© 2021 Yucheng Liang

LEARNING-BASED TUMOR SEGMENTATION USING METABOLIC IMAGING FEATURES

BY

YUCHENG LIANG

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Adviser:

Professor Zhi-Pei Liang

ABSTRACT

The ultimate objective of brain tumor imaging is to distill patient-specific knowledge that guides therapy planning and medical care. Magnetic resonance imaging (MRI) is a prevailing technique to visualize tumors due to its excellent contrast of soft tissue and non-invasiveness. Decades of research have helped brain tumor segmentation performance dramatically. However, precise segmentation is still considered hard partly due to the limitation in resolution, signal-to-noise ratio, and possible artifacts. While some tumors are easier to delineate, more infiltrating ones like gliomas have ragged and obscure boundaries that are harder to define. In recognition of this hardship, researchers have started exploring the use of Proton Magnetic Resonance Spectroscopic Imaging (MRSI) for better tumor prognosis, diagnosis, and characterization.

MRSI investigates the spatial distribution of metabolic changes by leveraging its unique temporal information. The wealth of this spectroscopic information is beneficial in classifying tumor subregions and aiding ongoing research investigations in tumor heterogeneity and related topics. Several studies have reported an increase in choline-containing compounds level and a reduced N-acetyl-aspartate level in brain tumors. Spectroscopic techniques can pick up these metabolic changes, and they might be the missing pieces of better MRI-based tumor segmentation solutions.

This study shows a successful application of deep learning and MRSI to identify tumor and edema regions of human brains with glioblastomas. The deep learning framework of choice is nnU-Net. Most specialized solutions in applying deep neural models in the medical image domain depend on dataset properties and hardware constraints. nnU-Net is a framework that automatically adapts itself to various medical image segmentation tasks. Therefore, it ensures a fair comparison of experiments. This work shows an improved segmentation result after incorporating high-resolution metabolite maps de-

rived from MRSI data acquired by the SPICE sequence. The high resolution, rapidness, and near whole-brain performance of SPICE should assist radiologists and oncologists in delimiting the pathological area better and providing more appropriate medical help.

To my family, for their love and support.

ACKNOWLEDGMENTS

First, I thank my advisor, Dr. Zhi-Pei Liang, for his guidance and support throughout this research and my higher education. His unwavering enthusiasm for science and technologies has deeply affected me. His suggestions have helped me tremendously in becoming better at research and work.

My sincerest thanks also go to our collaborator, Dr. Yao Li, for her mentorship and wisdom. I also thank my excellent colleagues, especially Mengmeng Cao, Rong Guo, Yudu Li, Ruihao Liu, Ziyu Meng, Chengyan Wang, and Yibo Zhao, for their help in acquiring the experimental data and sharing knowledge and insights.

I also appreciate the arrival of Gatlin, my pet rabbit. His presence has enlightened my quarantine life during the COVID time.

Finally, I would like to thank my family and many friends for their pouring love, understanding, and support.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Motivations

In conventional MRI, finding the precise boundary of gliomas is challenging due to the heterogeneous structure of the affected areas [1]. This heterogeneous nature confuses the automatic methods to overestimate or underestimate the presence of active tumor [2], and such a lousy outcome can lead to undesired treatment planning and prognosis consequences. Proton magnetic resonance spectroscopic imaging (1H-MRSI) can potentially remedy this challenging situation because it visualizes the metabolites distribution in human bodies. MRSI marks a fundamental paradigm shift from structural to metabolic imaging. Several studies have reported increased choline (Cho)-containing compounds level and the reduced presence of N-acetyl-aspartate (NAA) in brain tumors. Choline-containing compounds include choline, phosphocholine, and glycerophosphocholine. They increase as the level of membrane synthesis in rapidly dividing tumor cells elevates. NAA is regarded as a neuronal marker mainly contained within neurons, not in tumors. Therefore, it decreases in these abnormal regions. In principle, metabolic maps of NAA and Cho provide rich information differentiating necrosis, solid tumor, and varying degrees of tumor infiltration and tissue edema. In this thesis, we investigate the potential of metabolite maps for more precise brain tumor segmentation.

1.2 Problem Formulation

Brain tumor segmentation requires the separation of tumor and edema from normal brain tissues (gray matter, white matter, and cerebrospinal fluid).

In this study, we compare structural scans only and a combination of structural scans and metabolite maps derived from MRSI in their effectiveness of segmenting brain tumors and edema.

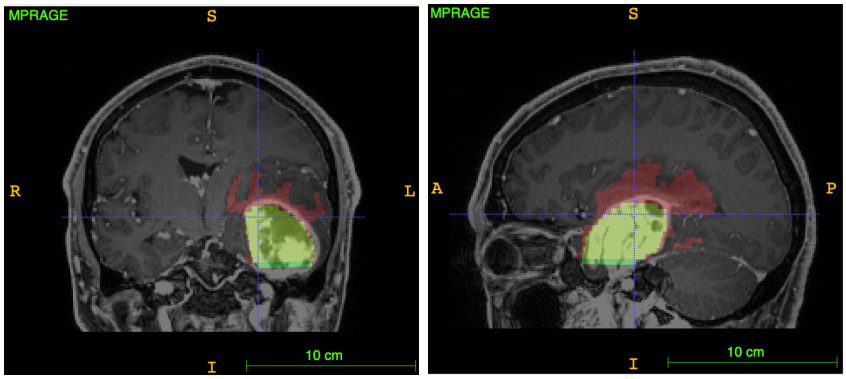

Figure 1.1: Left: a coronal view of a 3D MRI scan of a human brain with glioblastomas. MPRAGE isa type of MRI sequence that emphasizes the contrast between gray and white tissues. Right: sagittal view.

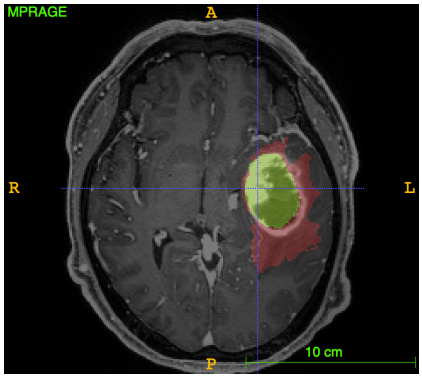

Figure 1.2: Horizontal view. Red is edema. Green is tumor.

1.3 Summary of Contributions

This thesis provides an overview of brain tumor segmentation, emphasizing machine learning and MRSI. Then, an experiment with eight glioblastoma scans demonstrates the effectiveness of MRSI in identifying brain tumor and edema regions, featuring an automated machine learning pipeline nnU-Net. A novel MRSI technique, SPICE, obtains all metabolite maps, which generates an ultra-high spatiotemporal resolution dataset in a short 7-min scan. This study is unique in analyzing brain tumor sub-regions with this level of detail and richness of information.

1.4 Organization of the Thesis

Chapter 2 presents an introduction to background material and explains the motivation for this study. This section focuses on a literature review of brain tumor segmentation based on MR structural images. It first introduces brain tumors and the important role of automated segmentation techniques. It then presents different paradigms in current segmentation techniques with their pros and cons.

