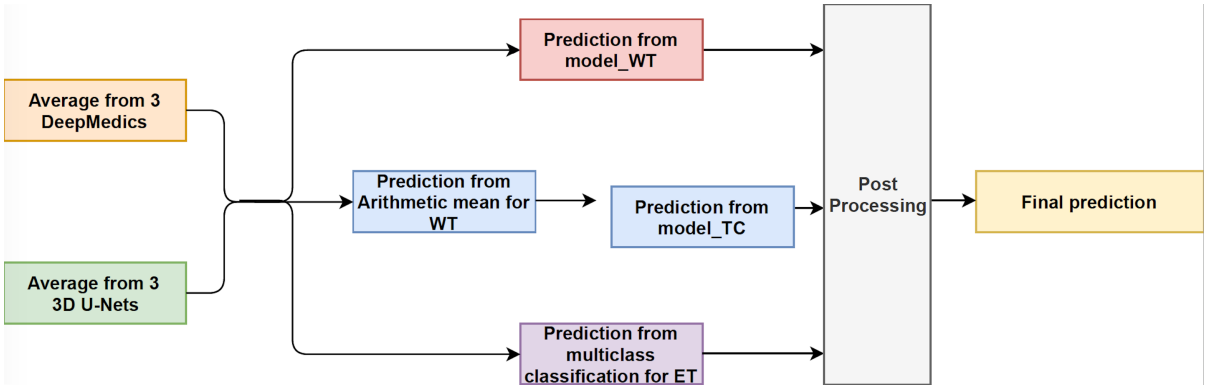


Figure 3.7: The workflow of Fusion Classification method. For post processing step, we classify the voxels into four classes with background, whole tumor, tumor core and necrosis & non-enhancing tumor in decreasing order of priority. For example, if a voxel is classified into both whole tumor and tumor core, we give the final label as that of tumor core according to the preference mentioned before.

Table 3.4: Results of the third experiment on BraTS 2018 validation set. The results are reported as mean. Bold numbers highlight the best performance between different ensemble methods.

	DSC			HD95		
Ensemble Methods	ET	WT	TC	ET	WT	TC
Arith. mean	**78.3**	90.6	81.3	3.72	**4.35**	7.77
TLMC	**78.3**	90.7	81.0	**2.81**	4.38	7.80
TLBC	76.6	90.7	82.2	7.93	4.39	8.34
TLFC	78.2	**90.8**	**82.3**	2.96	4.39	**6.91**

3.8.4 Experiment 4: Compare to the State-of-the-art Methods

In this experiment, we compare the brain tumor segmentation performance of the proposed method described in Section. 3.8.3 with the state-of-the-art methods on both

BraTS 2017 and BraTS 2018 dataset. The quantitative results are shown in Table. 3.5. The proposed method has the best tumor segmentation performance compared to other state-of-the-art methods in BraTS 2017 with a similar number of models in the ensemble. Also, the proposed method has a competitive tumor segmentation performance compared to other state-of-the-art methods in BraTS 2018 with fewer models in the ensemble.

Table 3.5: The first three rows show the results of our proposed method and the state-of-the-art methods on BraTS 2017 validation set, and the bottom four rows show the results of our proposed method and the state-of-the-art methods on BraTS 2018 validation set. The results are reported as mean. Bold numbers highlight the best performance in each dataset. These results are directly copied from their paper.

Methods	# of models	DSC			HD95		
		ET	WT	TC	ET	WT	TC
Kamnitsas et al. [56]	7	73.8	90.1	**79.7**	4.50	**4.23**	**6.56**
Isensee et al. [58]	5	73.2	89.6	**79.7**	4.55	6.97	9.48
Proposed method	6	**74.3**	**90.4**	78.5	**3.49**	4.46	8.45
Myronenko [59]	10	**82.3**	**91.0**	**86.6**	3.93	4.52	**6.85**
Isensee et al. [60]	10	81.0	90.8	85.4	**2.54**	4.97	7.04
Kao et al. [63]	26	78.8	90.5	81.3	3.81	**4.32**	7.56
Proposed method	6	78.2	90.8	82.3	2.96	4.39	6.91

3.9 Discussion and Summary

Due to the computational limitation, we are not able to input the whole brain volume of size $240 \times 240 \times 155$ to a neural network for training purpose. Alternatively, we randomly crop sub-regions of the brain and input these sub-regions to the neural network for training. For the current patch-based neural networks, we noted that these neural networks lack location information of the brain for both training and test procedure. That is, these patch-based neural networks do not have the information about where the patch comes from the brain. Therefore, we proposed the location fusion method which explicitly carries location information of the brain into patch-based neural networks such as 3D U-Net and DeepMedic. An existing structural brain parcellation atlas, HarvardOxford Sub-cortical Atlas, is used as additional location information to these patch-based neural networks in both training and test.

From Table 3.2, we demonstrate that the proposed location fusion method improves the brain tumor segmentation performance of both single state-of-the-art model. Also, we demonstrate that the proposed location fusion method improves the ensemble of multiple same types of state-of-the-art models in Table 3.3. The proposed location fusion method yields a smoother prediction for both 3D U-Net and DeepMedic compared to the resulting prediction without location information (see Figure 3.8 and 3.9).

From Table 3.4, the proposed ensemble method, two-level fusion classification (TLFC) method, has the best performance compared to other ensemble methods including arithmetic mean, two-level multi-class classification (TLMC), and two-level binary classification (TLBC). TLFC takes advantages of TLMC and TLBC. Moreover, Figure 3.10 shows the predictions of brain tumor lesions from different ensemble methods, and TLFC method has the best performance among other methods.

From Table 3.5, the proposed method has the best tumor segmentation performance

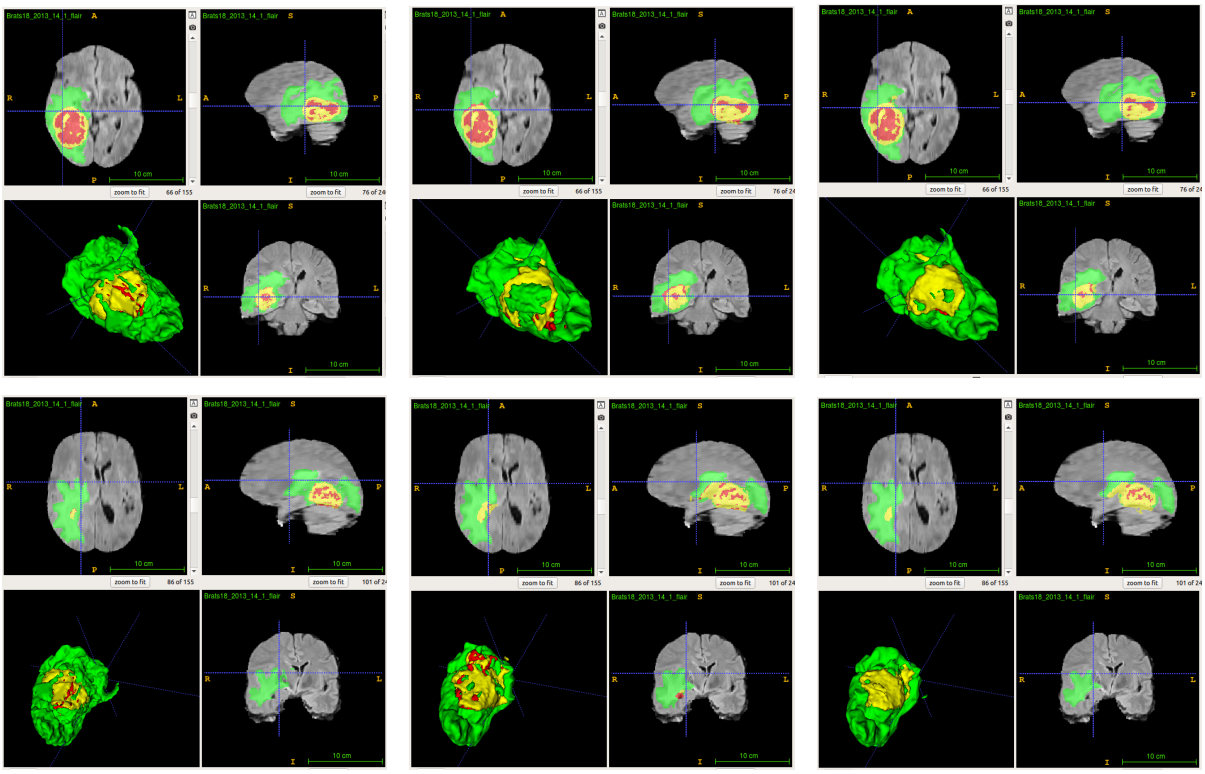


Figure 3.8: Examples of predictions from single model with different inputs. Top row shows the predictions from DeepMedic, and bottom row shows the predictions from 3D U-Net (from left to right: ground-truth lesions, prediction from single model, and prediction from single model with additional brain parcellation masks.) Red: enhancing tumor, yellow: necrosis & non-enhancing tumor, and green: edema. ITK-SNAP [7] is used to visualize the MR images and lesion masks.

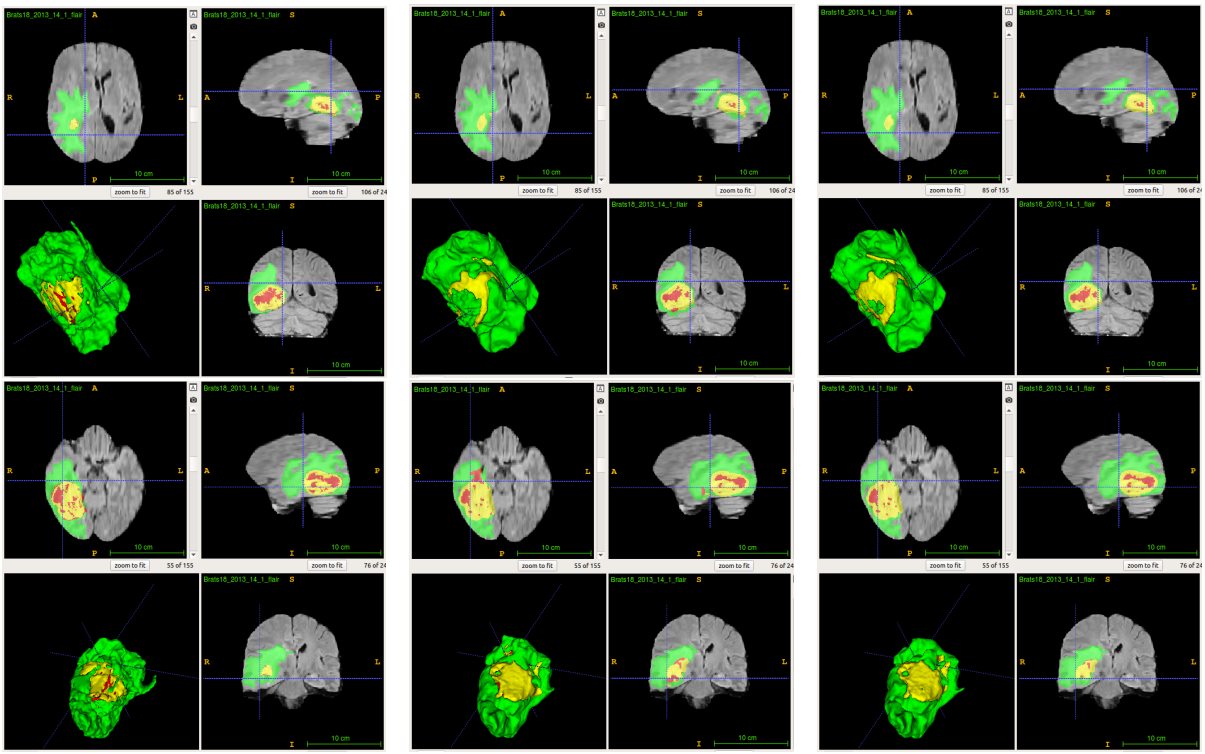


Figure 3.9: Examples of predictions from ensemble with different inputs. Top row shows the predictions from ensemble of DeepMedics, and bottom row shows the predictions from ensemble of 3D U-Nets (from left to right: ground-truth lesions, prediction from ensemble, and prediction from ensemble with additional brain parcellation masks.) Red: enhancing tumor, yellow: necrosis & non-enhancing tumor, and green: edema. ITK-SNAP [7] is used to visualize the MR images and lesion masks. four classes

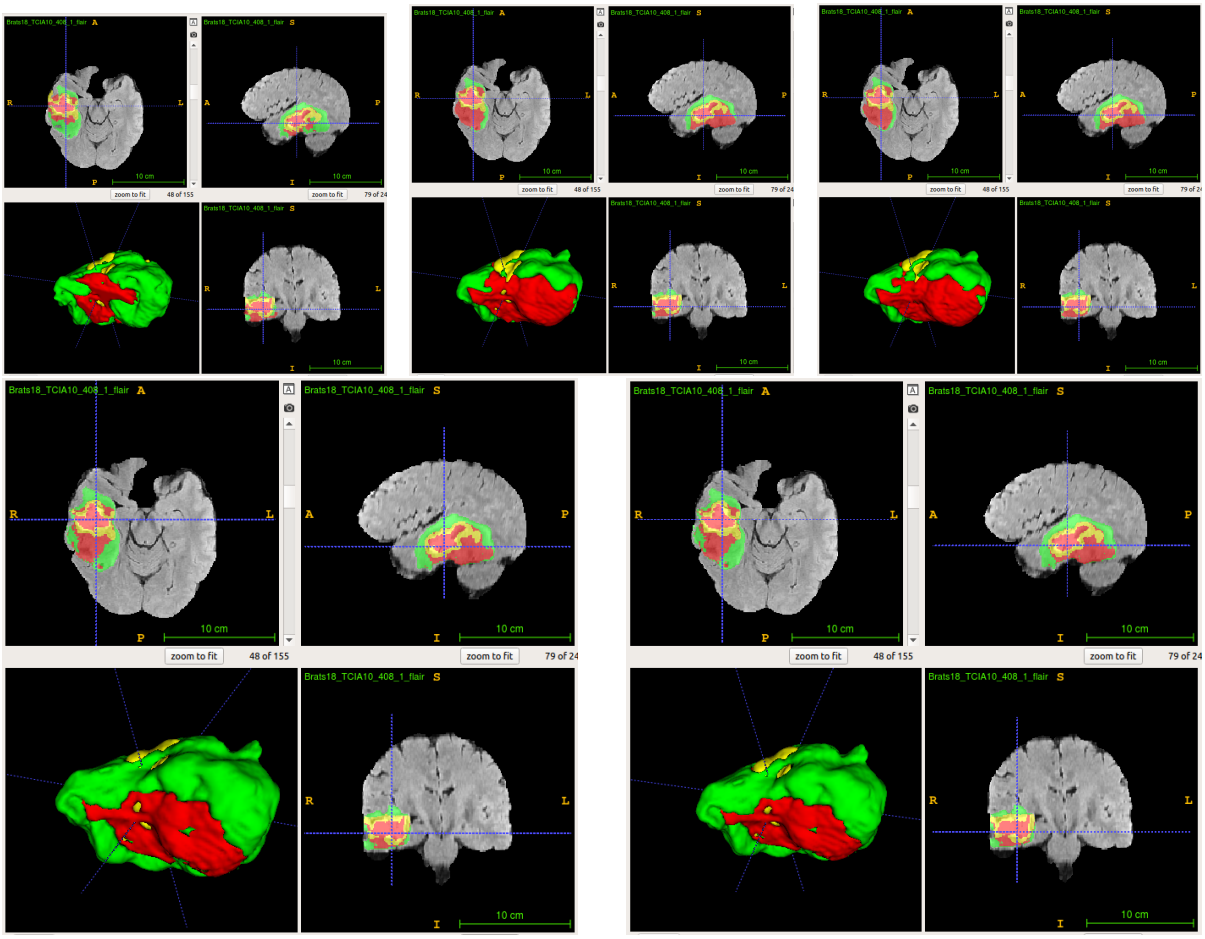


Figure 3.10: Examples of predictions from different ensemble methods. Top left image shows the ground-truth lesion mask, and top middle image shows the predictions using arithmetic mean. Top right image shows the prediction using two-level multi-class classification (TLMC) method. Bottom left image shows the prediction using two-level binary classification (TLBC) method, and bottom right image shows the prediction using two-level fusion classification (TLFC) method. Red: enhancing tumor, yellow: necrosis & non-enhancing tumor, and green: edema. ITK-SNAP [7] is used to visualize the MR images and lesion masks. four classes

compared to other state-of-the-art methods in BraTS 2017 with a similar number of models in the ensemble. Also, the proposed method has a competitive tumor segmentation performance compared to other state-of-the-art methods in BraTS 2018 with fewer models in the ensemble. It is noted that the model of Myronenko [59] requires a huge amount of GPU memory (32 GB) for training, and Isensee et al. [60] trained the models with additional public and institutional data. In addition, Myronenko [59] and Isensee et al. [60] have 10 models in their ensemble but our proposed ensemble only has 6 models. The proposed ensemble has much fewer models with a better segmentation performance compared to our previous work which has 26 models [63]. The test time of our previous ensemble takes approximately 30 minutes on an Nvidia 1080 Ti GPU and an Intel Xeon CPU E5-2696 v4 @ 2.20 GHz. However, the proposed ensemble only takes approximate 3 minutes on the same infrastructure. Our previous ensemble rank 6th out of 63 teams in BraTS 2018 segmentation challenge, and the proposed ensemble even has a better performance and less interference time compared to the previous ensemble.

In summary, we proposed a novel method to integrate location information about the brain into a patch-based neural network for improving brain tumor segmentation. Our experimental results demonstrate that the proposed location information fusion approach improves the segmentation performance of the baseline models including DeepMedic and 3D U-Net. Moreover, the proposed location information fusion method can be easily integrated with other patch-based network architectures to potentially enhance their brain tumor segmentation performance. We also proposed a two-level fusion classification method which reduces the uncertainty of prediction in the first level and takes advantage of different types of models in the second level. Also, the proposed ensemble method can be easily integrated with more different types of neural networks. The proposed ensemble helps the neurologists on delineating brain tumor and improves the quality of the neuro-surgery.