Chapter 3 provides more content on the problem of interest and the proposed solution. It first identifies the integration of the MRSI technique as one of the exciting directions in determining the extent of brain tumors. Then, it elaborates on the dataset, mask generation, scan registration, and the deep convolutional neural network model of choice. It also reviews works on using MR spectroscopy and spectroscopic imaging for brain tumors and draws connections between this thesis and other works.

Chapter 4 presents the experiment results. It evaluates the segmentation quality and discusses the potential for further improvement.

Chapter 5 summarizes the results and provides insight into future MRSI studies for brain tumor sub-region segmentation.

CHAPTER 2

BRAIN TUMOR SEGMENTATION BASED ON MR STRUCTURAL IMAGES

This chapter starts by introducing the significance of automatic brain tumor segmentation on MRI images. MRI stands for magnetic resonance imaging, and it is a non-invasive medical imaging technique that can form 2D and 3D representations of internal human bodies. The visualization helps understand body development and various diseases in research and clinical settings. Next, this chapter reviews different paradigms of solutions, including manual labeling, semi-automatic and fully automated methods, unsupervised and supervised learning.

2.1 MRI Brain Tumor Segmentation

Brain and other nervous system tumors are among the leading causes of death in the United States. Cancer causes the second most deaths in 2019, while the brain and other nervous system tumors cause the most cancer deaths for men aged <40 years and women aged <20 years [3, 4]. For this devastating disease, MR imaging remains the gold standard for neuroimaging due to noninvasiveness, excellent soft-tissue contrast, and versatility in sequence design to identify different key components of tumor physiology [5]. Take the three most common brain tumors as examples, the central nervous system metastases, meningiomas, and glioblastomas; all require MR imaging in the early detection, monitoring, and diagnosis. MR imaging of the brain with and without contrast is the gold standard for diagnosing brain metastases, which cover most diagnosed cases of brain tumors. The same technique helps to diagnose meningiomas, which comprise 35% of all primary brain tumors [6], with additional biopsy or resection sometimes required. It is also the first choice in visualizing glioblastomas, which account for 45% of malignant primary brain tumors [6], while the final diagnosis requires pathological

confirmation by taking a biopsy from the patient [5]. Structural scans are essential in brain tumor treatment, so are diffusion tensor imaging (DTI) tractography and MRSI. Preoperative DTI tractography can help neuronavigation, and MRSI can guide targeted biopsy in heterogeneous tumors.

Treatment of brain tumors requires a balance between minimizing the risk of perioperative morbidity and maximizing the extent of tumor resection. The most common practice is to perform surgical removal. Other therapies include radiation therapy, radiosurgery, chemotherapy, targeted drug surgery, and others. Knowing the extent of tumors is critical in therapy planning. Automatic segmentation also facilitates prognosis. Building an automated pipeline robust to pseudo-response and pseudo-progression is a crucial task.

Although humans can distinguish between healthy brain regions, tumors, and edema, automatic brain segmentation remains hard. First of all, the boundary of low-grade gliomas is blurred or nearly indistinguishable. Low-grade tumors are those with grades 1 and 2 as defined in table 2.1. Secondly, HGGs usually have ragged boundaries. Some tumors are easier to segment, like meningiomas, which have a smooth border and are space-occupying. Malignant gliomas are much harder. They infiltrate normal brain tissue and often cause edema, making the edge less distinguishable and uneven. Thirdly, automatic tumor segmentation requires different modalities, but they are usually in different orientations and require registration. Fourthly, obtaining high-resolution multi-modalities scans requires a long scan time. Lastly, the automatic tumor segmentation methods need to be robust enough to generalize well because MRI scanners and sequence choices (field-of-view, voxel resolution, gradient strength, field strength, etc.) can differ and cause different contrast.

Table 2.1: Tumor grade.

Grade	Description
x	grade isn't known
1	well differentiated, low grade
2	moderately differentiated, intermediate grade
3	poorly differentiated, high grade
4	undifferentiated, high grade

In the case of segmenting gliomas, there are even more challenges. Gliomas might be confused with a stroke. Also, it can take place at every location

of the brain, with very different shapes and sizes. Plus, gliomas change the surrounding brain tissue instead of moving them, blurring the boundaries.

MRI has a long history of being used for brain tumor segmentation. Tremendous progress has been made, and there are many review papers available considering different aspects of the process. Some review papers are focusing on the broad field [7, 8, 9, 10], and some on more specialized sub-areas like denoising [11], data argumentation [12], and others on the use of deep learning techniques [13]. Numerous methods are available to choose from, and open-source labeled data sets are available for researchers to benchmark. The most used data set is Brain Tumor Segmentation (BraTS) challenge. Many techniques mentioned in the thesis are past participants in this famous challenge.

Table 2.2: Tissue Intensity [14, 9].

Modality	Edema	Tumor	Region
FLAIR	Hyper-intense	Hyper-intense	non-enhanced tumor
T1	Hyper-intense	Hyper-intense	
T1C	Iso/Hyper-intense	Hyper-intense	tumor boundary
T2	Hyper-intense	Hyper-intense	non-enhanced tumor

Table 2.3: Tumor subregion and corresponding contrast in different modalities.

Enhancing tumor	T1c-w	hyperintense
Edema	T2w, T1w	Hyperintense
	FLAIR	Edema is dark, CSF is bright
Necrotic	T1c-w	Not enhance
	T2w	Hyperintense
	T1w	Hypointense
Non-enhancing tumor	T1c-w	Not enhance
	T2w	Lower intensity than necrotic
	T1w	Hypointense

The Medical Image Computing and Computer Assisted Intervention Society (MICCAI) has been holding BraTS challenges since 2012. MICCAI will release a labeled training data set for researchers to develop and train their machine learning-based classifiers every year. The focus of BraTS challenges is to foster better segmentation of sub-regions of gliomas in multi-modality

MR scans. These scans involve T1, post-contrast T1-weighted (T1Gd), T2-weighted, and T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) volumes. The early challenges involve synthetic data sets and semi-automated labels, but the recent data sets are all manually labeled or approved by experienced neuro-radiologists. Four labels are available: GD-enhancing tumor, the peritumoral edematous or invaded tissue, the necrotic tumor core, and others. Such open-source databases have given rise to the development of deep data-driven approaches.

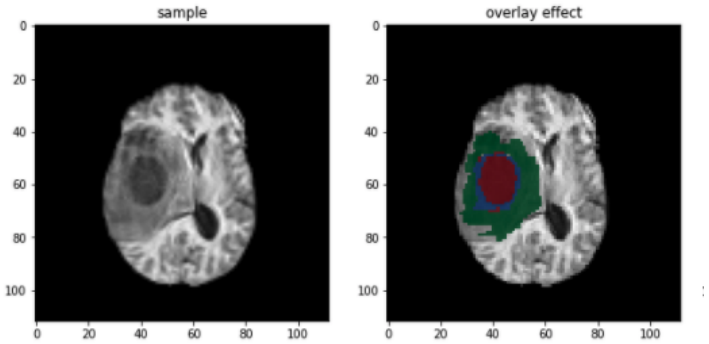

Figure 2.1: An example of a BraTs data. Whole tumor: a union of green, red, and blue. Tumor core: a union of blue and red. Enhanced tumor core: red.

2.2 Manual Annotation

Manual segmentation involves manual delineation of tissue boundaries and close examination of every slice. This process is time-consuming and error-prone even with the help of a specialized and thoughtfully designed graphical user interface. A trained expert has to go through the region of interest slice by slice, and the view is limited. Given the increasingly mature technique nowadays, more than eight 512x512 images might need to be examined and labeled. The process is also error-prone because of potential artifacts and noise. Quantitative evaluations also reveal high disagreement between human raters with the DICE score range between 74% and 85% [15]. Labeling slice by slice in one direction can produce jaggy labels judging from a different direction. The inter and intraoperator variability is also another issue in brain tumor segmentation [16] and other tissues [17, 18]. However, the manual segmentation from experts remains the ground truth in most cases.

2.3 Semi-automatic Methods

On the other hand, semi-automatic segmentation methods can alleviate the pressure on human annotators. This type of method does not require training examples. Still, it needs some human-labeled prior information about the segmentation task, i.e., a few example labels, region of interest (ROI), initialization seed, etc.

MRI Brain tumor segmentation is hard to tackle without training data because intensity information alone is unreliable. For instance, in T1-weighted image, tumor, gray matter, CSF might share very similar voxel intensities [19]. To make things worse, due to the heterogeneous nature of some cancerous tumors, the voxels inside the tumor region might have very different intensities. Therefore, simple thresholding cannot perform well. Semi-automatic methods involve human intervention to ensure good results.

Lim et al. [14] classify semi-automatic segmentation methods as region-based, clustering-based, and a mixture of both. The region-based methods consider the spatial characteristics of images, and therefore usually produce connected areas. Some example algorithms include random walker, watershed, region growth. The clustering-based methods cluster pixels according to their feature space values. *K nearest neighbors* (KNN), *Gaussian mixture model* (GMM) are examples of this type. Note that K nearest neighbor can also be supervised methods, depending on how the labeled data are obtained. GMM is an automatic method, but it is often combined with other components requiring human input. In the same paper, a mixture of both approaches is used. The proposed segmentation scheme comprises three parts: information modeling, information fusion, and visual object extraction. A region-based approach, random walk modified by adding homogeneity- and object feature-based components, is the main workhorse of the first part. In the second part, the weighted averaging method fuses information from different modalities. Finally, information-theoretic rough sets determine the boundary in transition areas.

As an example of requiring tumor seed points, Zhang et al. devised a two-step method to semi-automatically segment the brain tumor region [19], which they denoted as the multi-scale Otsu-based segmentation. The first step is automatic. The procedure starts with an edge-aware filter that smooths the MRI image. Then, multi-scale Otsu-based thresholding seg-

ments both the original and the smoothed images. K-nearest neighbor is used to fusing the two results. The second step requires human intervention. Human-labeled tumor seed points needed to be ready before the bi-directional region growing algorithm could work. Fig. 2.2 shows the overall workflow.

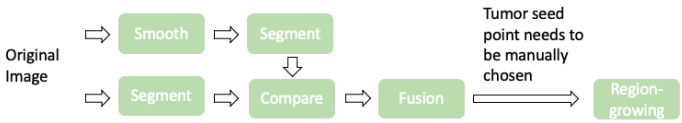

Figure 2.2: The general workflow of multi-scale Otsu based segmentation.

Another example of requiring operators' initial seed is the Snake model. Yezzi et al. proposed a Snake model for segmentation of medical imagery [20] in 1997. This method uses a hand-crafted feature and requires the user to provide initial seeds to start optimizing. Snake is widely accepted but usually serves as the first step of labeling. While making accurate classification for most voxels, it tends to fail at the boundary and is sensitive to noise, and therefore, manual adjustments follow.

Zhao et al. used the medical image analog to optical flow, structural trajectory, to generalize one annotated slide to the 3D volume. Their method requires human annotation on an automatically selected slice. The proposed method determines the first slice by evaluating the degree of asymmetry. Because consecutive areas should share similar label distribution, labels propagate through each slice from the nearest to furthest. Optimal flow estimation helps to find the structural correspondence between neighboring pieces. Markov random field optimizes each prediction with hard constraints on those labeled pixels identified by the optimal flow estimation.

Zhu et al. [21] built a graphic user interface for semi-automatically segmenting brain tumors using existing open-source software. The first stage makes use of voxel-based segmentation provided by the Functional MRI of the Brain (FMRIB) Software Library (FSL) and the FMRIB's Automated Segmentation Tool (FAST) [22, 23]. The next stage involves a level set-based segmentation that uses ITK-SNAP [24]. The general workflow looks like Fig. 2.3.

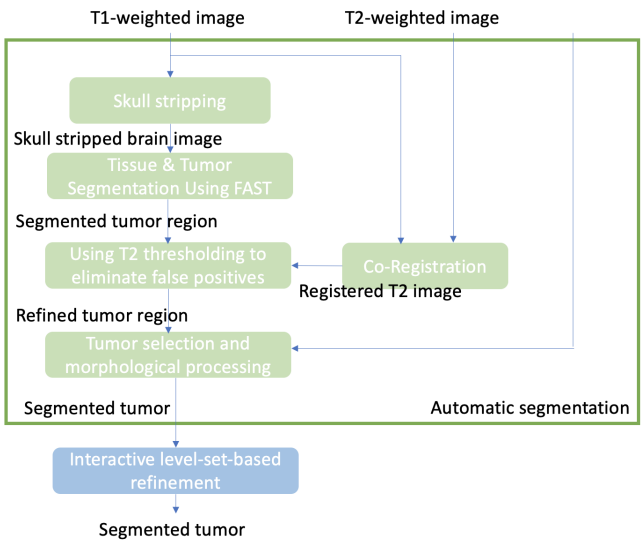

Figure 2.3: AFINITI pipeline: a graphic user interface design for semi-automatic brain tumor segmentation by utilizing existing softwares like FAST, FSL, and ITK-SNAP.

2.4 Fully Automated Methods

Fully automated methods refer to methods that require no human intervention. While some of them require data prior, this section will focus on strategies without training data. Those methods require a training step will be further divided their dependency on labels. Section 2.5 discusses automatic techniques with unlabeled training data, and section 2.6 reviews data-driven solutions that require labels.

Early work [25] uses spatial probabilistic brain atlas as prior information. This approach is prevalent in healthy brain tissue segmentation. The probabilistic atlas is often an average of many brain tissue masks from different people. The atlas serves two purposes. Not only it provides spatial prior probabilities, but it also helps estimate the intensity distribution of each healthy tissue class. Tumor prior probability is calculated using T1 post-contrast and preconstrast images. Because edema appears most often in the white matter region, the spatial prior for edema is 20% of the voxel's white matter probability. These prior probabilities for all normal and abnormal tissues become inputs to the expectation-maximization (EM) segmentation. In this work of Prastawa et al., the EM algorithm iteratively optimizes the

following tasks:

- classify voxel using the current tissue distribution and bias field estimates
- updates the bias field estimate using the current prediction classes
- re-estimate the probability distribution

Unlike the atlas-based approach, another approach in automating the segmentation process requires multiple modalities. Diaz et al. [10] propose ABTS, a four-staged automatic brain tumor segmentation tool using automatic histogram multi-thresholding and morphological operations. This method requires four types of modalities: T1-weighted spin echo, T1-weighted spin echo with gadolinium contrast agent, T2-weighted spin echo, and Fluid Attenuated Inversion Recovery. Having multiple scans with different contrast is extremely helpful because it avoids making false-positive predictions to the tissue boundaries due to the poor contrast. A voxel intensity prior is used to help with the thresholding method. The author also used a Savitzky-Golay FIR filter to separate background, skull, and brain parenchyma. The contrast difference in different modalities helps separate edema because it shows a low-intensity value in T1 and T1C and a high-intensity value in T2 and FLAIR. Gadolinium-enhanced lesions like tumors appear with low intensity in T1 and high intensity in T1C and T2.