Chapter 4

Overall Survival Prediction using Tractographic Feature

> Many of life's failures are people who did not realize how close they were to success when they gave up.
>
> *Thomas A. Edison*

This chapter focuses on predicting the overall survival of brain tumor patients given the lesion region. We introduce a novel methodology to integrate human brain connectomics, brain parcellation information, and the brain lesion mask for overall survival (OS) prediction. We leverage the average connectome information from the Human Connectome Project (HCP) and map each subject brain volume onto this common connectome space. From this, we compute tractographic features that describe potential neural disruptions due to the brain tumor lesion. These features are then used to predict the overall survival of the subjects. The main novelty in the proposed methods is the use of normalized brain parcellation data and tractography data from the Human Connectome

Project for analyzing MR images for segmentation and survival prediction. Experimental results are reported on the BraTS 2018 dataset.

4.1 Introduction

Glioblastomas, or Gliomas, are one of the most common types of brain tumor. They have a highly heterogeneous appearance and shape and may happen at any location in the brain. High-grade glioma (HGG) is one of the most aggressive types of brain tumor with a median survival of 15 months [83]. We introduce a novel methodology to predict overall survival of brain tumor patients using tractographic feature. The proposed pipeline consists of two steps: In the first step, we utilize the brain tumor segmentation method proposed in Chapter 3 to automatically segment the brain tumor lesion from structural MR images. In the second step, we extract the tractographic feature to predict the overall survival of brain tumor patients.

Shboul et al. [84] extracted 40 features from the predicted brain tumor mask and used a random forest regression to predict the glioma patient's OS. Jungo et al. [85] extracted four features from each subject and used a support vector machine (SVM) with a radial basis function (RBF) kernel to classify glioma patients into three different OS groups. We propose a novel method to extract the tractographic features from the lesion regions on structural MR images via an average diffusion MR image which is from a total of 1021 HCP subjects [86] (Q1-Q4, 2017). We then use these tractographic features to predict the patient's OS with an SVM classifier with a linear kernel.

4.2 Dataset

The BraTS 2018 dataset also includes the age (in years), survival (in days) and resection status for each of 163 subjects in the training dataset, and 59 of them have the resection status of Gross Total Resection (GTR). The validation dataset has 53 subjects with the age (in years) and resection status, and 28 of them have the resection status of GTR. The test dataset has 131 subjects with the age (in years) and resection status, and 77 of them have the resection status of GTR. For this task, we only predict the overall survival (OS) for glioma patients with resection status of GTR. That is, 59, 28, and 77 subjects are considered in the training, validation, and test set, respectively.

4.3 Overall Survival Prediction using Tractographic Feature

Our proposed training pipeline, shown in Figure 4.1, includes three stages: In the first stage, we use the proposed ensemble from the section 3 to obtain the predicted tumor mask for each subject. In the second stage, we extract the tractographic features explained in Section. 4.3.2 from each subject. We then perform feature normalization and selection. In the final stage, we train an support vector machine (SVM) classifier with a linear kernel using the tractographic features extracted from the training subjects. We evaluate the overall survival classification performance of tractographic features on the BraTS 2018 training dataset with the 1000-time repeated stratified 5-fold cross-validation, validation dataset, and test dataset.

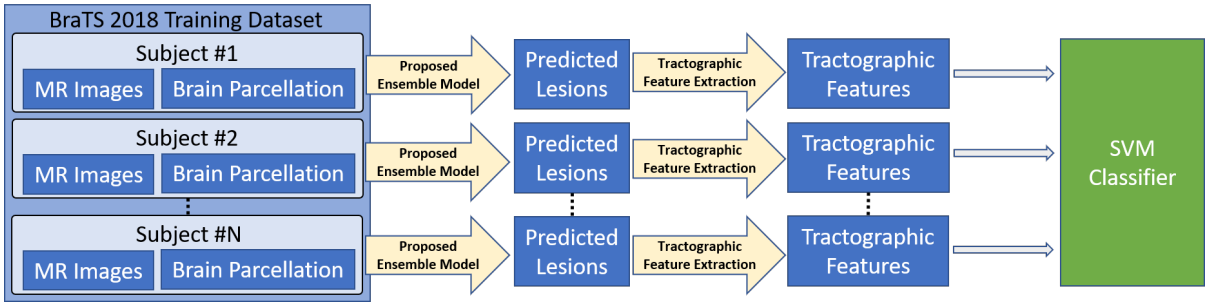


Figure 4.1: Training pipeline for overall survival prediction of brain tumor patients.

4.3.1 Glioma Segmentation

To segment the glioma, we use the proposed ensemble in Chapter 3 to obtain the prediction of three different types of the lesion including necrosis & non-enhancing tumor, edema, and enhancing tumor.

4.3.2 Tractographic Feature Extraction from the Glioma Segmentation

After we obtain the predicted lesion mask, we extract the tractographic feature from the whole tumor region which is the union of all different lesions for each subject.

Tractographic Feature: Tractographic feature describes the potentially damaged parcellation regions impacted by the brain tumor through fiber tracking. Figure 4.2 shows the workflow for building a connectivity matrix for each subject. First, the predicted whole tumor mask and the average diffusion orientation distribution function from HCP-1021, created by QSDR [87], are obtained for each subject. FLIRT [51–53] from FSL is used to map the whole tumor mask from the subject space to the MNI 152 1mm space. Second, we use a deterministic diffusion fiber tracking method [54] to create approximately 1,000,000 tracts from the whole tumor region. Finally, a structural brain atlas is

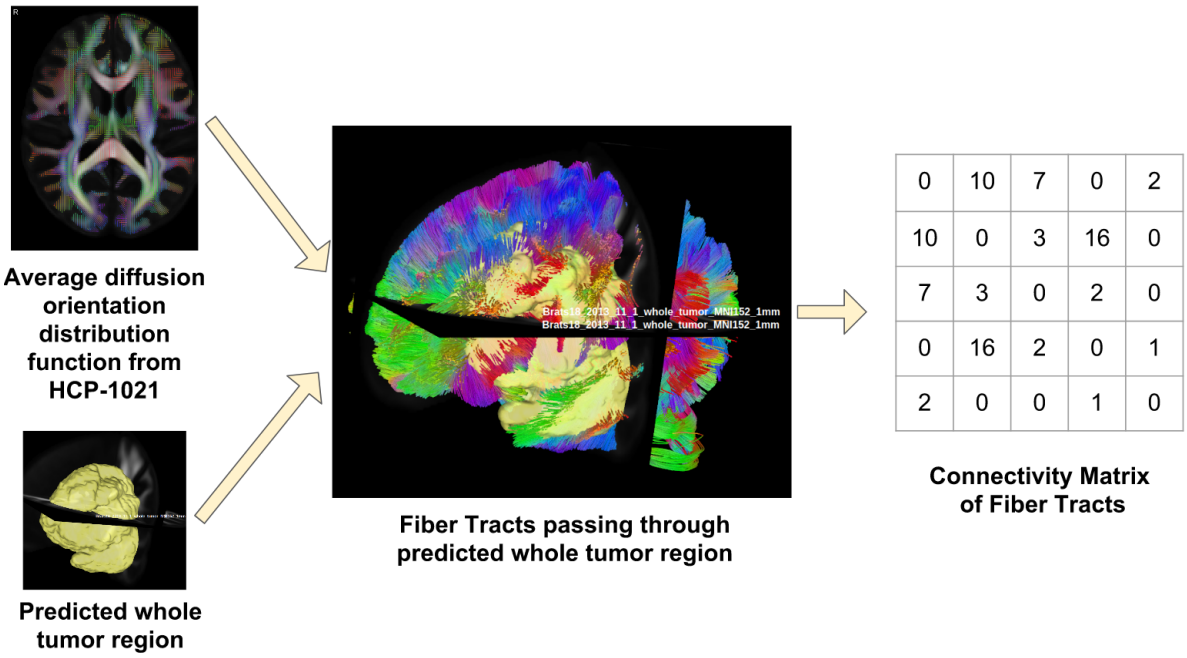


Figure 4.2: Workflow for building a connectivity matrix for each subject. The fiber tracts are created by DSI Studio (`http://dsi-studio.labsolver.org/`), and ITK-SNAP [8] is used for visualizing the 3D MR images and 3D labels.

used to create a connectivity matrix $\vec{W}_{ori}$ for each subject. This matrix contains information about whether a fiber connecting one region to another passed through or ended at those regions, as shown:

$\vec{W}_{ori}$ is a $N \times N$ matrix, and N is the number of parcellation in a structural brain atlas.

$$\vec{W}_{ori} = \begin{bmatrix} w_{ori,11} & w_{ori,12} & \dots & w_{ori,1N} \\ w_{ori,21} & w_{ori,22} & \dots & w_{ori,2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{ori,N1} & w_{ori,N2} & \dots & w_{ori,NN} \end{bmatrix} \tag{4.1}$$

If w_{ij} is pass-type, it shows the number of tracts passing through region j and region i. if w_{ij} is end-type, it shows the number of tracts starting from a region i and ending in a region j or starting from a region j and ending in a region i. From the original

connectivity matrix $\vec{W}_{ori}$, we create a normalized version $\vec{W}_{nrm}$ and a binarized version $\vec{W}_{bin}$.

$$\vec{W}_{nrm} = \vec{W}_{ori}/max(\vec{W}_{ori}) \tag{4.2}$$

/ is the element-wise division operator, and $max(\vec{W}_{ori})$ is the maximum value of the original connectivity matrix $\vec{W}_{ori}$.

$$\vec{W}_{bin} = \begin{bmatrix} w_{bin,11} & w_{bin,12} & \dots & w_{bin,1N} \\ w_{bin,21} & w_{bin,22} & \dots & w_{bin,2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{bin,N1} & w_{bin,N2} & \dots & w_{bin,NN} \end{bmatrix} \tag{4.3}$$

$w_{bin,ij} = 0$ if $w_{ori,ij} = 0$, and $w_{bin,ij} = 1$ if $w_{ori,ij} > 0$. Then, we sum up each column in a connectivity matrix to form a unweighted tractographic feature vector.

$$\vec{V} = \sum_{i=1}^{N} w_{ij} = \begin{bmatrix} v_1, v_2, \dots, v_N \end{bmatrix} \tag{4.4}$$

Furthermore, we weight every element in the unweighted tractographic feature vector with respect to the ratio of the lesion in a brain parcellation region to the volume of this brain parcellation region.

$$\vec{V}_{wei} = \vec{\alpha} \odot \vec{V}, \vec{\alpha} = \begin{bmatrix} t_1/b_1, t_2/b_2, \dots, t_N/b_N \end{bmatrix} \tag{4.5}$$

$\odot$ is the element-wise multiplication operator, t_i is the volume of the whole brain tumor in the i-th brain parcellation, and b_i is the volume of the i-th brain parcellation. This vector $\vec{V}_{wei}$ is the tractographic feature extracted from brain tumor.

In this problem, automated anatomical labeling (AAL) [5] is used for building the connectivity matrix. AAL has 116 brain parcellation regions, so the dimension of the

connectivity matrix $\vec{W}$ is 116×116 and the dimension of each tractographic feature $\vec{V}_{wei}$ is 1×116. In the end, we extract six types of tractographic features for each subject. Six types of tractographic features are computed from 1) the pass-type of the original connectivity matrix, 2) the pass-type of the normalized connectivity matrix, 3) the pass-type of the binarized connectivity matrix, 4) the end-type of the original connectivity matrix, 5) the end-type of the normalized connectivity matrix and 6) the end-type of the binarized connectivity matrix.

4.3.3 Feature Normalization and Selection

First, we remove features with the low variance between subjects and then apply a standard z-score normalization on the remaining features. In the feature selection step, we combine recursive feature elimination with the 1000-time repeated stratified 5-fold cross-validation and an SVM classifier with a linear kernel. These feature processing steps are implemented by using scikit-learn [88].

4.3.4 Overall Survival Prediction

We first divide all 59 training subjects into three groups: long-survivors (>15 months), short-survivors (<10 months), and mid-survivors (between 10 and 15 months). Then, we train a SVM classifier with linear kernel on all training subjects with 1000-time repeated stratified 5-fold cross-validation to evaluate the performance of the proposed tractographic feature on overall survival prediction for brain tumor patients. We also evaluate the OS prediction performance of tractographic feature on the BraTS 2018 validation and test dataset.

4.4 Experimental Results and Discussion

We first examine the overall survival classification performance of our proposed tractographic feature compared to other types of features including age, volumetric features, spatial features, volumetric-spatial features, and morphological features.

Volumetric Features: The volumetric features include the volume and the ratio of brain to the different types of lesions, as well as the tumor compartments. 19 volumetric features are extracted from each subject.

Spatial Features: The spatial features describe the location of the tumor in the brain. The lesions are first registered to the MNI152 1mm space by using FLIRT, and then the centroids of the whole tumor, tumor core and enhancing tumor are extracted as our spatial features. For each subject, we extract 9 spatial features.

Volumetric-spatial Features: The volumetric-spatial features describe the volume of different tumor lesions in different brain regions. First, the Harvard-Oxford subcortical structural atlas brain parcellation regions are registered to the subject space by using FLIRT. The volumes of different types of tumor lesions in each of the parcellation regions left-brain region, middle-brain region, right-brain region, and other brain region are extracted as volumetric-spatial features. For each subject, we extract 78 volumetric-spatial features.

Morphological Features: The morphological features include the length of the major axis of the lesion, the length of the minor axis of the lesion and the surface irregularity of the lesions. We extract 19 morphological features from each subject.

In the first experiment, the ground-truth lesion is used to extract different types of

features, and the pass-type of the binarized connectivity matrix is built to compute the tractographic feature. Recursive feature elimination with cross-validation (RFECV) is used in the feature selection step to shrink the feature. A SVM classifier with the linear kernel is trained with each feature type, and stratified 5-fold cross-validation is conducted 1000 times to achieve a reliable metric. The average and standard deviation of overall survival classification accuracy for different types of features on the BraTS 2018 training dataset is shown in Figure 4.3. This figure demonstrates that the proposed tractographic features have the best overall survival classification performance compared to age, volumetric features, spatial features, volumetric-spatial features, and morphological features. Initial analysis based on feature selection indicates that 12 out of 116 AAL regions are more influential in affecting the overall survival of the brain tumor patient.

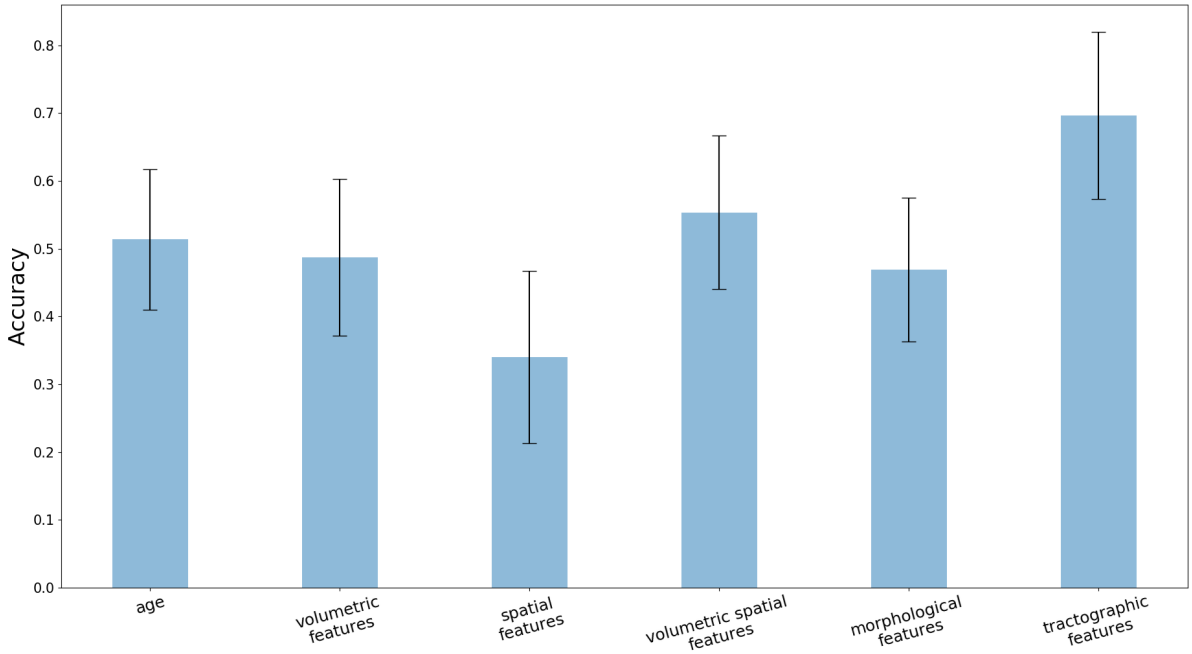


Figure 4.3: Overall survival classification accuracy between different types of features on BraTS 2018 training dataset. 1000-time repeated stratified 5-fold cross-validation is used to obtain the average classification accuracy.

Next, the pass-type of the binarized connectivity matrix is built from the predicted lesion and the tractographic feature is computed from this connectivity matrix. The overall survival classification performance of this tractographic feature is compared with the tractographic feature from our first experiment. In this experiment, we follow the same feature selection method and training strategy, using the same SVM classifier with a linear kernel. The average and standard deviation of overall survival classification accuracy on the BraTS 2018 training dataset is reported in Table 4.1. From this table, the average classification accuracy drops to 63 % when we use predicted lesions instead of ground-truth lesions to generate the tractographic features. This drop is likely caused by the imperfection of our tumor segmentation tool.