2.5 Unsupervised Methods

Unsupervised methods learn from untagged data, and clustering is one of the most popular categories. This section will cover both k-means clustering and fuzzy c-means (FCM) clustering. Afterward, a brief review of recent works is presented.

2.5.1 K-means clustering

Lloyd's algorithm for k-means clustering is an iterative clustering method that guarantees convergence. The user needs to define the number of clusters, K, and a distance metric. The algorithm initializes by first picking K random

points as cluster centers. The iteration starts with assigning data points $x^(i)$ to the closest cluster center. The most intuitive and widely adopted distance metric is the L-2 distance. Then, choose the centroids of each cluster as the new cluster centers. These two steps keep iterating until no point assignment changes. To formally describe this process, we say k-means clustering optimizes the following cost function:

$$\min_{\mu} \min_{r} \sum_{i \in \mathbf{D}} \sum_{k=1}^{K} \frac{1}{2} r_{ik} ||x^{(i)} - \mu_k||_2^2 \text{ s.t. } \begin{cases} r_{ik} \in 0, 1 \ \forall i, k \\ \sum_{k=1}^{K} r_{ik} = 1 \ \forall i \end{cases} \tag{2.1}$$

Lloyd's algorithm for k-means guarantees to find a local optimum in a finite number of iterations. The downside of this algorithm is its sensitivity to initialization. K-means++ [26] specifies a way to initialize the k-means and makes it more robust. Researchers also use kernel tricks to make k-means clustering more expressive. K-means is a prevalent building block of the brain tumor segmentation pipeline.

Sasibhusana Rao et al. [27] compares both k-means and FCM methods in their performance on brain tumor segmentation. According to their comparative study, FCM has a better performance in terms of mean-squared error, peak signal-to-noise ratio, and processing time.

2.5.2 Fuzzy C-means Clustering

Fuzzy c-means clustering [28, 29] is a soft version of k-means clustering. In FCM, each data has a non-hard assignment to a class. A point can have a positive coefficient of being in one cluster and another positive coefficient of being in another. Therefore, it makes more intuitive sense in the case of MRI brain tumor segmentation. The partial volume effect describes the loss of contrast when a voxel contains multiple tissue types. We can express this idea mathematically by having one data point belonging to multiple classes. To perform FCM clustering on a new dataset, the user needs to define the number of clusters, K, a hyper-parameter that controls the fuzziness, m, and a distance metric to start. Again, some popular choices for the distance metrics include Euclidean distance, cosine distance, and kernel-based distances. The first step is to randomly assign coefficients to each data point for being in the K clusters. Then, the algorithm computes the centroid of each cluster.

Next, the fuzzy membership of each point is re-calculated as

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} (\frac{d_{ij} d_{ik}}{)}^{\frac{2}{m-1}}} \tag{2.2}$$

where d_{ij} is the distance between current data point x_i and the centroid v_j, m is the fuzziness index with $m \in [1, \infty]$. Iterate these two steps until convergence.

2.5.3 Recent Works

Sauwen et al. [30] have compared the performance of several unsupervised classification methods for HGG segmentation using multiple modalities. The two main classes of choice are Non-negative matrix factorization (NMF) and clustering. Non-negative matrix factorization finds a low-rank matrix approximation of the original data, such that this approximation is a product of two non-negative matrices. K-means clustering then converts the decomposition matrices to hard segmentation. In this paper, hierarchical alternating least-squares NMF, convex NMF, and hierarchical NMF are the candidates for comparison. The clustering methods of their choice are FCM and the GMM. GMM is popular because the tissue intensity distribution closely follows the Gaussian distribution. Although in practice, the number of Gaussian distributions is usually larger than the number of tissues to achieve better performance. The last candidate is spectral clustering, a graph-based approach. The author uses two private datasets and finds that hNMF performs the best on the tumor, edema, and core region, while spectral clustering segments necrosis the best.

Just like in semi-automatic brain tumor segmentation, atlas prior is essential to the success of many automatic methods. Many research centers have made public their versions of the brain atlas. However, capturing the distribution of tumors is often impossible because of the different extents, shapes, and locations among different instances. Researchers managed to overcome this missing tumor prior issue with creative solutions. Prastawa et al. [25] perform segmentation using an expectation-maximization approach with spatial prior derived from the ICBM digital brain atlas.

2.6 Supervised Methods

Supervised methods refer to computational methods with parameters tuned by labeled training data. Recent years have witnessed the success of such computational models in brain segmentation. Different types of models have contributed to the field, including Artificial Neural Network (ANN), random forest (RF), support vector machine (SVM), k-nearest neighbor, convolutional neural network (CNN), etc. Deep neural networks started to gain popularity due to the increasing computation power and availability of large labeled datasets like BraTS. BraTS have just begun in the recent decade, and it shows a shift in attention from feature engineering to network topology engineering. Based on a study carried out by Ghaffari et al. [31], the CNN-based approach only appeared in BraTS after 2014. In 2012 and 2013, most participants were using the random forest classifier-based approach. Since 2015, most of the submissions have been using CNN. U-Net was proposed in 2015 and started to strive in BraTS from 2016 and onwards. Forty-two of more than 50 papers are variants of U-net or DeepMedic in BraTS 2017, and more than half of the submissions are U-net based in the 2018 competition.

2.6.1 Random Forest

Random forest, as the name suggests, comprises many decision trees. A decision tree is a tree structure where each inference starts at the root and follows the decision, and the leaf node reveals the prediction. To construct such a tree, there are multiple specific learning algorithms, including iterative dichotomiser 3 (ID3) [32], C4.5 [33], and CART (classification and regression trees) [34]. The general flow is to choose a variable that best splits the data items according to a chosen criteria at a node. Then split the data according to the chosen rule, append two nodes, and repeat the first process. Stop when a node has a reasonably small amount of data and compute the statistics of this leaf node. Some criteria commonly used are Gini impurity, entropy, classification error, and information gain.

Random forest [35] is a parallel ensemble of multiple decision trees. Each input vector is fed to each tree when computing an inference from a random tree implementation. The class with the most votes becomes the final prediction. During training, each tree is grown by randomly sampling N cases

with replacements from the training set, where N is the number of cases. A much smaller set of features is allowed at each node. This number is held constant across nodes and trees as a hyper-parameter. Each tree grows with no constraints and no pruning. The performance of random forests depends on the correlation among trees and the strength of each tree.