Table 4.1: The overall survival classification performance of the proposed tractographic features from the ground-truth lesions and from the predicted lesions on the BraTS 2018 training dataset with 1000-time repeated stratified 5-fold cross-validation.

The source of tractographic features	Classification accuracy (mean±std)
Ground-truth Lesions	0.70 ± 0.12
Predicted Lesions	0.63 ± 0.13

For the training data, the tractographic features are computed using the ground-truth whole tumor, and a linear SVM classifier trained on these features. We used stratified 5-fold cross-validation on the training dataset, averaged over 1000 independent trials. The average OS classification accuracy using the tractographic features was 0.892 on the training set and 0.697 on the cross-validation set. However, when applied to the BraTS 2018 validation and test datasets, the accuracy dropped to 0.357 and 0.416, respectively [1]. Note that for the validation and test data, there is no ground-truth segmentation available. So we first predicted the whole tumor and then the tractography features are extracted from these predicted tumors, followed by the OS classification using the previously trained linear SVM. We speculate that the automated segmentation to predict the whole tumor is one possible reason for the significant variation in performance

between the training and validation/test data, besides, any data specific variations.

4.5 Summary

For overall survival prediction, the novel use of tractographic features appears to be promising for aiding brain tumor patients. To the best of our knowledge, this is the first time to integrate brain parcellation and human brain connectomics for overall survival prediction of brain tumor patients. The final publication of this chapter is available at Springer via `https://doi.org/10.1007/978-3-030-11726-9_12`.

Chapter 5

Utilizing Tractographic Feature to Predict the Clinical Outcome of Stroke Patients

> Every great dream begins with a dreamer. Always remember, you have within you the strength, the patience, and the passion to reach for the stars to change the world.
>
> *Harriet Tubman*

In this chapter, we focus on predicting the clinical outcome of stroke patients given the stroke lesion mask. The volume of stroke lesion is the gold standard for predicting the clinical outcome of stroke patients. However, the presence of stroke lesion may cause neural disruptions to other brain regions, and these potentially damaged regions may also affect the clinical outcome of stroke patients. In this chapter, we also utilize the tractographic feature to capture these potentially damaged regions and predict the

modified Rankin Scale (mRS), which is a widely used outcome measure in stroke clinical trials. The tractographic feature is built from the stroke lesion and average connectome information from a group of normal subjects. The tractographic feature takes into account different functional regions that may be affected by the stroke, thus complementing the commonly used stroke volume features. The proposed tractographic feature is tested on a public stroke benchmark Ischemic Stroke Lesion Segmentation (ISLES) 2017 and achieves higher accuracy than the stroke volume and the state-of-the-art feature on predicting the mRS grades of stroke patients. Also, the tractographic feature yields a lower average absolute error than the commonly used stroke volume feature.

5.1 Introduction

According to the World Health Organization, stroke is the second leading cause of death and the third leading cause of disability worldwide [89]. 15 million people worldwide suffer a stroke each year, and 5.8 million people die from it [90]. Around 87% strokes are ischemic strokes, which happen as a result of an obstruction within a blood vessel in the brain [91]. The corresponding lack of oxygen results in different degrees of disability of people, and the modified Rankin Scale (mRS) shown in Table 5.1 is commonly used to measure the degree of disability or dependence in the daily activities of stroke patients [13]. The mRS has been widely used to measure the clinical outcome for stroke clinical trials [92–94].

Several studies [92,95–101] demonstrate significant correlations between stroke volume and mRS grades, with larger lesions predicting more severe disability. That is, the volume of stroke lesion is a gold predictor on the clinical outcome. There are few recent studies which extract different features other than the volume from stroke lesion to predict the mRS grades of stroke patients. Forkert et al. [102] used the problem-specific brain regions

Table 5.1: The modified Rankin Scale (mRS)

Score	Description
0	No symptoms.
1	No significant disability. Able to carry out all usual activities, despite some symptoms.
2	Slight disability. Able to look after own affairs without assistance, but unable to carry out all previous activities.
3	Moderate disability. Requires some help, but able to walk unassisted.
4	Moderately severe disability. Unable to attend to own bodily needs without assistance, and unable to walk unassisted.
5	Severe disability. Requires constant nursing care and attention, bedridden, incontinent.
6	Deceased.

for lesion overlap quantification and the multi-class support vector machine to predict the mRS grades of stroke patients. Maier and Handels [9] extracted 1650 image features and 12 shape characteristics from the stroke, the volume surrounding the stroke and the remaining brain volume, and applied a random forest regressor with 200 trees on these 1664 features to predict the mRS scores of stroke patients. Choi et al. [103] used a shallow convolutional neural network to extract image features, fitted a logistic regression model with clinical parameters, and integrated the outputs from both the neural network and the regressor to predict the mRS scores of stroke patients. Mahmood and Basit [104] extracted the local image features and stroke characteristics, and used a regression forest to predict the mRS scores of stroke patients. All these recent studies focus on extracting the first-order features of the stroke lesion. However, the presence of stroke lesion may damage other brain regions that may affect the clinical outcome of stroke patients.

In this chapter, we construct the second-order feature from the stroke lesion of a patient and the average connectome information from a group of normal subjects without diffusion MR images of the stroke patient. This so-called tractographic feature describes the potential damage brain regions due to the neural disruptions of stroke lesion. Then,

we use the tractographic features of stroke patients to predict their mRS grades. In this chapter, we show that the tractographic feature is a better predictor on the clinical outcome of stroke patients. Also, we discover that the stroke lesion in the left brain has a higher impact on mRS grades of stroke patients compared to the stroke lesion in the right brain. The details of the tractographic feature we use for stroke patients are described in Section 5.3 and the quantitative results are provided in Section 5.5.

The main contribution of this chapter is we demonstrate that the tractographic feature has potential to be a better predictor on the clinical outcome of stroke patients compared to the volume of stroke, other first-order features, and the state-of-the-art feature.

5.2 Dataset

Ischemic Stroke Lesion Segmentation (ISLES) 2017 [33,34] provides 43 subjects in the training dataset. Each subject has (1) two diffusion maps including diffusion-weighted imaging (DWI), and apparent diffusion coefficient (ADC), (2) five perfusion maps containing cerebral blood volume (CBV), cerebral blood flow (CBF), mean transit time (MTT), time-to-peak (TTP) and, time-to-maximum (TMAX), (3) one ground-truth lesion mask, and (4) clinical parameters. The ground-truth lesion mask was built in the follow-up anatomical sequence (MR-T2w or MR-FLAIR). The clinical parameters include mRS score ranging from 0 to 4, time-to-mRS (88+ days), thrombolysis in cerebral infarction (TICI) scale grade from 0 to 3, time-since-stroke (in minutes), and time-to-treatment (in minutes). TICI scale grade, time-since-stroke and, time-to-treatment were missing for some subjects. Therefore, these three clinical parameters are not considered in our experiments. We only focus on the subjects who obtain an mRS score at 3 months (90 days) following hospital discharge since ascertainment of disability at 3-month post-stroke is an essential component of outcome assessment in stroke patients [105]. Therefore, only

37 subjects from ISLES 2017 are considered.

5.3 Tractogrpahic Feature

Tractographic feature describes the potentially damaged region impacted by the presence of the stroke lesion through the average connectome information [106] from 1021 Human Connectome Project (HCP) [86] subjects. For each HCP subject, q-space diffeomorphic reconstruction (QSDR) [87] is used to compute the diffusion orientation distribution function. Figure 5.1 shows the workflow of building a tractographic feature for a stroke patient.

Given the stroke lesion in the subject space, we first map the stroke lesion to the Montreal Neurological Institute (MNI) space by using FLIRT [51–53] from FSL (low image quality may cause problems on the image registration.) Second, we place one seed within each voxel of the whole brain region, and a deterministic diffusion fiber tracking method [54] is used to find the possible tracts passing through the stroke volume inside the brain from the average connectome information of 1021 HCP subjects. Topology-informed pruning (TIP) [107] is used to remove false-positive tracts. The detailed parameters for fiber tracking are depicted in the following sections. Third, an existing brain parcellation atlas is used to create the connectivity matrix C, which describes the degree of damage between different brain parcellation regions. C is a $N \times N$ connectivity matrix, and N is the number of brain parcellation regions in an existing atlas.

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1N} \\ c_{21} & c_{22} & \dots & c_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{N1} & c_{N2} & \dots & c_{NN} \end{bmatrix} \tag{5.1}$$

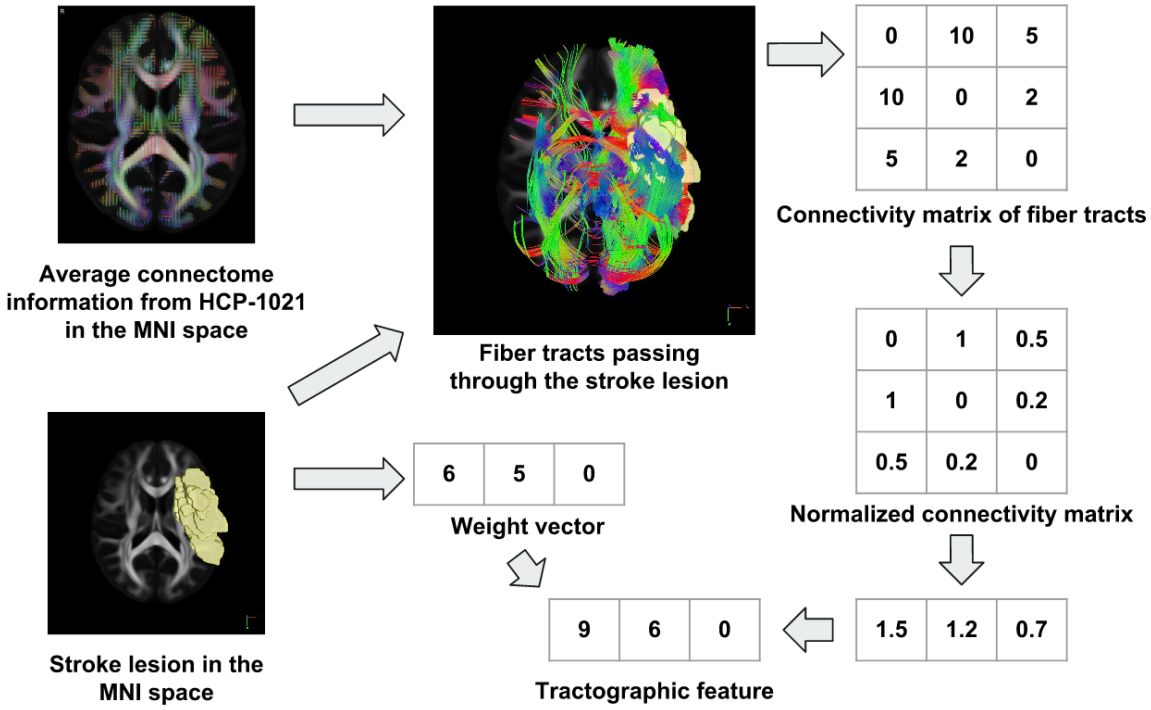


Figure 5.1: The workflow for constructing a tractographic feature from a stroke region. First, the stroke lesion is registered to the average diffusion orientation distribution function in the MNI space, and we place one seed in each voxel inside the whole brain region to find all possible tracts passing through the stroke volume. Second, an existing brain parcellation atlas is used to form a connectivity matrix. Here, we show a simple example of dividing the brain into three sub-regions. Therefore, the size of the connectivity matrix of fiber tracts is 3×3. Each value in this connectivity matrix stands for the number of fiber tracts starting from a brain parcellation region and ending in another brain parcellation region. Then, we normalize this connectivity matrix with respect to its maxima (10 in our example.) Afterward, we sum up each column of this normalized connectivity matrix to form a row vector. We then use the distribution of stroke lesion in different brain parcellation regions to build a weight vector whose size is 1×3. Finally, this row vector is multiplied by the weight vector element-wisely to form the tractographic feature.

c_{ij} notes the number of tracts starting from a region i and ending in a region j. Note that this connectivity matrix is symmetric. Then, we normalize this connectivity matrix by dividing every element on this matrix by its maximum value.

$$\hat{C} = C \odot \frac{1}{c_m} \tag{5.2}$$

$\hat{C}$ is the normalized connectivity matrix, $\odot$ is the element-wise multiplication operator, and c_m is the maximum element of the connectivity C. Afterward, we sum up each column in this normalized connectivity matrix $\hat{C}$ to form a row vector $\vec{L}$.

$$\vec{L} = \sum_{i=1}^{N} \hat{c}_{ij} = \Big[l_1, l_2, \ldots, l_N \Big] \tag{5.3}$$

From the stroke lesion, we build a weight vector $\vec{\gamma}$ which is the volume of the stroke lesion in each brain parcellation region.

$$\vec{\gamma} = \Big[s_1, s_2, \ldots, s_N \Big] \tag{5.4}$$

s_i is the volume of the stroke lesion in the i-th brain parcellation region. In the end, the row vector $\vec{L}$ from Eq. (5.3) is multiplied by this weight vector $\vec{\gamma}$ element-wisely to form the tractographic feature $\vec{T}$.

$$\vec{T} = \vec{\gamma} \odot \vec{L} \tag{5.5}$$

$\odot$ is the element-wise multiplication operator. This vector $\vec{T}$ is the tractographic feature extracted from stroke lesion without any diffusion information of a patient. The Automated Anatomical Labeling (AAL) [5] template is used to define 116 brain regions so the dimension of the tractographic feature is 116. The reason for choosing AAL rather than other existing atlases because this atlas contains an optimal number ($\sim$ 100) of brain

regions that could make each region large enough to compensate possible stroke-induced lesion effect or distortion, and this atlas contains cortical, subcortical and cerebellar regions, which could be equally important for mRS prediction [108].

5.3.1 Parameters of Fiber Tracking

DSI Studio [38] is used to build the fiber tracts for each subject. Table 5.2 shows the tracking parameters[1] we used in our experiments. The type of stroke lesion is set to ROI (–roi=stroke_lesion) that found all possible tracts passing through the stroke lesion.

Table 5.2: Tracking parameters of building the fiber tracts for stroke patients in this chapter. More details of parameters can be found at `http://dsi-studio.labsolver.org/Manual/Fiber-Tracking`.

Parameter	Value
Termination Index	qa
Threshold	0.15958
Angular Threshold	90
Step Size (mm)	0.50
Smoothing	0.50
Min Length (mm)	3.0
Max Length (mm)	500.0
Topology-Informed Pruning (iteration)	1
Seed Orientation	All orientations
Seed Position	Voxel
Randomize Seeding	Off
Check Ending	Off
Direction Interpolation	Tri-linear
Tracking Algorithm	Streamline(Euler)
Terminate if	2,235,858 Tracts
Default Otus	0.60

[1] parameter_id=F168233E9A99193F32318D24ba3Fba3Fb404b0FA43D21D22cb01ba02a01d

5.3.2 Parameters of Connectivity Matrix

DSI studio is used to create the connectivity matrix (`http://dsi-studio.labsolver.org/Manual/command-line-for-dsi-studio`) followed by fiber tracking. Automated Anatomical Labeling [5] is chosen to form a 116×116 connectivity matrix. The type of the connectivity matrix is set to end, the value of each element in the connectivity matrix is the count of fiber tracts, and the threshold to remove the noise in the connectivity matrix is set to 0.

5.4 Experiments

5.4.1 Experiment 1: tractographic feature vs. other first-order features

In our first experiment, we compare the mRS prediction performance of the tractographic feature with other first-order features extracted from the lesion mask. These features include the volumetric feature, spatial feature, morphological feature and volumetric-spatial feature described in Table 5.3. The pipeline of our first experiment is depicted in Figure 5.2.

We first register the stroke lesion from the subject space to the MNI space to overcome the differences of the voxel spacing and image dimension between different subjects, and features are extracted from this normalized stroke lesion. From this normalized stroke lesion, we extract tractographic feature and other features shown in Table 5.3. After feature extraction, we apply a standard feature normalization on the extracted features to ensure that each dimension of the features has the same dynamic range. Then, we apply two feature selection methods on these different types of features. The first feature selection is to remove the dimensions of the feature with the low variance

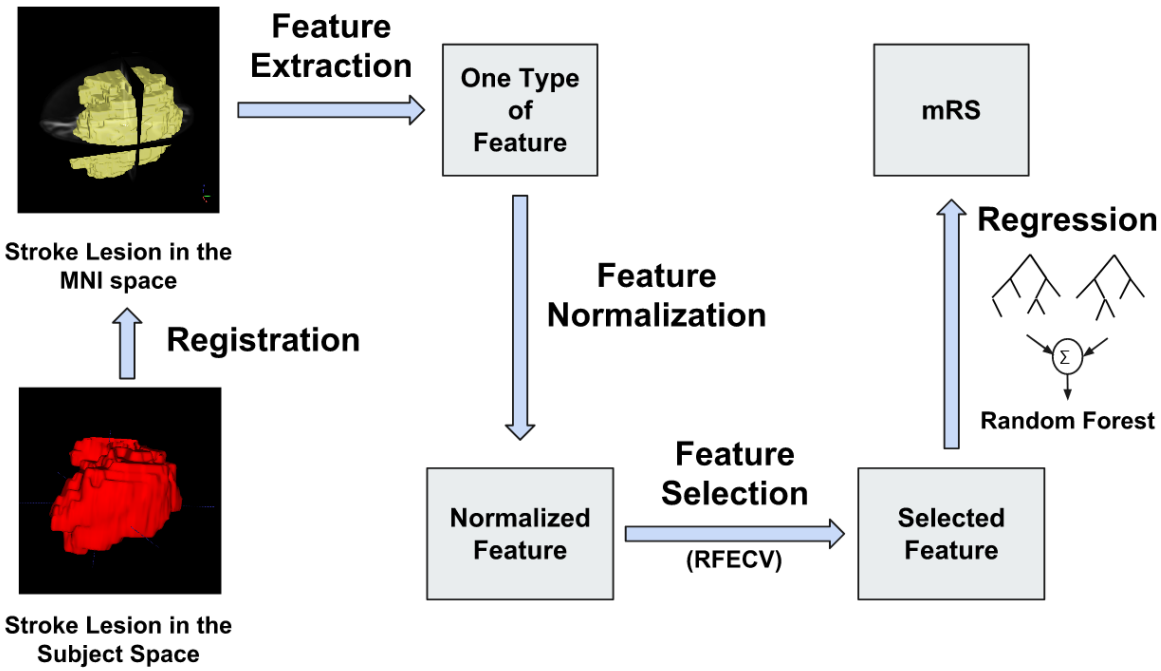


Figure 5.2: The pipeline of the first experiment for predicting the clinical outcome of stroke patients. The stroke lesion is first mapped to the MNI space, and five different types of features are extracted from this normalized stroke lesion. Then, we apply the standard feature normalization on this extracted feature to ensure each dimension of the feature has the same scale. After feature normalization, we remove the dimensions of the feature with a lower variance between subjects and apply recursive feature elimination with cross-validation (RFECV) on the feature selection. In the end, we train one random forest regressor with each type of features and use these random forest regressors to predict the mRS grades with each type of features for the stroke patients.

between subjects. The second step is to apply the recursive feature elimination with leave-one-out cross-validation and a random forest regressor that search for the best subset of feature which yields the lowest average absolute error. After feature selection, we train one random forest regressor for each type of features. That is, five random forest regressors are trained with five different types of features. Each random forest regressor has 300 trees, and the maximum depth of each tree is 3 with the same initial seed. In the end, we use different types of features with corresponding random forest regressors to predict the mRS grades of stroke patients.

Table 5.3: First-order features extracted from the stroke lesion.

Type of feature	**Descriptions**
Volumetric feature	Volumetric feature is the volume of the lesion (in mm^3) in the MNI space, and it only has one dimension.
Spatial feature	Spatial feature describes the location of the lesion in the brain. The centroid of the lesion is extracted as the spatial feature for each subject, and the spatial feature has three dimensions.
Morphological feature	Morphological feature describes shape information of the lesion. The length of the major axis and minor axis of the lesion, the ratio of the length of the major axis and minor axis of the lesion, the solidity and roundness of the lesion, and the surface of the lesion are extracted as the morphological feature. The morphological feature has six dimensions for each subject.
Volumetric-spatial feature	Volumetric-spatial feature describes the distribution of lesion in different brain parcellation regions from an existing structural atlas. Automated Anatomical Labeling (AAL) [5] is used to build the volumetric-spatial feature so the dimension of the volumetric-spatial feature is 116.

5.4.2 Experiment 2: tractographic feature vs. state-of-the-art feature

We first extract 1650 image features and 14 shape features from the lesion volume and the ADC map and concatenate these two different types of features together to build a feature with 1664 dimensions. Then, we apply a standard feature normalization to this feature to ensure each dimension of the feature has the same scale. Afterward, we remove the dimensions of the feature with the low variance between subject and apply a recursive feature elimination with leave-one-out cross-validation to search for the best subset of feature which gives us the best mRS prediction performance. In the end, we apply a random forest regressor with 300 trees whose maximum depth is 3 on this selected feature to predict the clinical outcome of stroke patients. We use the same pipeline as our first experiment to extract the tractographic feature from the stroke lesion and predict the mRS grades of stroke patients.

5.4.3 Experiment 3: performance of feature selection

In the third experiment, we would like to examine the performance of feature selection on predicting the mRS grades of stroke patients. From Figure 5.2 and Figure 5.3, feature selection is applied on features extracted from the stroke lesion before predicting the mRS grades of stroke patients. The feature selection includes removing the dimensions of the feature with the low variance between subjects and recursive feature elimination with leave-one-out cross-validation. In this experiment, we directly apply the random forest regressor with 300 trees whose maximum depth is 3 on the normalized features without feature selection to predict the mRS of stroke patients.

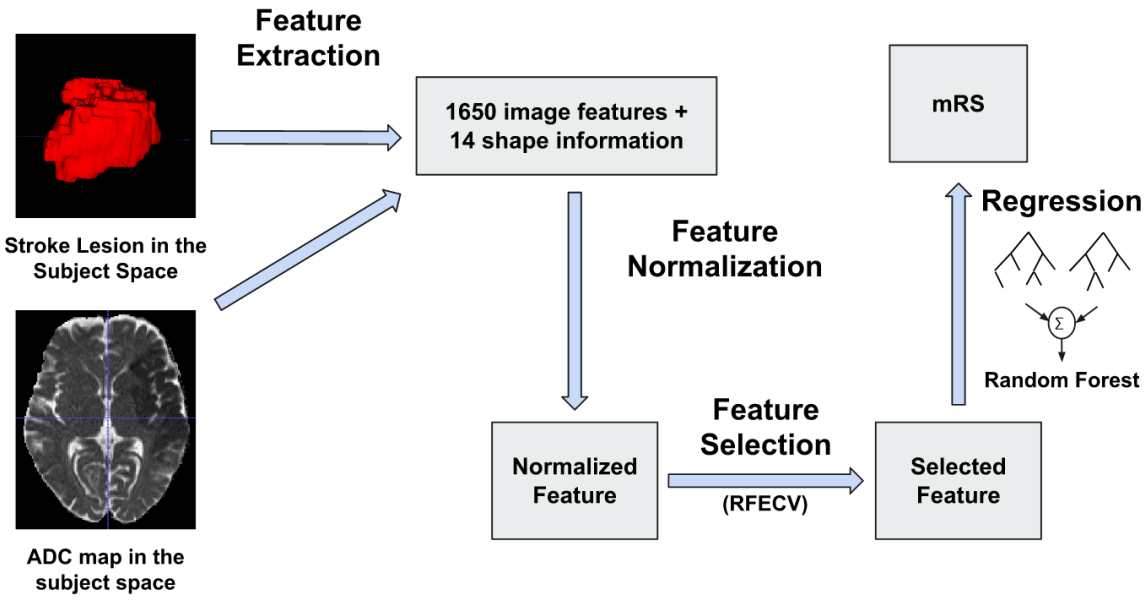


Figure 5.3: Experiment 2 pipeline. We adjust the feature extraction method proposed by Maier and Handels [9] to work on ISLES 2017. First, we extract 1650 image features and 14 shape information from the stroke lesion and the ADC map. These two different types of features are concatenated together to form a feature with 1664 dimension. Then, we normalize this extracted feature to ensure each dimension of this feature has the same scale. After feature normalization, we remove the dimensions of the feature with the low variance between subjects and apply recursive feature elimination with leave-one-out cross-validation to find the best subset of feature which has the smallest average absolute error. Then, we use a random forest regressor with 300 trees whose maximum depth is 3 on the selected feature to predict the mRS grades of stroke patients.

5.4.4 Evaluation Metrics

The employed evaluation metrics are (i) the accuracy, which is the percentage of the predicted labels matching the corresponding ground-truth labels, and (ii) the average absolute error between the predicted labels and the corresponding ground-truth labels.

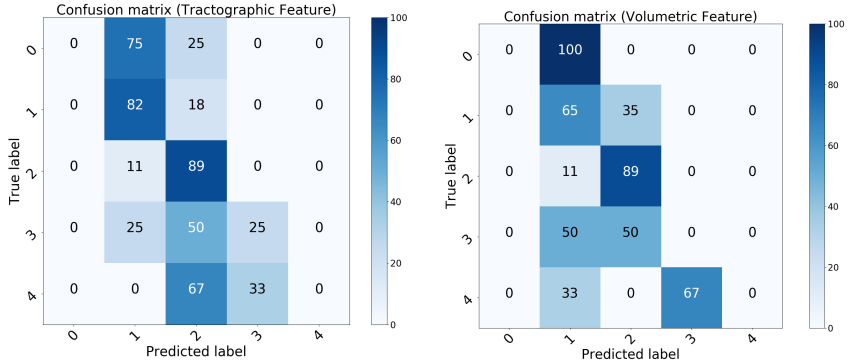
5.5 Experimental Results

5.5.1 Experiment 1: tractographic feature vs. other first-order features

We evaluate the mRS prediction performance of different types of features with leave-one-out cross-validation on the ISLES 2017 training dataset. The quantitative results are reported in Table 5.4 and the confusion matrices are shown in Figure 5.4. From Table 5.4, the tractographic feature has the highest accuracy on predicting the mRS scores of stroke patients compared to the volumetric features, volumetric-spatial feature, morphological features, and spatial features. Moreover, the tractographic feature yields a lower average absolute error than volumetric feature, volumetric-spatial feature ($p < 0.1$), morphological feature ($p < 0.05$) and spatial feature ($p < 0.01$). Also, Figure 5.4 shows that the tractographic feature has a better performance in predicting mRS score 2 and 3 compared to other first-order features.

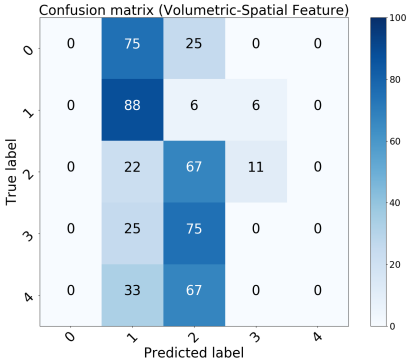
5.5.2 Experiment 2: tractographic feature vs. state-of-the-art feature

The quantitative results of the state-of-the-art feature are shown in the last row of Table 5.4 and the Figure 5.4f. From Table 5.4, the tractographic feature also achieves

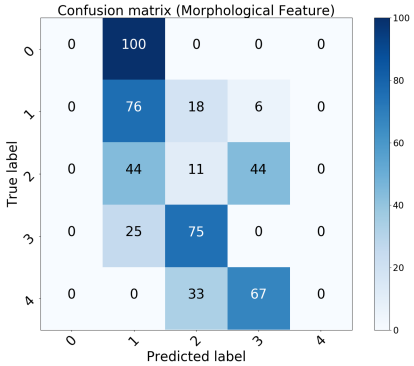


(a) Tractographic Feature

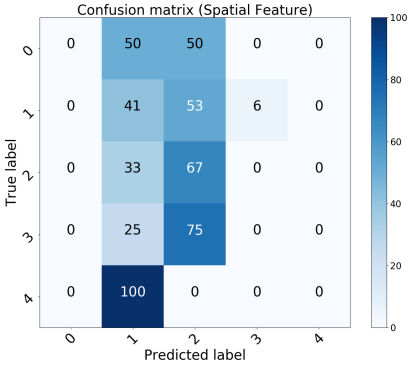
(b) Volumetric Feature



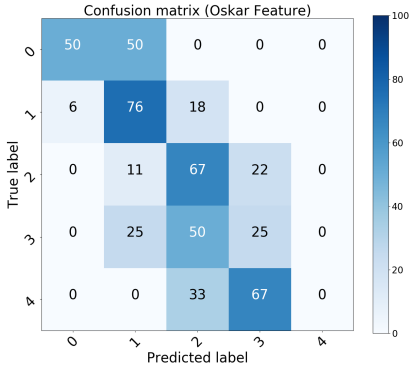
(c) Volumetric-Spatial Feature



(d) Morphological Feature



(e) Spatial Feature



(f) Maier and Handels [9]

Figure 5.4: The confusion matrices of the predicted mRS grades and the corresponding ground-truth mRS grades using different types of features. These confusion matrices are normalized by the number of subjects in each mRS grades. The diagonal elements represent the percentages for which the predicted mRS grade is equal to the ground-truth mRS grade. The higher (darker) the diagonal values of the confusion matrix the better, indicating many correct predictions. Feature selection steps are applied on these features. A random forest regressor is trained to predict the mRS grades for stroke patients with leave-one-out cross-validation.

Table 5.4: The mRS prediction performance of different types of features on the ISLES 2017 training dataset with leave-one-out cross-validation. Feature selection steps are applied on these features. A random forest regressor is used to predict the mRS score for each stroke patient. The random forest regressor has 300 trees, and the maximum depth of each tree is 3. The bold numbers show the best performance. (The average absolute error is reported as mean ± std.)

Type of feature	**Accuracy**	**Average absolute error**
Tractographic feature	**0.622**	0.487 ± 0.683
Volumetric feature	0.514	0.595 ± 0.715
Volumetric-spatial feature	0.568	0.621 ± 0.817
Morphological feature	0.378	0.703 ± 0.609
Spatial feature	0.351	0.919 ± 0.882
Maier and Handels [9]	0.595	$\mathbf{0.460 \pm 0.597}$

higher accuracy and similar average absolute error ($p = 0.81$) compared to the state-of-the-art feature. From Figure 5.4, the tractographic feature has a better performance on predicting mRS grade 1 and 2 compared to the state-of-the-art feature.

5.5.3 Experiment 3: performance of feature selection

The mRS prediction performance of different types of features without feature selection is reported in Table 5.5. Compared with Table 5.4 and Table 5.5, it is shown that the feature selection steps improve the performance of tractographic feature, volumetric-spatial feature, morphological feature and the state-of-the-art feature, particularly for the tractographic feature ($p < 0.5$) and state-of-the-art feature ($p < 0.01$). Also, it is noted that without feature selection, the tractographic feature has the best accuracy compared to other features and similar average absolute error ($p = 0.76$) compared to the volumetric feature (see Table 5.5).

Table 5.5: The mRS prediction performance of different types of features without feature selection steps on the ISLES 2017 training dataset with leave-one-out cross–validation. A random forest regressor is used to predict the mRS score of each stroke patient. The random forest regressor has 300 trees, and the maximum depth of each tree is 3. The bold numbers show the best performance. (The average absolute error is reported as mean ± std.)

Description	Accuracy	Average absolute error
Tractographic Feature	**0.541**	0.622 ± 0.748
Volumetric feature	0.514	**0.595 ± 0.715**
Volumetric-Spatial Feature	0.514	0.649 ± 0.779
Morphological Feature	0.351	0.730 ± 0.600
Spatial feature	0.351	0.919 ± 0.882
Maier and Handels [9]	0.405	0.730 ± 0.684

5.6 Discussion

From the first experiment, the tractographic feature has the best mRS prediction accuracy and the lowest average absolute error compared to other first-order features(See Table 5.4). The main reason is that the tractographic feature integrates volumetric-spatial information of the stroke lesion and average diffusion information from a group of normal subjects. Average diffusion information gives the tractographic feature potentially damaged regions impacted by the stroke lesion. Potentially damaged brain regions can be discovered by finding the possible tracts passing through the lesion volume. These potentially damaged regions are formatted in the connectivity matrix from Eq. (5.1). The weight $\vec{\gamma}$ from Eq. (5.4) brings spatial and volumetric information of the lesion to the tractographic feature. Also, it is worth noting that the volumetric-spatial feature is similar to the tractographic feature without average connectome information, and the mRS prediction performance of volumetric-spatial feature has a significant improvement ($p < 0.1$) by considering average connectome information from a group of normal subjects.

From the second experiment, the tractographic feature also has better accuracy than

the state-of-the-art features proposed by Maier and Handels [9] on predicting the mRS of stroke patients. Their feature contains 99% image features from the ADC maps and only 1% shape information of the stroke lesion. That is, their feature does not work well without the ADC maps. Moreover, the inconsistent voxel values between different MR images might cause some problems to their image features. However, the tractographic feature is more robust compared to image features since the tractographic feature is generated from a lesion mask and it does not require intensity information from MR images.

In the first two experiments, we apply the recursive feature selection with cross-validation on different types of features and this procedure reduces one dimension of feature recursively until finding the best subset of the feature. This feature selection step significantly improves the performance of the tractographic feature ($p < 0.05$) whose dimension shrinks to 8 from 116. This selected tractographic feature comes from eight AAL regions (left and right inferior temporal gyrus, right Rolandic operculum, left middle frontal gyrus, orbital part and triangular part of right inferior frontal gyrus, left angular gyrus and left putamen) shown in Figure 5.5. After feature selection, we use a random forest regressor to predict the mRS grades of stroke patients with one type of feature. The random forest regressor can give importance to each dimension within a type of feature. The importance of each AAL region within the selected tractographic feature is shown in Figure 5.6. From this figure, left inferior temporal gyrus yields higher average importance compared to the other seven regions within 37 ISLES 2017 training subjects on the task of predicting the mRS grades. Left inferior temporal gyrus is connected to areas that are important for language processing and speech production. That could explain the reason why this region has the greatest effect on the mRS of stroke patients [109].