According to [31], the best performing models of BraTS 2012 are using random forest. Four out of the ten participants in BraTS 2013 are RF-based, and three out of four RF-based approaches ranked top. RF remains popular and keeps appearing in BraTS competition, although deep convolutional neural networks now achieve the best performances.

2.6.2 Neural Network

A general workflow of many machine learning-based approaches includes data preprocessing, choosing the desired network architecture, choosing a loss function and optimizing it, and running inference and post-processing the prediction. This section reviews some essential training techniques and methodologies like back-propagation and stochastic gradient descent. This section is by no means a comprehensive review; readers are redirected to more readings here [36, 37, 38, 39].

A neural network is a composite function with vector input and output. In the case of brain tumor segmentation, the input is often some array representations of brain MR images with different acquisition schemes. The output is obtained by the *forward propagation*, which describes applying the composite function defined by the network architecture on the input array. The ideal output should be the same as the target image, also called ground truth images, usually multi-channel, with each channel representing a sub-region.

Back-propagation is used to find better parameters at each iteration. The goal of tuning parameters is to reduce the loss defined by the cost function using the current data batch. Let us consider the parameter's current value as a point in the parameter space. In each iteration, we calculate the gradient at the current location. Gradients always point in the greatest ascend. We want to take a small step in the greatest descent direction to find the minimum. Adding the product of gradient and a small negative value called *step size*, also known as the *learning rate*, can direct us to a local minimum. There

are many schemes in determining the learning rate. *Back-propagation* refers to an implementation of the gradient finding process. Of course, there are many ways of finding minima in a continuous space. The implementation of those optimization methods is called *optimizers*. But the back-propagation and stochastic gradient descent are widely chosen for their empirically good performance. They are more likely to find the global minimum and take less computation time and resources to find it.

A deep neural network has many layers, and therefore, a large number of parameters. As a result, the model is nearly always data-hungry. Medical images are usually scarce due to poor accessibility to the general population, time-consuming labeling process, and privacy concerns. Therefore, researchers have developed ways to expand the number of training data from the limited labeled scans, like rotation, flipping, and more complicated operation. This process is called *data augmentation*. Another vital trick to overcome the data scarcity problem is to use the training data multiple times. Every time the whole dataset is trained on the network is called an *epoch*. If the model has seen the entire training dataset twice, the epoch number is two.

Unlike other machine learning tasks, brain tumor segmentation is unique in its colossal data size. With increasingly mature techniques, the MR scanner can produce 1 mm^3 or even finer resolution. The large size of input data and the deep structure of neural networks lead to a significant amount of calculation. The computational device has to have a considerable memory such that each layer can hold the calculated floating-point values. As a workaround, researchers use patches instead of the whole image. Image patch size, batch size, and network size are critical trade-offs in experiment design. A bigger size indicates richer spatial features possible to be learned, but the training might take much longer, and more training data might be required. In nowadays, machine learning society, *stochastic gradient descent* is the most popular training scheme. In each iteration, only a batch of training data is used to tune the parameters instead of using the whole training corpus at a time. By having less training data at a time, we allow more complex models.

Different combinations of them have been explored extensively throughout the years. Such representation learning methods usually learn the input distribution quite well but have difficulties generalizing, i.e., they are often limited to the image modalities of the training set.

To dive into the building components of neural network models, let us start with *fully connected layer*. A fully connected layer can be seen as a matrix multiplication operation. The user needs to decide the input size and output size of a fully connected layer, and the number of parameters will be $(n+1)*m$. Let us denote x as the input vector with size $1*n$, the weight matrix will become $(n+1)*m$. The output derives from $\hat{x}w = y$ where

$$\begin{bmatrix} 1 \\ x \end{bmatrix} = \hat{x} \tag{2.3}$$

and the output is y with size $m*1$. The fully connected layer was used heavily in artificial neural networks before the rise of convolutional neural networks [40]. The fully connected layer can easily change sizes at different layers, capturing global features. However, it does not scale well if we want deeper and wider networks. To deal with the scaling issue, the researchers have to make use of the convolution operation. A *convolutional layer* has a trainable parameter w, called filters, and bias b. The user needs to define the width, height, depth, and number of filters. An operation that plays an opposite role is *transposed convolution*, where the input matrix is expanded using convolution. A detailed discussion can be found at [41]

The *maximum pooling* does not have a training parameter. To define a maximum pooling layer, the researcher needs to determine the size of the receptive field and the step size. The maximum pooling layer reduces the size from layer to layer while preserving essential signals.

Until here, all layers described above produces linear transformation of the input vector. To add more nonlinearity, researchers have come up with *activation functions* like ReLU and tanh. A *ReLU* is an element-wise operation performed on the input array. The output of each value is

$$y = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{otherwise} \end{cases} \tag{2.4}$$

There is also a modification of ReLU called leaky ReLU [42], which is

$$y = \begin{cases} ax, & \text{if } x < 0 \\ x, & \text{otherwise} \end{cases} \tag{2.5}$$

where a is a small constant.

Softmax is another way of introducing non-linearities, often used in the last layer of model in a classification task. It takes in a vector and produces another vector. The input is often the scores for each class, and the output is the class possibilities.

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \text{ for } i = 1, ..., K \text{ and } \mathbf{z} = (z_1, ..., z_K) \in \mathbb{R}^K \tag{2.6}$$

Batch normalization [43], instance normalization [44], and layer normalization [45] are ways to solve covariate shift and to mitigate training instability. All normalization methods require shifting and scaling the input signals by mean and standard deviation respectively.

$$\hat{x} = \frac{x - \mu}{\sigma} \tag{2.7}$$

Batch normalization calculates the two statistics per channel. In each channel, it uses all samples and all spatial dimensions. Instance normalization calculates per channel per sample, with all spatial dimensions included in the calculation. Layer normalization calculates per layer. It uses all individual samples across all channels and all spatial dimensions.

There are many loss functions used in the machine learning community. And the design of the loss function is task-dependent. In brain tumor segmentation, people use *cross-entropy* and *DICE loss*. The cross-entropy between two probability distributions measures the average number of bits needed to identify an event in information theory. Its calculation is

$$H(p, q) = -\sum_{x \in \chi} p(x) \log q(x) \tag{2.8}$$

where p and q are two distributions, and χ is its event space. In the context of machine learning, p and q are vectors, and χ is each dimension.