In the third experiment, we use different types of features with a random forest regressor to predict the mRS grades of stroke patients without feature selection steps. As

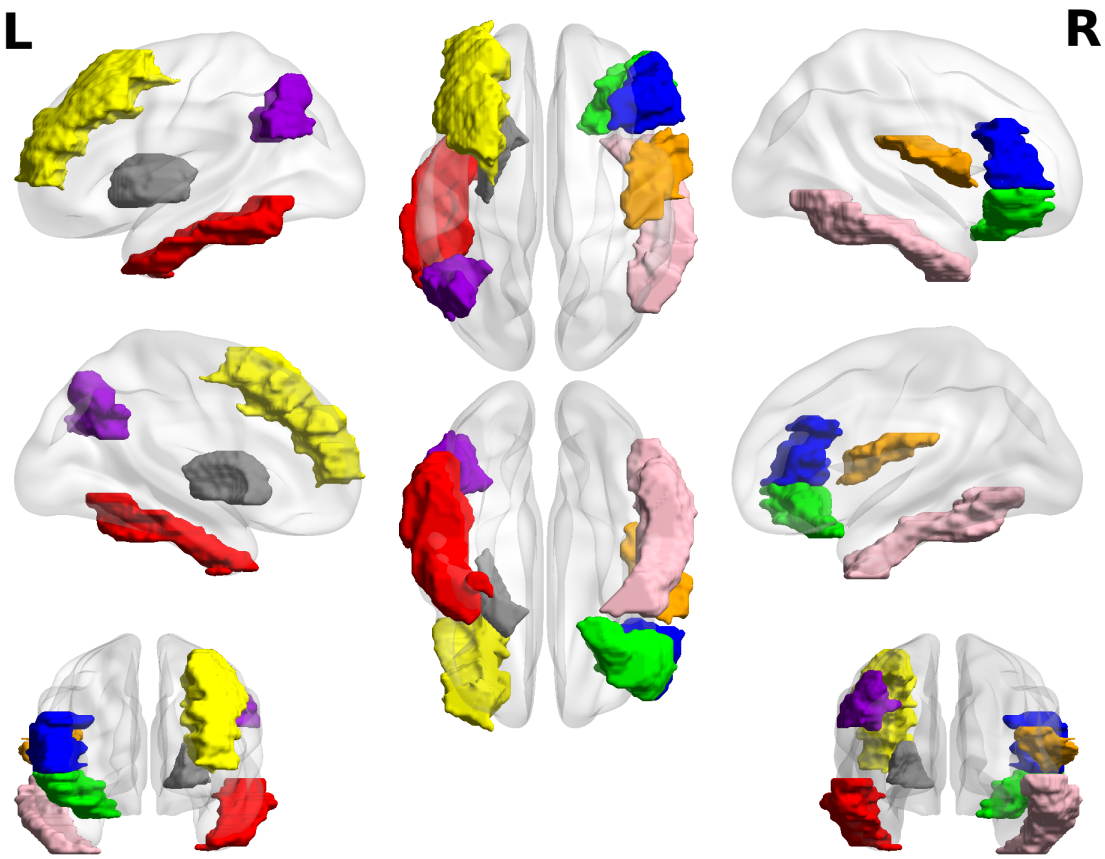


Figure 5.5: Selected tractographic feature from eight AAL regions for predicting the clincial outcome of stroke patients. These regions include left (in red) and right (in pink) inferior temporal gyrus red, right Rolandic operculum (in orange), left middle frontal gyrus (in yellow), orbital part (in green) and triangular part (in blue) of right inferior frontal gyrus, left angular gyrus (in purple) and left putamen (in grey) after applying the recursive feature selection with cross-validation on the original tractographic features. These tractographic features are extracted from 37 ISLES 2017 training subjects. The figure is drawn by using BrainNet Viwer [10].

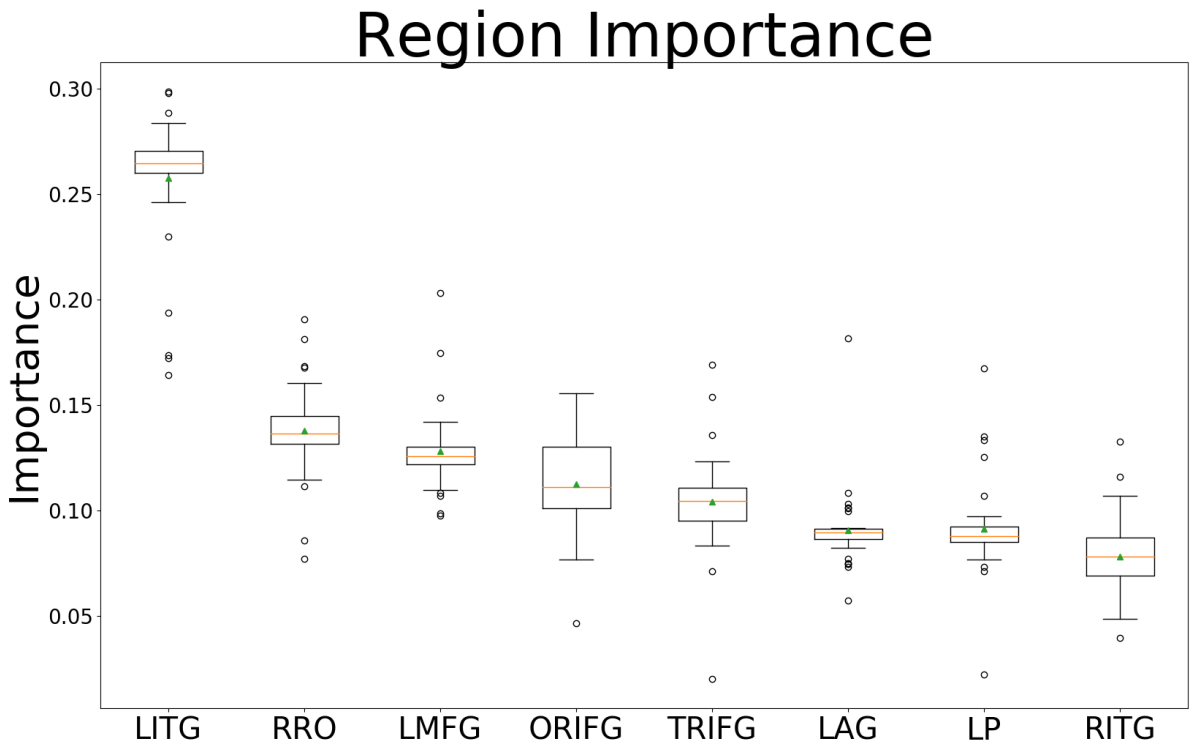


Figure 5.6: Region importance of eight selected AAL brain parcellation regions for predicting the clinical outcome of stroke patients. These regions are given by a random forest regressor with 300 trees whose maximum depth is 3. The average values are marked in the green triangles. Left inferior temporal gyrus (LITG) yields a higher mean importance (0.26) than right Rolandic operculum (RRO, 0.14), left middle frontal gyrus (LMFG, 0.13), orbital part (ORIFG, 0.11) and triangular part (TRIFG, 0.10) of right inferior frontal gyrus, left angular gyrus (LAG, 0.09), left putamen (LP, 0.09) and right inferior temporal gyrus (RITG, 0.08) within 37 ISLES 2017 training subjects on the task of predicting the mRS grades of stroke patients.

mentioned above, a random forest regressor can give the importance to each dimension within a type of feature. That is, the importance to each 116 AAL regions from the original tractographic feature is given by the random forest regressor for the task of clinical outcome prediction. The importance of all AAL regions are shown in Table 5.6. From this table, left inferior temporal gyrus again is the most important areas affecting the mRS of stroke patients compared to other regions. Figure 5.7 shows the fiber tracts passing through the left inferior temporal gyrus, and a large number of fibers are going across the splenium of the corpus callosum that may explain the reason why the left inferior temporal gyrus is the most crucial regions in predicting the mRS of stroke patients [110]. In addition, it should be noted that the seven most important regions (left inferior temporal gyrus, left middle frontal gyrus, right Rolandic operculum, left putamen, triangular part and orbital part of right inferior frontal gyrus and left angular gyrus) are selected after applying two feature selection steps on the original tractographic feature (see Figure 5.5), and the total importance of these regions is 0.58. Left inferior temporal gyrus, left middle frontal gyrus and left angular gyrus are on the left and eloquent areas that mean they should likely have a profound effect on the patient's neurological status [111]. Left putamen is one of the structures that compose the basal ganglia. Stroke damage in the left putamen is strongly associated with motor disorders and cognitive disorders [112]. Moreover, from Table 5.6, the importance of the left hemispheres is 0.55, and the right hemispheres is 0.37. Namely, the lesion in the left hemispheres has a higher impact on the clinical outcome of stroke patients compared to the lesion in the right hemispheres [113]. Again, the stroke lesion in the left brain, particularly in inferior temporal gyrus, has a higher impact on predicting the mRS grades of stroke patients.

Table 5.6: Region importance of AAL brain parcellation regions averaged from 37 ISLES 2017 subjects. Regions whose importance is less than 0.01 are not listed in this table. The importance of left hemispheres is 0.55, and right hemispheres are 0.37. The importance of regions is given by a random forest regressor trained with the original tractographic features of 37 ISLES 2017 subjects with leave-one-out cross-validation. The values are reported as mean.

Region	**Importance**
Left inferior temporal gyrus	0.19
Left middle frontal gyrus	0.09
Right Rolandic operculum	0.08
Left putamen	0.07
Triangular part of right inferior frontal gyrus, Orbital part of right inferior frontal gyrus, Left angular gyrus	0.05
Right inferior temporal gyrus	0.04
Opercular part of right inferior frontal gyrus	0.03
Orbital part of right superior frontal gyrus, Right insula, Left hippocampus, Left middle occipital gyrus, Left precuneus, Right globus pallidus	0.02
Left precentral gyrus, Right precentral gyrus, Right superior frontal gyrus, Right middle frontal gyrus, Opercular part of left inferior frontal gyrus, Triangular part of left inferior frontal gyrus, Right hippocampus, Left amygdala, Right middle occipital gyrus, Left postcentral gyrus, Left inferior parietal lobule, Left globus pallidus, Left thalamus, Left middle temporal gyrus, Right middle temporal pole	0.01

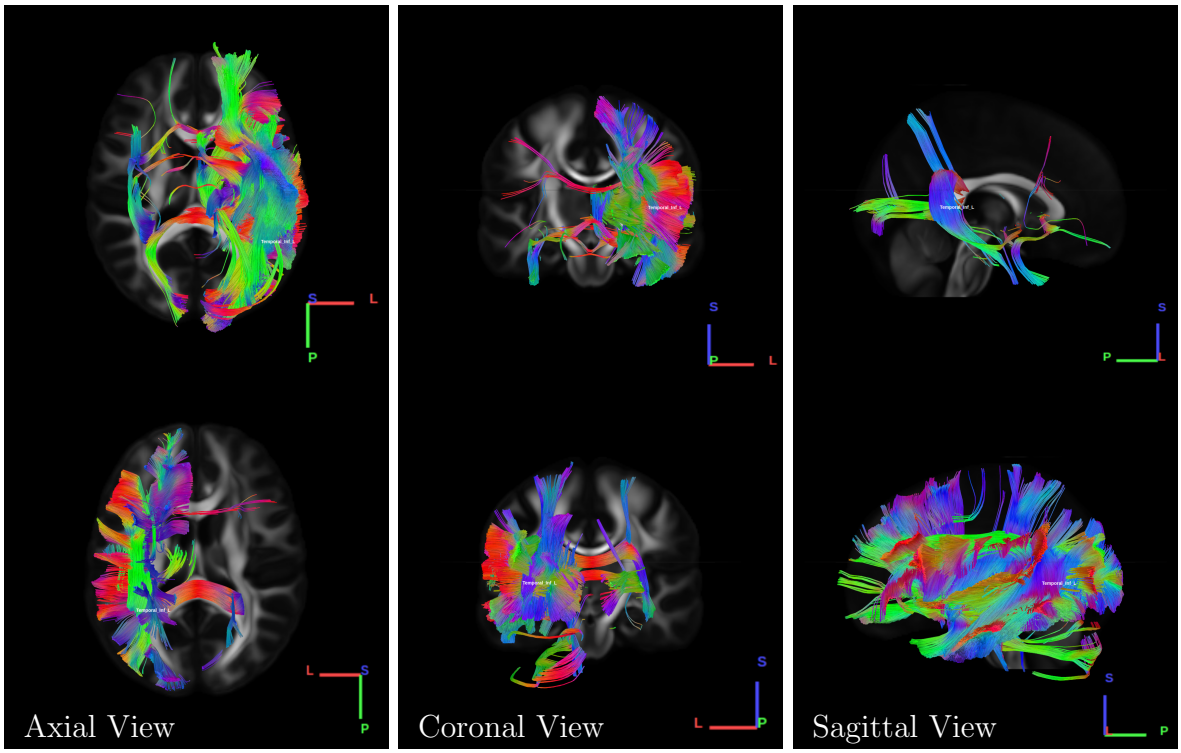


Figure 5.7: The fiber tracts passing through the left inferior temporal gyrus from the average connectome information of 1024 HCP subjects. We place a seed in each voxel inside the whole brain to find all possible tracts passing through the left inferior temporal gyrus.

5.7 Summary

Tractographic feature leads to promising mRS prediction results in ISLES 2017 dataset. However, this dataset contains only 37 stroke subjects and further validation on a larger and diverse data is needed. Besides, it has been shown that fusion of multiple features can boost the performance of large-scale classification and recognition [114–117], and it is worth a further investigation with a larger and more comprehensive dataset. Also, the tractographic feature can be extracted from any CT or MR scan with a brain lesion, e.g., traumatic brain injury, brain tumor. The region importance table (Table 5.6) would be a helpful reference for researchers. The final publication of this chapter will be available in the coming proceedings of 5th International MICCAI Brainlesion Workshop, BrainLes 2019 by Springer.

Chapter 6

Unsupervised 3D Feature Learning for Predicting with Mild Traumatic Brain Injury

> Your time is limited, so don't waste it living someone else's life. Don't be trapped by dogma - which is living with the results of other people's thinking.
>
> *Steve Jobs*

In this chapter, we focus on predicting the outcome of mild traumatic brain injury patients. It is noted that it is hard to observe the hematoma on the structural MR images of patients with mild traumatic brain injury. Therefore, our proposed tractographic feature does not work in this condition. Alternatively, we present an unsupervised three-dimensional feature clustering algorithm to gather the Mild Traumatic Brain Injured Outcome Prediction (mTOP) 2016 challenge data into 3 groups. We use the brain MR-

T1, diffusion tensor fractional anisotropy, and diffusion tensor mean diffusivity images provided by the mTOP 2016 competition. A distance-based size constraint method for data clustering is used. The proposed approach achieves 0.267 adjusted rand index and 0.3556 homogeneity score within the 15 labeled test subjects, corresponding to 10 correctly classified data items. Our proposed method was ranked third in the mTOP 2016 challenge (`https://tbichallenge.wordpress.com/`) held with the international conference on medical image computing and computer-assisted intervention (MICCAI) 2016.

6.1 Introduction

Traumatic brain injury (TBI) is an important public health concern in the United States and worldwide. TBI is defined as an acute brain injury resulting from mechanical energy to the head from external physical forces. It is estimated that approximately 1.5-2 million Americans suffer from TBI annually. TBI also contributes to approximately 30% of all deaths in the USA annually. TBI often results in residual symptoms that affect an individual's cognition, movement, sensation, and/or emotional functioning. Recovery and rehabilitation from TBI may require considerable resources and may take years. Some individuals never fully recover, and some require lifetime ongoing care and support. TBI has an enormous social and financial cost, with estimates of the annual financial burden associated with TBI ranging between 9 and 10 billion US dollars [118, 119].

Predicting the outcome of patients suffering from TBI could be both facilitate clinical decision making and support the development of new therapeutic concepts. However, prognostics are often based on clinical or cognitive symptoms, which might be biased by personal perception. Therefore, there is an urge for more objective methods that detect radiological evidence after TBI. However, the strong heterogeneity of the injury

pattern and the complex change of pathology over time pose a persistent challenge. This especially holds true for mild traumatic brain injury (mTBI), where lesions are non-prevalent and conventional MRI often appears normal, but the injury can cause post-concussional symptoms and neuro-cognitive dysfunction. Despite its valuable potential for outcome prediction, MRI-based features which describe mTBI pathology are not fully understood. Diffusion MRI could be the key to reveal biological factors of strong predictive power. Indeed, common techniques such as local region or whole-brain analysis are highly dependent on adequate image registration and brain parcellation and can be restricted by the high dimensionality of the data.

This chapter addresses the challenge of feature detection and classification of subject data based on brain imaging for mTBI. The imaging data include the MR-T1 and diffusion-weighted images (DWI). Bellotti et al. [120] combined the graph-based and k-nearest neighbors (K-NN) methods to predict the outcome of mTBI patients. Cai and Ji [121] developed an injury prediction or classification pipeline based on diffusion tensor imaging (DTI) by combining a novel deep learning approach with statistical permutation tests. While there is extensive work on applying unsupervised learning to clustering 2D image features [122–125], the problems posed by the mTBI dataset are particularly challenging since the features of interest are likely very localized. Furthermore, the subject categorization is derived not necessarily from the image data but from other observations, making this problem very distinct from the traditional works in natural image processing.

Our proposed workflow includes four stages. The first stage performs data preparation and pre-processing on mild traumatic brain injury outcome prediction (mTOP) 2016 dataset. The second stage performs learning 3D features from brain MR-T1, diffusion tensor fractional anisotropy (DT-FA) and diffusion tensor mean diffusivity (DT-MD) images from 27 subjects of mTOP 2016 dataset. The third stage performs feature representation for each subject, and the last stage performs group clustering based on these

feature representations. The main contributions of this chapter are two-fold. First, we propose a fully unsupervised methodology to learn the 3D features from the volumetric data. Second, we propose a novel distance-based size constraint methodology for data clustering.

6.2 Dataset

The mTOP 2016 dataset consists of MR-T1, DT-FA and DT-MD images, see Fig 6.1. The training set contains 27 subjects belonging to 3 different categories (healthy, patient category 1 or patient category 2) each consisting of 9 subjects. The testing set contains 15 subjects belonging to 3 different categories (healthy, patient category 1 or patient category 2) each consisting of 5 subjects. The mTBI patients are categorized into one of two groups based on their long term recovery status following the injury. The imaging data includes for MR-T1 image at $182 \times 218 \times 182$ voxels, with $1mm \times 1mm \times 1mm$ voxel resolution, and the dimension for DT-FA and DT-MD image is $91 \times 109 \times 91$ with $2mm \times 2mm \times 2mm$ voxel resolution.

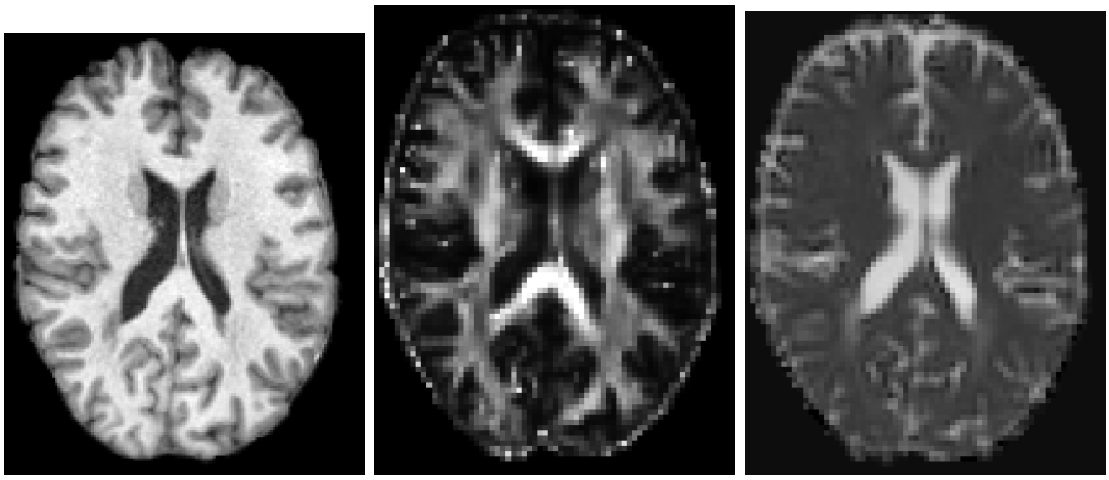


Figure 6.1: An example of the mTOP 2016 dataset. Left: MR-T1 image, Mid: DT-FA image, Right: DT-MD image.

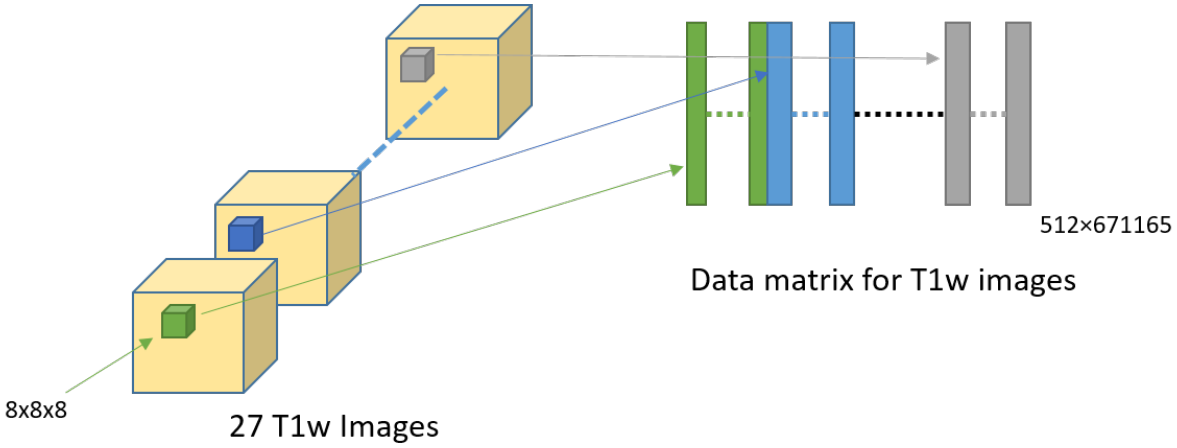


Figure 6.2: Data preparation for MR-T1 images.

6.3 Data Preparation and Pre-processing

Data preparation for MR-T1 images is shown in Fig 6.2. For MR-T1 images, we consider $8 \times 8 \times 8$ voxel volume represented as a 512 dimensional vector of voxel values, $\tilde{x}_{T1}^{(i)} \in \mathbb{R}^{512}$, where i indexes the 3D patch. The overlap between the volumes in a sliding window is 50%, and those volumes that have more than 75% zero values are discarded. Thus, a large number of data vectors are generated that are then organized as column vectors in a matrix. Moreover, these vectors are normalized to zero mean and unit standard deviation:

$$x^{(i)} = \frac{\tilde{x}^{(i)} - mean(\tilde{x}^{(i)})}{std(\tilde{x}^{(i)})}$$

where $\tilde{x}^{(i)}$ is a unnormalized column vector and "mean" and "std" are the mean and standard deviation of the element of $\tilde{x}^{(i)}$. Let X_{T1} represent this matrix that includes data from all of the 27 subjects. Similarly, two other matrices X_{FA} and X_{MD} are constructed. However, since the spatial resolution of the data for these two cases is different from the MR-T1, we use a $4 \times 4 \times 4$ voxel volume. Therefore, the data vectors all represent a

$512\,mm^3$ spatial volume.

After normalization, we apply the standard *Zero Component Analysis* (ZCA) whitening transform [126] on each of the datasets X_{T1}, X_{FA}, and X_{MD}. This helps minimize the correlation among the components of the column vectors. For contrast-normalized data, we set the whitening parameter ϵ_{zca} to 0.01 for $8 \times 8 \times 8$ voxel patches and 0.1 for $4 \times 4 \times 4$ voxel patches.

6.4 3D Dictionary learning via K-means clustering

The next step is to learn a dictionary for each of the data matrices using the standard K-means clustering [127]. A separate dictionary is learned for each of the three matrices. Let the data matrix be $X \in \mathbb{R}^{N \times M}$ and the corresponding dictionary be $D \in \mathbb{R}^{N \times K}$. Then,

Loop until convergence:

$$c_j^{(i)} = \begin{cases} D^{(j)\top} x^{(i)}, & \text{if } j = \arg\min_l |D^{(l)\top} x^{(i)}| \quad \forall i, j. \\ 0, & \text{otherwise.} \end{cases}$$

$$D := XC^\top + D$$

$$D^{(j)} / ||D^{(j)}||_2 \quad \forall j$$

where $c_j^{(i)}$ is the code vector associated with the input $x^{(i)}$ (i^{th} column of X), and $D^{(j)}$ is the j^{th} column of the dictionary D that is a 3D feature we learned. In the end, we will learn K 3D features from a dataset($D \in \mathbb{R}^{N \times K}$). Note that $C \in \mathbb{R}^{K \times M}$. Let the three corresponding dictionaries be D_{T1}, D_{FA}, and D_{MD}.

6.5 Feature Representation using 3D Convolutional Network

Feature computation workflow schematic is shown in Figure 6.3. Input data includes the three types: brain MR-T1, DT-FA, and DT-MD, for each of the subjects. Each of these datasets is first normalized by subtracting the mean voxel value and dividing by the standard deviation within the brain region. The dictionary code words learned from the K-means clustering above are used as the weights for the first convolutional layer. The stride for MR-T1 is 2 voxels, and for DT-FA and DT-MD is 1 voxel. This is followed by a 3D max-pooling layer of size $3 \times 3 \times 3$. The final merge layer concatenates the features from the three different pooling layers, thus constructing a single feature vector for each of the subjects. The dimensions of the resulting 3D feature vector are $1536 \times 25 \times 32 \times 23$.

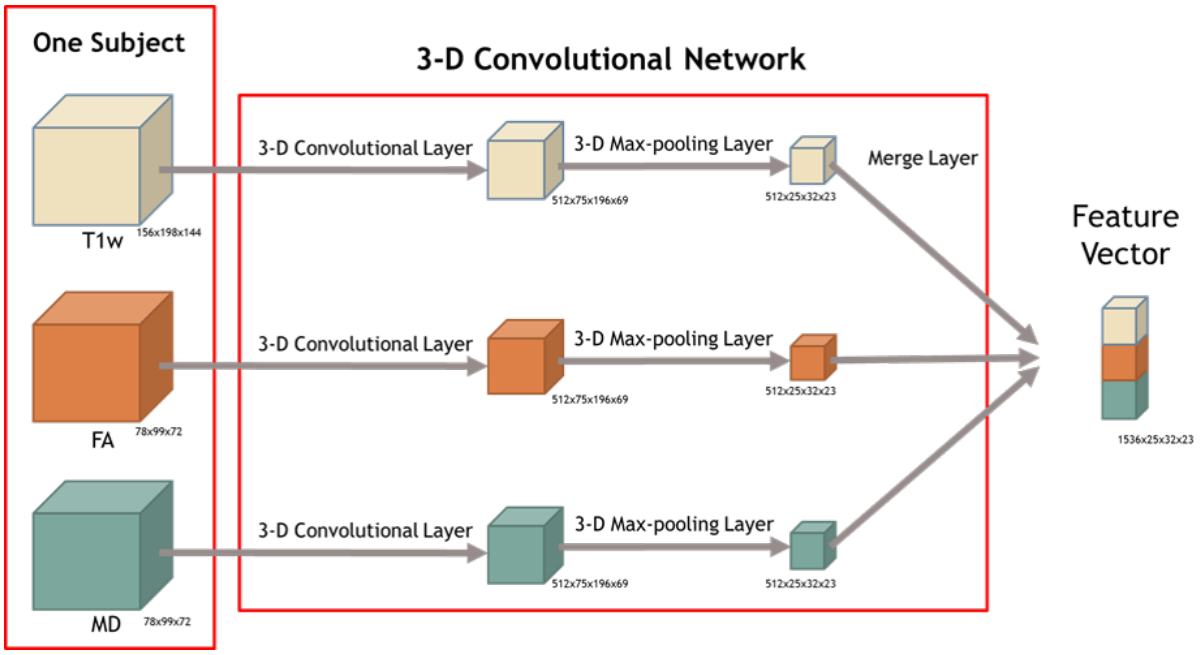


Figure 6.3: 3D Convolutional Network for Feature Extraction.

6.6 Group Clustering with Size Constraints

Ideally one would like to train the convolutional network to adjust the weights for discriminating the three different classes. However, given the number of data points (27 training subjects), this is currently not feasible. We explored training an SVM with cross-validation but the initial results were not promising. Instead, we now consider this problem as one of unsupervised clustering in the feature space computed by the above hand-tuned convolutional network.

For clustering, we use the standard K-means clustering with distance-based size-constraint, building upon the method described in [128]. However, Zhu et al. [128] do not provide a unique solution as it only uses the cluster labels. Instead, we modify the method to account for both labels and distances to the centroid as follows.

Given a dataset of N objects with P centroids (number of clusters), let $Dist$ be the $N \times P$ distance matrix,

$$Dist = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1P} \\ d_{21} & d_{22} & \dots & d_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ d_{N1} & x_{N2} & \dots & d_{NP} \end{bmatrix} \tag{6.1}$$

where d_{ip} is the distance between i object and p-th centroid. The objective is to compute a constrained $P \times N$ binary label matrix L,

$$L = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1N} \\ l_{21} & l_{22} & \dots & l_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ l_{P1} & l_{P2} & \dots & l_{PN} \end{bmatrix} \tag{6.2}$$

such that

$$\sum_{i=1}^{P} l_{ij} = 1, \quad j = 1, \ldots, n \quad \text{, and} \quad \sum_{j=1}^{N} l_{ij} = N_i, \quad i = 1, \ldots, p \tag{6.3}$$

where $l_{ij} = 1$ if the j-th object is assigned to cluster i, and cluster i is constrained to have exactly N_i points. This results in the following problem statement:

$$\text{minimize} \sum_{k=1}^{n} Dist_{(k)} L^{(k)} \tag{6.4}$$

where $Dist_{(k)}$ is the k^{th} row of $Dist$, and $L^{(k)}$ is the k^{th} column of L. This binary integer linear programming problem can be easily solved by any existing solver. The mTOP 2016 dataset has 27 subjects which are belonged to three different classes, and each class has nine subjects. Therefore, for this dataset, we set $N = 27, P = 3$, and $N_i = 9$ for each class.

6.7 Experimental Results and Discussion

Experiments are carried out with the following parameter settings: (i) whether to use whitening, (ii) the size of 3D patches (iii) the size of the 3D max-pooling kernel and (iv) the number of 3D features.

6.7.1 Evaluation Metrics

We use adjusted rand index (ARI) [129] and homogeneity score (HS) [130] to measure the performance. The adjusted rand index measures the similarity of two assignments (clustered labels vs. ground truth labels), which is invariant to permutations and normalized to chance. The similarity score is between 1.0 and -1.0. Random labelings have

an ARI close to 0.0, and 1.0 stands for a perfect match. Homogeneity score measures the purity of ground truth labels within a cluster. HS is between 1.0 and 0.0. 1.0 stands for perfectly homogeneous labeling.

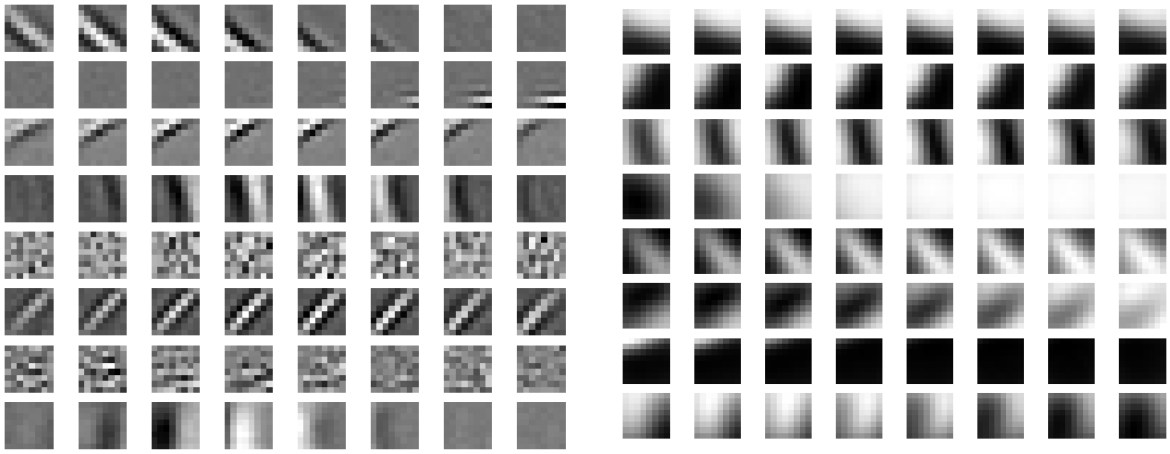


Figure 6.4: 3D features learned by K-means algorithm from MR-T1 images. Each row stands for a 3D feature and different columns stand for different axial planes. Left: Learned from whitened image patches. Right: Learned from un-whitened image patches.

6.7.2 Effect of whitening

In general, the whitening transformation helps improve accuracy. Figure 6.4 shows some example dictionary elements learned from K-means clustering and contrasts that to the original data. We observe that the ZCA transformation results in a sharper dictionary kernel. Figure 6.5 shows the clustering performance with and without whitening. The x-axis here shows the size of the dictionary. With the ZCA transform the results improve considerably as evidenced by the corresponding ARI and HS scores. This experiment used a stride size of 4 voxels and $8 \times 8 \times 8$ patch size for MR-T1 images, a stride size of 2 voxels and $4 \times 4 \times 4$ patch size for DT-FA and DT-MA image, and a $25 \times 32 \times 23$ kernel in the max-pooling layers.

6.7.3 Effect of 3D patch size

We also computed features at different 3D patch (volume) size settings and the results are plotted in Figure 6.6. Similar to the previous figure, the x-axis shows the size of the dictionary. The 3D feature size in the inset corresponds to the MR-T1 images. This experiment used ZCA transformed (whitened) data and $3 \times 3 \times 3$ kernels in max-pooling layers, 2 voxel stride size for MR-T1 image and 1 voxel stride size for DT-FA and DT-MD images. Overall, the $8 \times 8 \times 8$ features for MR-T1 image and the $4 \times 4 \times 4$ features for DT-FA and DT-MD image worked best. Therefore, increasing the max-pooling kernel decreased classification accuracy.