Dice loss, as explained in section 3.5 is calculated as

$$S = \frac{2|X + Y|}{|X| + |Y|} \tag{2.9}$$

and Dice loss is

$$\text{DiceLoss} = 1 - \frac{2|X + Y|}{|X| + |Y|} \tag{2.10}$$

where

$$|X \cap Y| = \begin{bmatrix} a & b \\ c & d \end{bmatrix} * \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ c & d \end{bmatrix} \tag{2.11}$$

CHAPTER 3

IMPROVED BRAIN TUMOR SEGMENTATION WITH METABOLIC INFORMATION

3.1 Magnetic Resonance Spectroscopic Imaging

Magnetic Resonance Spectroscopic Imaging is the spectroscopic variant of Magnetic Resonance Imaging. It produces spatially localized spectra and delivers metabolic information about voxels in the region of interest. Each voxel contains a spectrum with multiple peaks due to the chemical shift effect. Chemical shift refers to the resonant frequency of the hydrogen proton in most MRI sequences. The surrounding chemical environments determine it. To provide some more quantitative understanding, let us first recall Larmor frequency

$$\omega_L = \gamma B_0 \tag{3.1}$$

where ω_L is the Larmor frequency, which describes the precessional frequency of a proton spin. γ is the gyromagnetic ratio, and B_0 is the background magnetic field. However, the shielding effect causes each proton to experience a slightly different background magnetic field. The shielding effect refers to the reduction in effective nuclear charge due to the attraction between electrons and nuclei. The effective magnetic field is

$$B_{\text{eff}} = (1 - \sigma)B_0 \tag{3.2}$$

where σ is the shielding constant. Because every proton group experiences a different field, they resonate at different frequencies. This chemical phenomenon makes resonant frequencies of proton groups shift away from one common frequency. Thus, the name, chemical shift.

Let us take one subject as an example. Visualization of some spectrum shows the distribution of different resonant frequencies. Because spectra are

complex, we will draw both the magnitude and the real part. The average signals of pure edema, tumor, contralateral normal, and tissue surrounding the affected regions are attached below.

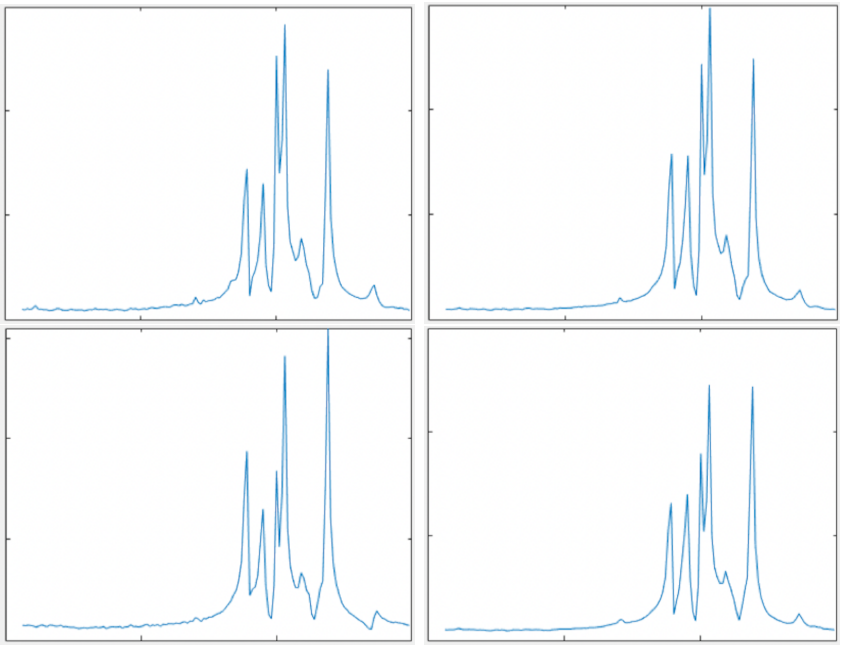

Figure 3.1: The average of spectrum magnitude with water and lipid signal removal. From top to bottom, and left to right: tumor, pure edema, contralateral normal, and neighbor voxels.

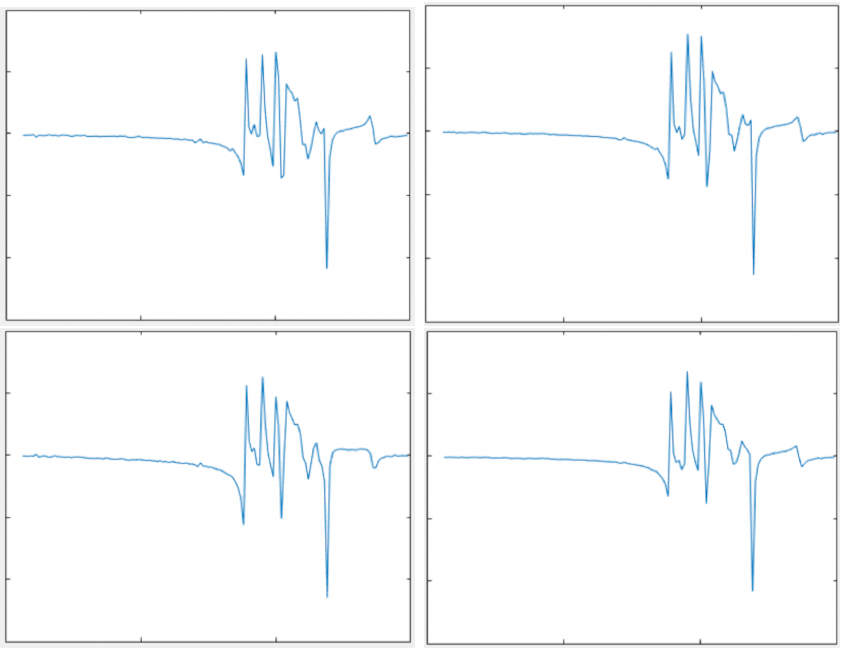

Figure 3.2: The average of spectrum real part with water and lipid signal removal. From top to bottom, and left to right: tumor, pure edema, contralateral normal, and neighbor voxels.

Fig. 3.3 and 3.5 shows some example spectra. The horizontal axis unit is ppm, which stands for parts per million, and it is measured relative to a reference compound. Tetramethylsilane is the most common reference compound. The equation to calculate the chemical shift in ppm is

$$\frac{\upsilon - \upsilon_{ref}}{\upsilon_{ref}} * 10^6 \tag{3.3}$$

where υ is the Larmor frequency. This representation is popular because of its field strength independence. For example, no matter the field strength an MRI scanner provides, the water peak is always at around 4.7 ppm.

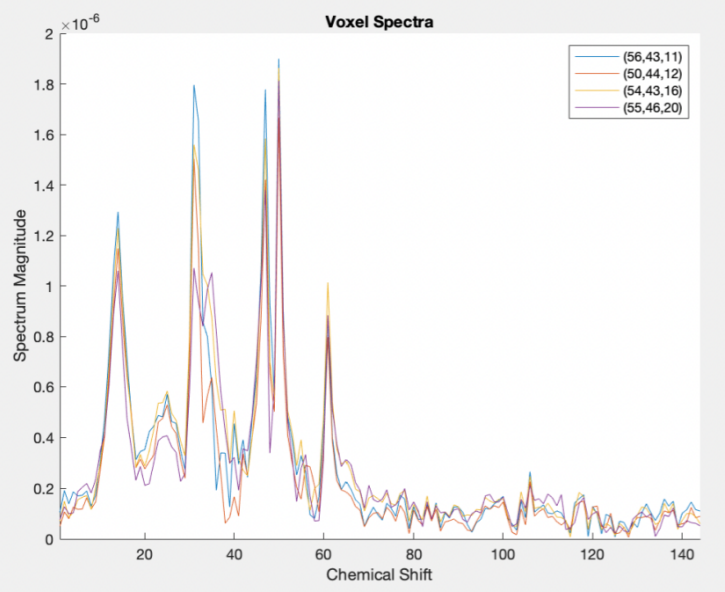

Figure 3.3: Randomly chosen edema spectrum.