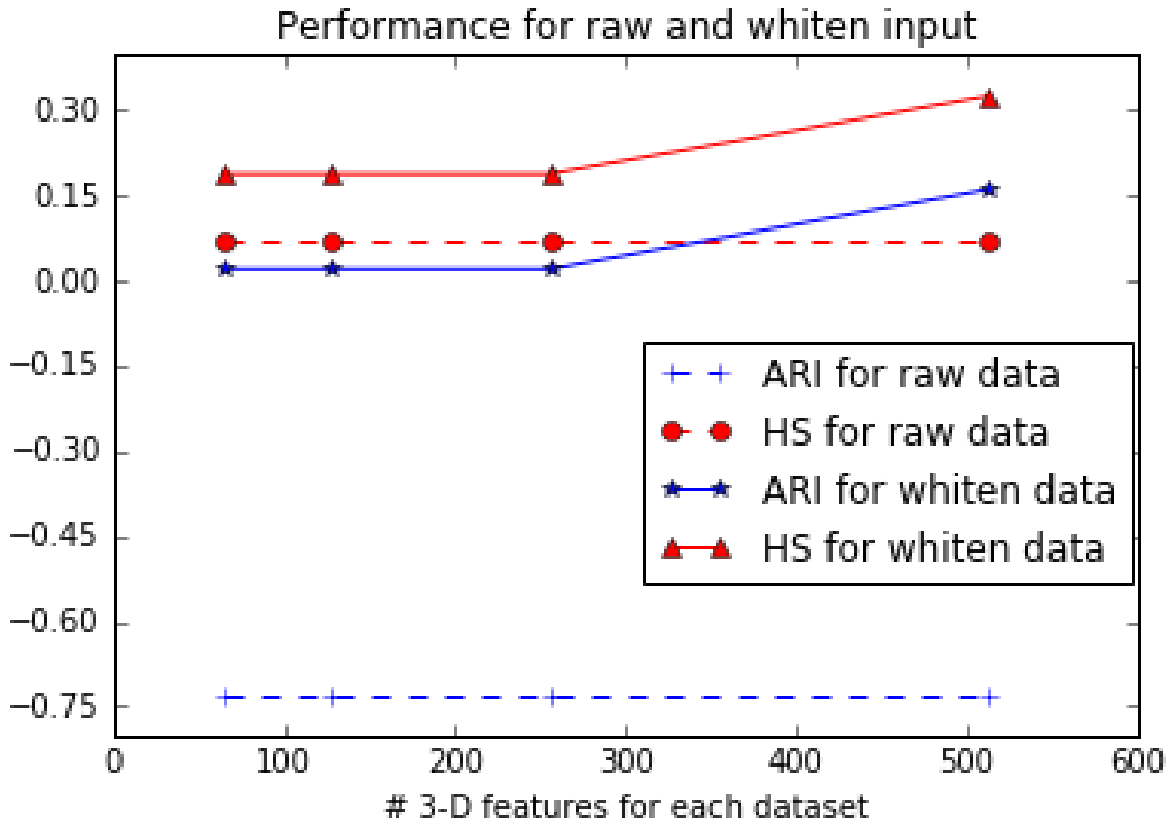


Figure 6.5: The effect of whitening

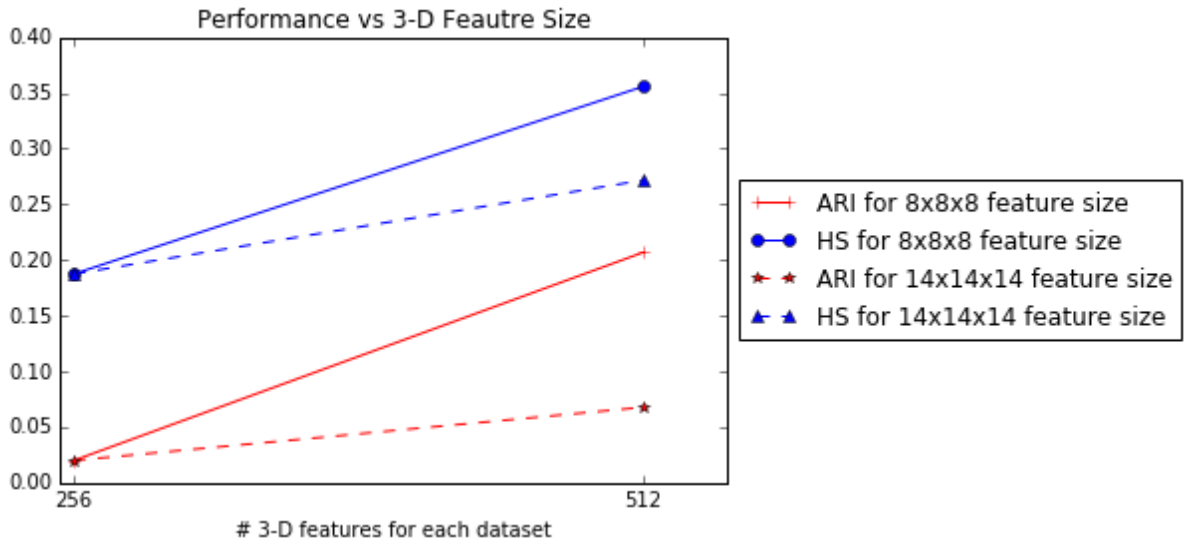


Figure 6.6: The effect of 3D features size

6.7.4 Effect of the size of 3D max-pooling kernel

In figure 6.7, we compared the results between $3 \times 3 \times 3$, and $25 \times 32 \times 23$ maximum pooling kernel size. The x-axis also shows the size of the dictionary. In our experiments, we observe that $3 \times 3 \times 3$ maximum pooling kernels have the best performance. This experiment used whitened datasets, $8 \times 8 \times 8$ feature kernels, and a stride size of 2 voxels for MR-T1 image, and $4 \times 4 \times 4$ feature kernels and 1 voxel for DT-FA and DT-MD images.

6.7.5 Effect of Dictionary size

We considered feature representations with 64, 128, 256, and, 512 3D dictionary items. Figures 6.5, 6.6 and 6.7 clearly show that a dictionary size of 512 gives the best results. Going beyond 512 did not result in much improvement.

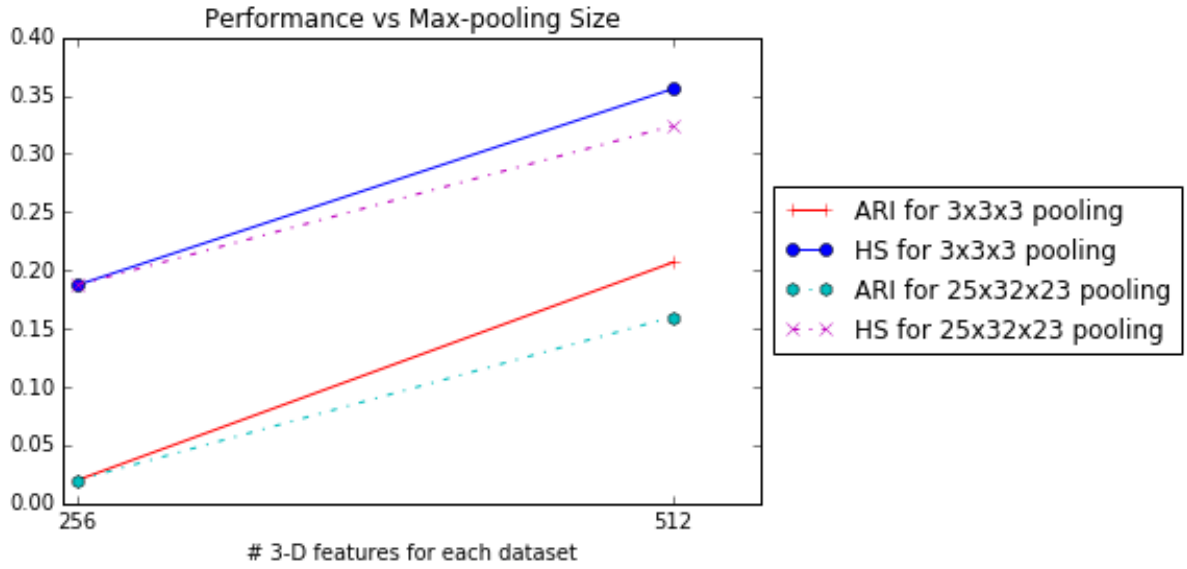


Figure 6.7: The effect of max-pooling size

6.8 Summary

We explored the unsupervised classification of the mTBI challenge dataset. Given the small number of samples, it is not feasible to train a deep learning network for feature extraction and classification. Instead, we focused on computing volume features and using them for classification. In the end, the best classification results correctly classified 10 out of 15 samples for which the labels are known, and the corresponding unsupervised clustering scores are ARI= 0.267 and HS = 0.3556. The proposed method was placed third in the MICCAI-mTOP 2016 challenge. The final publication of this chapter is available at Springer via `https://doi.org/10.1007/978-3-319-55524-9_26`.

Chapter 7

Conclusions and Future Work

> The past is a place of reference, not a place of residence; the past is a place of learning, not a place of living.
>
> *Roy T. Bennett*

In this dissertation, we focused on predicting the outcome of patients with brain disorders including brain tumor, stroke, and mild traumatic brain injury. For the patients with a brain lesion, we propose a novel tractographic feature to capture the potentially damaged regions due to the presence of the brain lesion. The tractographic feature is built from the brain lesion and average connectome information from a group of normal subjects. The tractographic feature takes into account different functional regions that may be affected by the lesion, thus complementing the commonly used lesion volume features. The tractographic feature is tested on a public brain tumor dataset and a public ischemic stroke dataset and achieves a better outcome prediction performance than the gold standards, other commonly used features, and the state-of-the-art features [63, 131].

To automate the outcome prediction pipeline using the tractographic feature, a

brain lesion segmentation is needed. Towards addressing this problem, we also propose a location information fusion method which integrates location information of the brain with the state-of-the-art patch-based neural networks including 3D U-Net [26] and DeepMedic [27, 28]. The proposed feature fusion method is tested on a public brain tumor segmentation dataset, and it improves the segmentation performance of these patch-based neural networks. Also, our proposed ensemble won the 6th place out of 63 teams in BraTS 2018 challenge [63].

For predicting the outcome of mild traumatic brain patients, we explore the unsupervised classification of the mTBI challenge dataset. Given the small number of samples, it is not feasible to train a deep learning network for feature extraction and classification. Instead, we focused on computing volumetric features and using it for classification in an unsupervised manner. The proposed method was ranked third in the MICCAI-mTOP 2016 challenge [132].

7.1 Future Directions

Brain Lesion Segmentation

Squeeze-and-Excitation (SE) [133] block, which adaptively recalibrates channel-wise feature responses by explicitly modeling interdependencies between channels, is shown to bring significant improvements in performance for existing state-of-the-art convolutional neural networks at a slight additional computational cost. The integration of SE unit, brain location information, and the state-of-the-art patch-based neural networks has the potential to significantly improve the segmentation performance of existing state-of-the-art patch-based neural networks.

In addition, utilizing a long short-term memory (LSTM) [134], a bidirectional LSTM [135] or a Convolutional LSTM [136] to learn the 3D spatial relations of brain with 3D

convolutional neural networks seems a promising direction for the brain lesion segmentation. Also, the brain tumor segmentation performance of 3D U-Net has a significant improvement by utilizing lesion prior directly [137] but the lesion prior fusion method needs to be validated in other patch-based neural network architectures.

Tractographic Feature

The proposed tractographic feature is built from average connectome information of a group of normal subjects. Another approach is to build the disruption matrices for every normal subject given the brain lesion mask and calculate the average disruption matrices based on individual matrices. This approach has the potential to improve the outcome prediction performance of the tractographic feature but it takes more time to generate the fiber tracts from each subject. Also, using the tractography-based atlases to generate the disruption matrix has the potential to improve the outcome performance of the tractographic feature. Besides, the tractographic feature can also apply to the traumatic brain injury (TBI) patients with hematoma to predict their outcome. The computerized tomography (CT) scan is commonly used to diagnose the TBI patients. We have implemented a registration tool [138] which maps the CT scan of head in the subject space to the MNI 152 space [139]. This tool is a good starting point to extract the proposed tractographic feature for TBI patients with CT scans. In addition, if the patient-specific tractography data is available, we can apply similar concepts to construct the patient-specific tractographic features. The expectation is that these patient-specific tractographic features will provide more accurate information for predictive computation compared to the average HCP-based tractographic features.

Bibliography

[1] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki, *et. al.*, *Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge*, *arXiv preprint arXiv:1811.02629* (2018).

[2] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos., *Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection.*, 2017.

[3] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos., *Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection.*, 2017.

[4] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, *Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features*, *Scientific data* **4** (2017) 170117.

[5] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, *Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain*, *Neuroimage* **15** (2002), no. 1 273–289.

[6] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, *et. al.*, *An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest*, *Neuroimage* **31** (2006), no. 3 968–980.

[7] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, *et. al.*, *3d slicer as an image computing platform for the quantitative imaging network*, *Magnetic resonance imaging* **30** (2012), no. 9 1323–1341.

[8] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J. C. Gee, and G. Gerig, *User-guided 3D active contour segmentation of anatomical*

structures: Significantly improved efficiency and reliability, *Neuroimage* **31** (2006), no. 3 1116–1128.

[9] O. Maier and H. Handels, *Predicting stroke lesion and clinical outcome with random forests*, in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 219–230, Springer, 2016.

[10] M. Xia, J. Wang, and Y. He, *Brainnet viewer: a network visualization tool for human brain connectomics*, *PloS one* **8** (2013), no. 7 e68910.

[11] R. T. Higashida and A. J. Furlan, *Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke*, *Stroke* **34** (2003), no. 8 e109–e137.

[12] J. Fugate, A. Klunder, and D. F. Kallmes, *What is meant by "tici"?*, *American Journal of Neuroradiology* **34** (2013), no. 9 1792–1797.

[13] J. Van Swieten, P. Koudstaal, M. Visser, H. Schouten, and J. Van Gijn, *Interobserver agreement for the assessment of handicap in stroke patients.*, *Stroke* **19** (1988), no. 5 604–607.

[14] "Brain disorders." `https://www.healthline.com/health/brain-disorders`. Accessed: 2019-08-26.

[15] "Brain tumor." `https://www.healthline.com/health/brain-tumor`. Accessed: 2019-08-26.

[16] D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, B. W. Scheithauer, and P. Kleihues, *The 2007 who classification of tumours of the central nervous system*, *Acta neuropathologica* **114** (2007), no. 2 97–109.

[17] "About stroke." `https://www.stroke.org/en/about-stroke`. Accessed: 2019-08-26.

[18] "Traumatic brain injury." `https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Traumatic-Brain-Injury`. Accessed: 2019-08-26.

[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, *Imagenet: A large-scale hierarchical image database*, in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[20] H. Garavan *et. al.*, *Recruiting the abcd sample: Design considerations and procedures*, *Developmental cognitive neuroscience* **32** (2018) 16–22.

[21] D. J. Hagler *et. al.*, *Image processing and analysis methods for the adolescent brain cognitive development study*, *bioRxiv* (2018).

[22] M. Luciana *et. al.*, *Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (abcd) baseline neurocognition battery*, *Developmental cognitive neuroscience* **32** (2018) 67–79.

[23] A. Pfefferbaum *et. al.*, *Altered brain developmental trajectories in adolescents after initiating drinking*, *American journal of psychiatry* **175** (2017), no. 4 370–380.

[24] N. D. Volkow *et. al.*, *The conception of the abcd study: From substance use to a broad nih collaboration*, *Developmental cognitive neuroscience* **32** (2018) 4–7.

[25] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, *et. al.*, *A large annotated medical image dataset for the development and evaluation of segmentation algorithms*, *arXiv preprint arXiv:1902.09063* (2019).

[26] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, *3d u-net: learning dense volumetric segmentation from sparse annotation*, in *International conference on medical image computing and computer-assisted intervention*, pp. 424–432, Springer, 2016.

[27] K. Kamnitsas, E. Ferrante, S. Parisot, C. Ledig, A. V. Nori, A. Criminisi, D. Rueckert, and B. Glocker, *Deepmedic for brain tumor segmentation*, in *International workshop on Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries*, pp. 138–149, Springer, 2016.

[28] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, *Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation*, *Medical image analysis* **36** (2017) 61–78.

[29] F. G. Davis, S. Freels, J. Grutsch, S. Barlas, and S. Brem, *Survival rates in patients with primary malignant brain tumors stratified by patient age and tumor histological type: an analysis based on surveillance, epidemiology, and end results (seer) data, 1973–1991*, *Journal of neurosurgery* **88** (1998), no. 1 1–10.

[30] A. Flowers, *Brain tumors in the older person*, *Cancer control* **7** (2000), no. 6 523–538.

[31] M. Pogorzala, J. Styczynski, and M. Wysocki, *Survival and prognostic factors in children with brain tumors: long-term follow-up single center study in poland*, *Anticancer research* **34** (2014), no. 1 323–326.

[32] L. Weninger, O. Rippel, S. Koppers, and D. Merhof, *Segmentation of brain tumors and patient survival prediction: Methods for the brats 2018 challenge*, in *International MICCAI Brainlesion Workshop*, pp. 3–12, Springer, 2018.

[33] M. Kistler, S. Bonaretti, M. Pfahrer, R. Niklaus, and P. Büchler, *The virtual skeleton database: an open access repository for biomedical research and collaboration*, *Journal of medical Internet research* **15** (2013), no. 11.

[34] O. Maier, B. H. Menze, J. von der Gablentz, L. Häni, M. P. Heinrich, M. Liebrand, S. Winzeck, A. Basit, P. Bentley, L. Chen, *et. al.*, *Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri*, *Medical image analysis* **35** (2017) 250–269.

[35] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, *et. al.*, *Advances in functional and structural mr image analysis and implementation as fsl*, *Neuroimage* **23** (2004) S208–S219.

[36] M. W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, and S. M. Smith, *Bayesian analysis of neuroimaging data in fsl*, *Neuroimage* **45** (2009), no. 1 S173–S186.

[37] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, *Fsl*, *Neuroimage* **62** (2012), no. 2 782–790.

[38] "Dsi studio." `http://dsi-studio.labsolver.org/`. Accessed: 2019-12-11.

[39] M. Symms, H. Jäger, K. Schmierer, and T. Yousry, *A review of structural magnetic resonance neuroimaging*, *Journal of Neurology, Neurosurgery & Psychiatry* **75** (2004), no. 9 1235–1244.

[40] N. D. Volkow, G.-J. Wang, J. S. Fowler, D. Tomasi, and R. Baler, *Neuroimaging of addiction*, in *Imaging of the human brain in health and disease*, pp. 1–26. Elsevier, 2014.

[41] "Structural mri imaging." `http://fmri.ucsd.edu/Howto/3T/structure.html`. Accessed: 2019-09-19.

[42] "Magnetic resonance imaging (mri) of the brain and spine: Basics." `https://casemed.case.edu/clerkships/neurology/Web%20Neurorad/MRI%20Basics.htm`. Accessed: 2019-09-19.

[43] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et. al.*, *The multimodal brain tumor image segmentation benchmark (brats)*, *IEEE transactions on medical imaging* **34** (2015), no. 10 1993–2024.

[44] G. Grabner *et. al.*, *Symmetric atlasing and model based segmentation: an application to the hippocampus in older adults*, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 58–66, Springer, 2006.

[45] "Nih launches the human connectome project to unravel the brain's connections." https://www.nih.gov/news-events/news-releases/nih-launches-human-connectome-project-unravel-brains-connections. Accessed: 2019-09-10.

[46] "The human connectome project." https://neuroscienceblueprint.nih.gov/human-connectome/connectome-programs. Accessed: 2019-09-10.

[47] "Mind mapping: Inside the brain's wiring." https://www.bbc.com/news/av/science-environment-21489097/mind-mapping-inside-the-brain-s-wiring. Accessed: 2019-09-10.

[48] L. Geddes, *Human brain mapped in unprecedented detail*, *Nature* (2016).

[49] "$40 million awarded to trace human brain's connections." https://www.nih.gov/news-events/news-releases/40-million-awarded-trace-human-brains-connections. Accessed: 2019-09-10.