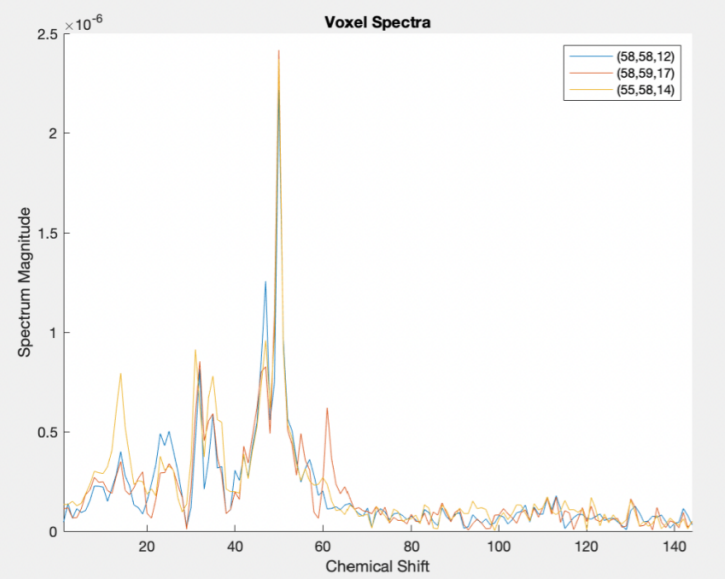

Figure 3.4: Randomly chosen tumor spectrum.

Because resonance frequency depends on the chemical environment, we can identify and measure different chemical compounds. The area under a peak has a linear relationship with the absorption intensity. Therefore it is proportional to the nucleus concentration. This relationship is valid for comparisons between molecules. Also, because MRSI provides spatial coordinates of the signals, we can map metabolites. With the help of these

two properties, we can obtain spatial distributions of various metabolites. MRSI becomes promising because metabolic changes can reflect and even may proceed anatomic changes.

MRSI and MRS are methods to understand tumors because they reflect the metabolic profile of molecules. As early as 1996, researchers have found MRSI powerful in diagnosing brain tumor type [46]. Also, MRSI is non-invasive in identifying the heterogeneity structure of brain tumors to guide biopsy planning, such that clinicians can have a better chance of knowing the actual grade of the tumor [47]. MR spectroscopy can also help identify IDH-mutated gliomas [48]. IDH stands for isocitrate dehydrogenase, and it is an enzyme that moderates the rate of the Krebs cycle and is part of the energy metabolism. Patients with IDH wild type often carry a worse prognosis than those with IDH mutant. Also, another MR spectroscopy-detectable metabolite, glycine, is proposed to be a marker of glioma aggressiveness. Although still under investigation, this could be another metabolite potent in brain tumor research and clinical use cases [49].

3.2 Useful MRSI-Detectable Metabolites

MR spectroscopy and spectroscopic imaging are common tools in studying brain tumors because they give us a glimpse into the neurochemical state of the brain. They can help to diagnose and differentiate various brain diseases. As introduced in this section, researchers have found several critical MRSI-detectable metabolites that respond to malignant tissue proliferation.

3.2.1 NAA

N-Acetylaspartic acid, or N-acetylaspartate or NAA, is a derivative of aspartic acid. The chemical formula of NAA is $C_6H_9NO_5$. The molecular graph is

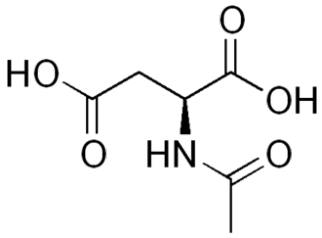

Figure 3.5: NAA molecular graph. Figure adapted from Wikipedia.

NAA presents in neurons, oligodendrocytes, and myelin in the adult brain, and it is a marker of neuronal density and viability. Some most widely-accepted primary functions of NAA include precursing neurotransmitter N-Acetylaspartylglutamic acid (NAAG), participating in myelin lipids formation, and regulating osmosis. A decrease in NAA indicates a loss or injury of neurons when tumors replace neurons. On the MRSI spectrum, NAA's peak is at around 2 ppm.

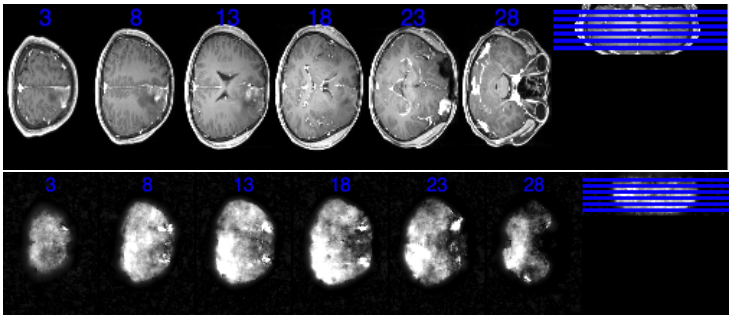

Figure 3.6: Up: MPRAGE of a patient. Down: the NAA distribution of this patient.

3.2.2 Choline

Choline refers to several soluble components of brain myelin and indicates fluid-cell membranes and integrity. Choline is an essential nutrient for brain development, and it is part of the S-adenosylmethionine synthesis, which regulates gene expression and changes brain function. Pathological alterations to choline appear in membrane turnover, for example, tumors. Choline's peak is at 3.2 ppm.

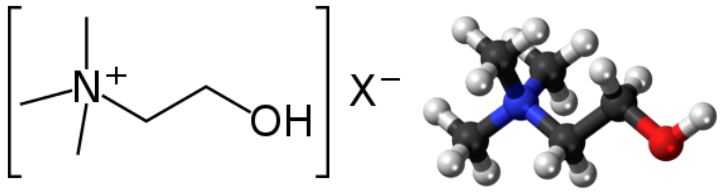

Figure 3.7: Choline molecular graph. Figure adapted from Wikipedia.

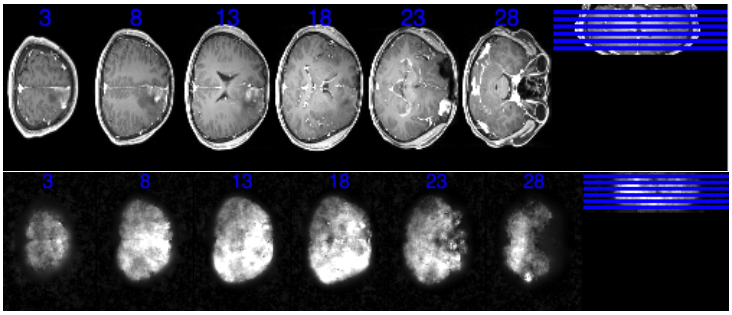

Figure 3.8: Up: MPRAGE of a patient. Down: the choline distribution of this patient.