[50] "1200 subjects data release." https://www.humanconnectome.org/study/hcp-young-adult/document/1200-subjects-data-release. Accessed: 2019-09-11.

[51] M. Jenkinson and S. Smith, *A global optimisation method for robust affine registration of brain images*, *Medical image analysis* **5** (2001), no. 2 143–156.

[52] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, *Improved optimization for the robust and accurate linear registration and motion correction of brain images*, *Neuroimage* **17** (2002), no. 2 825–841.

[53] D. N. Greve and B. Fischl, *Accurate and robust brain image alignment using boundary-based registration*, *Neuroimage* **48** (2009), no. 1 63–72.

[54] F.-C. Yeh, T. D. Verstynen, Y. Wang, J. C. Fernández-Miranda, and W.-Y. I. Tseng, *Deterministic diffusion fiber tracking improved by quantitative anisotropy*, *PloS one* **8** (2013), no. 11 e80713.

[55] R. Noiphithak and K. Veerasarn, *Clinical predictors for survival and treatment outcome of high-grade glioma in prasat neurological institute*, *Asian journal of neurosurgery* **12** (2017), no. 1 28.

[56] K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, *et. al.*, *Ensembles of multiple models and architectures for robust brain tumour segmentation*, in *International MICCAI Brainlesion Workshop*, pp. 450–462, Springer, 2017.

[57] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, *Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks*, in *International MICCAI Brainlesion Workshop*, pp. 178–190, Springer, 2017.

[58] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, *Brain tumor segmentation and radiomics survival prediction: contribution to the brats 2017 challenge*, in *International MICCAI Brainlesion Workshop*, pp. 287–297, Springer, 2017.

[59] A. Myronenko, *3d mri brain tumor segmentation using autoencoder regularization*, in *International MICCAI Brainlesion Workshop*, pp. 311–320, Springer, 2018.

[60] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, *No new-net*, in *International MICCAI Brainlesion Workshop*, pp. 234–244, Springer, 2018.

[61] R. McKinley, R. Meier, and R. Wiest, *Ensembles of densely-connected cnns with label-uncertainty for brain tumor segmentation*, in *International MICCAI Brainlesion Workshop*, pp. 456–465, Springer, 2018.

[62] C. Zhou, S. Chen, C. Ding, and D. Tao, *Learning contextual and attentive information for brain tumor segmentation*, in *International MICCAI Brainlesion Workshop*, pp. 497–507, Springer, 2018.

[63] P.-Y. Kao, T. Ngo, A. Zhang, J. W. Chen, and B. Manjunath, *Brain tumor segmentation and tractographic feature extraction from structural mr images for overall survival prediction*, in *International MICCAI Brainlesion Workshop*, pp. 128–141, Springer, 2018.

[64] K. Kamnitsas, L. Chen, C. Ledig, D. Rueckert, and B. Glocker, *Multi-scale 3d convolutional neural networks for lesion segmentation in brain mri*, *Ischemic stroke lesion segmentation* **13** (2015) 13–16.

[65] O. Ronneberger, P. Fischer, and T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[66] J. Jiang, P.-Y. Kao, S. A. Belteton, D. B. Szymanski, and B. Manjunath, *Accurate 3d cell segmentation using deep feature and crf refinement*, *arXiv preprint arXiv:1902.04729* (2019).

[67] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, *H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes*, *IEEE transactions on medical imaging* **37** (2018), no. 12 2663–2674.

[68] L. Yu, X. Yang, H. Chen, J. Qin, and P. A. Heng, *Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images*, in *Thirty-first AAAI conference on artificial intelligence*, 2017.

[69] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, *arXiv preprint arXiv:1502.03167* (2015).

[70] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[71] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, in *International Conference on Learning Representations*, 2015.

[72] S. J. Reddi *et. al.*, *On the convergence of adam and beyond*, in *International Conference on Learning Representations*, 2018.

[73] Y. Wu and K. He, *Group normalization*, in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19, 2018.

[74] P. Bühlmann, *Bagging, boosting and ensemble methods*, in *Handbook of Computational Statistics*, pp. 985–1022. Springer, 2012.

[75] T. G. Dietterich, *Ensemble methods in machine learning*, in *International workshop on multiple classifier systems*, pp. 1–15, Springer, 2000.

[76] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, *Journal of computer and system sciences* **55** (1997), no. 1 119–139.

[77] T. Hastie, S. Rosset, J. Zhu, and H. Zou, *Multi-class adaboost*, *Statistics and its Interface* **2** (2009), no. 3 349–360.

[78] J. H. Friedman, *Greedy function approximation: a gradient boosting machine*, *Annals of statistics* (2001) 1189–1232.

[79] J. H. Friedman, *Stochastic gradient boosting*, *Computational statistics & data analysis* **38** (2002), no. 4 367–378.

[80] T. Chen and C. Guestrin, *Xgboost: A scalable tree boosting system*, in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, ACM, 2016.

[81] B. Lakshminarayanan, A. Pritzel, and C. Blundell, *Simple and scalable predictive uncertainty estimation using deep ensembles*, in *Advances in Neural Information Processing Systems*, pp. 6402–6413, 2017.

[82] D. Nielsen, *Tree boosting with xgboost-why does xgboost win" every" machine learning competition?*, Master's thesis, NTNU, 2016.

[83] J. P. Thakkar, T. A. Dolecek, C. Horbinski, Q. T. Ostrom, D. D. Lightner, J. S. Barnholtz-Sloan, and J. L. Villano, *Epidemiologic and molecular prognostic review of glioblastoma*, *Cancer Epidemiology and Prevention Biomarkers* (2014).

[84] Z. A. Shboul, L. Vidyaratne, M. Alam, and K. M. Iftekharuddin, *Glioblastoma and survival prediction*, in *International MICCAI Brainlesion Workshop*, pp. 358–368, Springer, 2017.

[85] A. Jungo, R. McKinley, R. Meier, U. Knecht, L. Vera, J. Pérez-Beteta, D. Molina-García, V. M. Pérez-García, R. Wiest, and M. Reyes, *Towards uncertainty-assisted brain tumor segmentation and survival prediction*, in *International MICCAI Brainlesion Workshop*, pp. 474–485, Springer, 2017.

[86] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium, *et. al.*, *The wu-minn human connectome project: an overview*, *Neuroimage* **80** (2013) 62–79.

[87] F.-C. Yeh and W.-Y. I. Tseng, *Ntu-90: a high angular resolution brain atlas constructed by q-space diffeomorphic reconstruction*, *Neuroimage* **58** (2011), no. 1 91–99.

[88] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine learning in Python*, *Journal of Machine Learning Research* **12** (2011) 2825–2830.

[89] W. Johnson, O. Onuma, M. Owolabi, and S. Sachdev, *Stroke: a global response is needed*, *Bulletin of the World Health Organization* **94** (2016), no. 9 634.

[90] R. I. Lindley, *Stroke*. Oxford University Press, 2 ed., 2017.

[91] A. Maasz and B. Melegh, *Three periods of one and a half decade of ischemic stroke susceptibility gene research: lessons we have learned*, *Genome medicine* **2** (2010), no. 9 64.

[92] J. L. Banks and C. A. Marotta, *Outcomes validity and reliability of the modified rankin scale: implications for stroke clinical trials: a literature review and synthesis*, *Stroke* **38** (2007), no. 3 1091–1096.

[93] M. Goyal, B. K. Menon, W. H. Van Zwam, D. W. Dippel, P. J. Mitchell, A. M. Demchuk, A. Dávalos, C. B. Majoie, A. van der Lugt, M. A. De Miquel, *et. al.*, *Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials*, *The Lancet* **387** (2016), no. 10029 1723–1731.

[94] J. L. Wilson, A. Hareendran, M. Grant, T. Baird, U. G. Schulz, K. W. Muir, and I. Bone, *Improving the assessment of outcomes in stroke: use of a structured interview to assign grades on the modified rankin scale*, *Stroke* **33** (2002), no. 9 2243–2246.

[95] L. Derex, N. Nighoghossian, M. Hermier, P. Adeleine, Y. Berthezène, F. Philippeau, J. Honnorat, J.-C. Froment, and P. Trouillas, *Influence of pretreatment mri parameters on clinical outcome, recanalization and infarct size in 49 stroke patients treated by intravenous tissue plasminogen activator*, *Journal of the neurological sciences* **225** (2004), no. 1-2 3–9.

[96] M. Kluytmans, K. Van Everdingen, L. Kappelle, L. Ramos, M. Viergever, and J. Van Der Grond, *Prognostic value of perfusion-and diffusion-weighted mr imaging in first 3 days of stroke*, *European radiology* **10** (2000), no. 9 1434–1441.

[97] M. H. Lev, A. Z. Segal, J. Farkas, S. T. Hossain, C. Putman, G. J. Hunter, R. Budzik, G. J. Harris, F. S. Buonanno, M. A. Ezzeddine, *et. al.*, *Utility of perfusion-weighted ct imaging in acute middle cerebral artery stroke treated with intra-arterial thrombolysis:: Prediction of final infarct volume and clinical outcome*, *Stroke* **32** (2001), no. 9 2021–2028.

[98] M. Parsons, T. Li, P. Barber, Q. Yang, D. Darby, P. Desmond, R. Gerraty, B. Tress, and S. Davis, *Combined 1h mr spectroscopy and diffusion-weighted mri improves the prediction of stroke outcome*, *Neurology* **55** (2000), no. 4 498–506.

[99] K. Van Everdingen, J. Van der Grond, L. Kappelle, L. Ramos, and W. Mali, *Diffusion-weighted magnetic resonance imaging in acute stroke*, *Stroke* **29** (1998), no. 9 1783–1790.

[100] N. Vila, X. Filella, R. Deulofeu, C. Ascaso, R. Abellana, and A. Chamorro, *Cytokine-induced inflammation and long-term stroke functional outcome*, *Journal of the neurological sciences* **162** (1999), no. 2 185–188.

[101] G. Vogt, R. Laage, A. Shuaib, and A. Schneider, *Initial lesion volume is an independent predictor of clinical stroke outcome at day 90: an analysis of the virtual international stroke trials archive (vista) database*, *Stroke* **43** (2012), no. 5 1266–1272.

[102] N. D. Forkert, T. Verleger, B. Cheng, G. Thomalla, C. C. Hilgetag, and J. Fiehler, *Multiclass support vector machine-based lesion mapping predicts functional outcome in ischemic stroke patients*, *PloS one* **10** (2015), no. 6 e0129569.

[103] Y. Choi, Y. Kwon, H. Lee, B. J. Kim, M. C. Paik, and J.-H. Won, *Ensemble of deep convolutional neural networks for prognosis of ischemic stroke*, in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 231–243, Springer, 2016.

[104] Q. Mahmood and A. Basit, *Prediction of ischemic stroke lesion and clinical outcome in multi-modal mri images using random forests*, in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 244–255, Springer, 2016.

[105] A. I. Qureshi, S. A. Chaudhry, B. L. Sapkota, G. J. Rodriguez, and M. F. K. Suri, *Discharge destination as a surrogate for modified rankin scale defined outcomes at 3-and 12-months poststroke among stroke survivors*, *Archives of physical medicine and rehabilitation* **93** (2012), no. 8 1408–1413.

[106] F.-C. Yeh, S. Panesar, D. Fernandes, A. Meola, M. Yoshino, J. C. Fernandez-Miranda, J. M. Vettel, and T. Verstynen, *Population-averaged atlas of the macroscale human structural connectome and its network topology*, *NeuroImage* **178** (2018) 57–68.

[107] F.-C. Yeh, S. Panesar, J. Barrios, D. Fernandes, K. Abhinav, A. Meola, and J. C. Fernandez-Miranda, *Automatic removal of false connections in diffusion mri tractography using topology-informed pruning (tip)*, *Neurotherapeutics* **16** (Jan, 2019) 52–58.

[108] L. Liu, H. Zhang, J. Wu, Z. Yu, X. Chen, I. Rekik, Q. Wang, J. Lu, and D. Shen, *Overall survival time prediction for high-grade glioma patients based on large-scale brain functional networks*, *Brain imaging and behavior* (2018) 1–19.

[109] S. M. Antonucci, P. M. Beeson, D. M. Labiner, and S. Z. Rapcsak, *Lexical retrieval and semantic knowledge in patients with left inferior temporal lobe lesions*, *Aphasiology* **22** (2008), no. 3 281–304.

[110] M. G. Knyazeva, *Splenium of corpus callosum: patterns of interhemispheric interaction in children and adults*, *Neural plasticity* **2013** (2013).

[111] N. Yassi, L. Churilov, B. C. Campbell, G. Sharma, R. Bammer, P. M. Desmond, M. W. Parsons, G. W. Albers, G. A. Donnan, and S. M. Davis, *The association between lesion location and functional outcome after ischemic stroke*, *International Journal of Stroke* **10** (2015), no. 8 1270–1276.

[112] M. Giroud, M. Lemesle, G. Madinier, T. Billiar, and R. Dumas, *Unilateral lenticular infarcts: radiological and clinical syndromes, aetiology, and prognosis*, *Journal of Neurology, Neurosurgery & Psychiatry* **63** (1997), no. 5 611–615.

[113] V. S. Hedna, A. N. Bodhit, S. Ansari, A. D. Falchook, L. Stead, K. M. Heilman, and M. F. Waters, *Hemispheric differences in ischemic stroke: is left-hemisphere stroke more common?*, *Journal of Clinical Neurology* **9** (2013), no. 2 97–102.

[114] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, *Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition*, *IEEE Transactions on Information Forensics and Security* **11** (2016), no. 9 1984–1996.

[115] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, *Fully automatic face normalization and single sample face recognition in unconstrained environments*, *Expert Systems with Applications* **47** (2016) 23–34.

[116] M. Liu, D. Zhang, D. Shen, and A. D. N. Initiative, *Hierarchical fusion of features and classifier decisions for alzheimer's disease diagnosis*, *Human brain mapping* **35** (2014), no. 4 1305–1319.

[117] H.-I. Suk, S.-W. Lee, D. Shen, A. D. N. Initiative, *et. al.*, *Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis*, *NeuroImage* **101** (2014) 569–582.

[118] A. Gardner and R. Zafonte, *Neuroepidemiology of traumatic brain injury*, in *Handbook of clinical neurology*, vol. 138, pp. 207–223. Elsevier, 2016.

[119] J. A. Langlois, W. Rutland-Brown, and M. M. Wald, *The epidemiology and impact of traumatic brain injury: a brief overview*, *The Journal of head trauma rehabilitation* **21** (2006), no. 5 375–378.

[120] R. Bellotti, A. Lombardi, C. Guaragnella, N. Amoroso, A. Tateo, and S. Tangaro, *Mild traumatic brain injury outcome prediction based on both graph and k-nn methods*, in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 271–281, Springer, 2016.

[121] Y. Cai and S. Ji, *Combining deep learning networks with permutation tests to predict traumatic brain injury outcome*, in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 259–270, Springer, 2016.

[122] A. Coates, A. Ng, and H. Lee, *An analysis of single-layer networks in unsupervised feature learning*, in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 215–223, 2011.

[123] A. Coates and A. Y. Ng, *Learning feature representations with k-means*, in *Neural networks: Tricks of the trade*, pp. 561–580. Springer, 2012.

[124] G. E. Hinton, S. Osindero, and Y.-W. Teh, *A fast learning algorithm for deep belief nets*, *Neural computation* **18** (2006), no. 7 1527–1554.

[125] B. A. Olshausen and D. J. Field, *Emergence of simple-cell receptive field properties by learning a sparse code for natural images*, *Nature* **381** (1996), no. 6583 607.

[126] A. Krizhevsky, *Learning multiple layers of features from tiny images*, tech. rep., Citeseer, 2009.

[127] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, *An efficient k-means clustering algorithm: Analysis and implementation*, *IEEE Transactions on Pattern Analysis & Machine Intelligence* (2002), no. 7 881–892.

[128] S. Zhu, D. Wang, and T. Li, *Data clustering with size constraints*, *Knowledge-Based Systems* **23** (2010), no. 8 883–889.

[129] L. Hubert and P. Arabie, *Comparing partitions*, *Journal of classification* **2** (1985), no. 1 193–218.

[130] A. Rosenberg and J. Hirschberg, *V-measure: A conditional entropy-based external cluster evaluation measure*, in *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*, 2007.

[131] P.-Y. Kao, J. W. Chen, and B. Manjunath, *Predicting clinical outcome of stroke patients with tractographic feature*, *arXiv preprint arXiv:1907.10419* (2019).

[132] P.-Y. Kao, E. Rojas, J. W. Chen, A. Zhang, and B. Manjunath, *Unsupervised 3-d feature learning for mild traumatic brain injury*, in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 282–290, Springer, 2016.

[133] J. Hu, L. Shen, and G. Sun, *Squeeze-and-excitation networks*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.

[134] S. Hochreiter and J. Schmidhuber, *Long short-term memory*, *Neural computation* **9** (1997), no. 8 1735–1780.

[135] A. Graves and J. Schmidhuber, *Framewise phoneme classification with bidirectional lstm and other neural network architectures*, *Neural networks* **18** (2005), no. 5-6 602–610.

[136] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, *Convolutional lstm network: A machine learning approach for precipitation nowcasting*, in *Advances in neural information processing systems*, pp. 802–810, 2015.

[137] P.-Y. Kao, J. W. Chen, and B. Manjunath, *Improving 3d u-net for brain tumor segmentation by utilizing lesion prior*, *arXiv preprint arXiv:1907.00281* (2019).

[138] H. J. Kuijf, J. M. Biesbroek, M. A. Viergever, G. J. Biessels, and K. L. Vincken, *Registration of brain ct images to an mri template for the purpose of lesion-symptom mapping*, in *International Workshop on Multimodal Brain Image Analysis*, pp. 119–128, Springer, 2013.

[139] P.-Y. Kao, *CT2MNI152: First release of the CT to MNI 152 space registration tool*, 2019.