3.2.3 Creatine

MRSI can detect total creatine, including creatine (Cr) and phosphocreatine (PCr). Both compounds appear to have peaks at 3 ppm and 3.9 ppm in MRSI. Total creatine plays an integral role in energy metabolism. They serve as a reservoir for adenosine triphosphate (ATP). PCr donates phosphate groups to adenosine diphosphate (ADP) such that it converts back to ATP. Total creatine usually stays constant in most diseases and with age, with one known exception of astrocytomas where it diminishes. Therefore, it is considered an excellent internal standard.

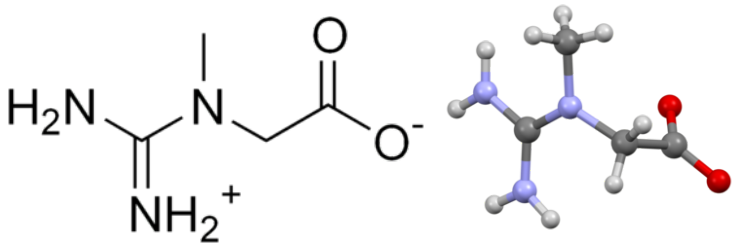

Figure 3.9: Creatine molecular graph. Creatine has the nominal formula $(H_2N)(HN)CN(CH_3)CH_2CO_2H$. Figure adapted from Wikipedia.

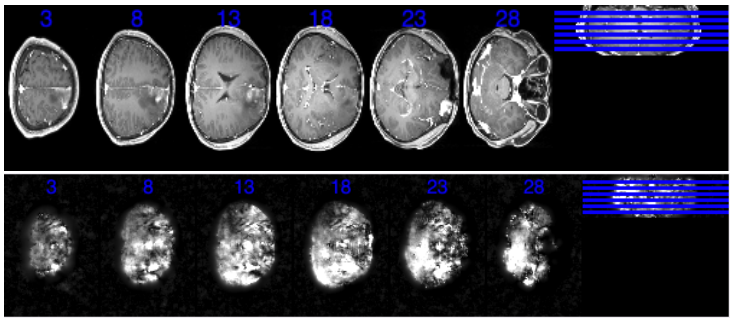

Figure 3.10: Up: MPRAGE of a patient. Down: the creatine distribution of this patient.

3.2.4 Lipid

Lipid can be produced in tissue breakdown processes, which stores energy, signals as in the lipid signaling process, and serves as structural components of cell membranes. The Lipid MAPS consortium has eight categories: fatty acids, glycerolipids, glycerophospholipids, sphingolipids, sterols, prenols, saccharolipids, and polyketides. An abnormal presence of lipid indicates brain tumor necrosis and myelin destruction, with possible contamination by subcutaneous fat from the skull. The peak can appear between 0.8 to 1.5 ppm and 2 ppm.

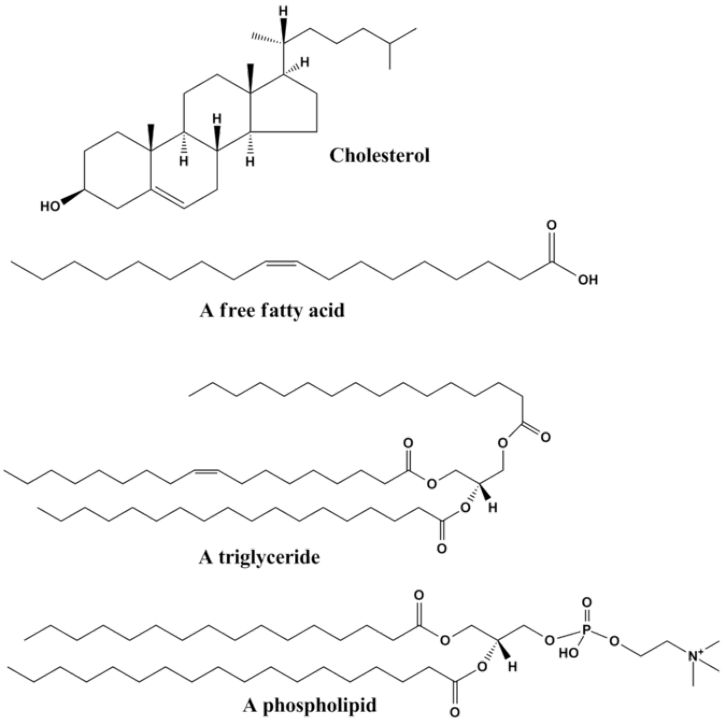

Figure 3.11: Molecule graphs of some common lipids. Figure adapted from Wikipedia.

3.2.5 myo-Inositol

myo-Inositol (MI) is a carbocyclic sugar, which mediates cell signal transduction and acts in osmoregulation. It is considered an inflammatory marker. MI appears to be low in high-grade tumors and high in low-grade tumors. Its peak is at 3.5-3.6 ppm.

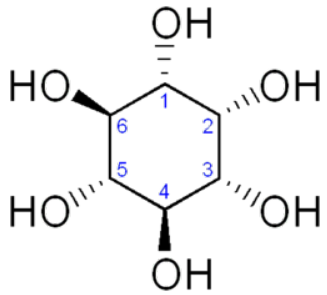

Figure 3.12: Molecule graph of myo-Inositol. Figure adapted from Wikipedia.

3.2.6 Lactate

Lactate (Lac) arises from anaerobic metabolism. It is the conjugate base of lactic acid, with the molecular formula $CH_3CH(OH)CO_2^-$. Lactate is proposed in a few reports to be the preferential energy source by human brain neurons. According to the lactate-shuttle hypothesis, glial cells transform glucose into lactate and provide it to the neurons. Lactate shows up in tumors containing zones of necrosis. On MRSI, it has doublet at 1.3 ppm.

3.3 Study Participants and Data Acquisition Procedures

A retrospective analysis of data from eight patients with histopathologic diagnosis of glioblastomas was performed, as approved by the Institutional Review Board of the Fifth People's Hospital of Shanghai, China.

The MRSI data were obtained using the SPICE sequence. More details about SPICE will be presented in this chapter later. As part of the experimental protocol, brain structural images were obtained using 3D contrast-enhanced MPRAGE (1.0 x 1.0 x 1.0 mm^3, FOV = 256 mm, TR/TE = 2400/2.13 ms). FLAIR, T1-weighted, and T2-weighted are also acquired.

3.4 Implementation

MRSI acquisition is slowed due to the curse of dimensionality. The long acquisition time and low SNR ratio are the main blocking barriers to the full utilization of the richness of MRSI signals. Researchers have proposed many ways to accelerate the acquisition process. Main approaches include reducing repetition time (TR), reducing the required samples, and increasing k-space coverage per echo.

SPectroscopic Imaging by exploiting spatiospectral CorrElation (SPICE) [50, 51] is a rapid, high-resolution, near whole-brain 3D proton MR spectroscopic imaging technique. In a 5-minute scan, the MRSI data were acquired at a nominal resolution of 2.0 x 2.4 x 3.0 mm^3 with whole-brain coverage (FOV = 230 x 230 x 72 mm^3). The relative amount of NAA, lactate, choline, and other commonly used metabolites can be inferred by applying

peak integration on SPICE scans. Variation of SPICE sequences can generate fMRI [52], macromolecule mapping, quantitative susceptibility maps (QSM) [53, 54], myelin water imaging [55], and MRSI with high resolution and SNR. There are also high-field versions [56, 57], phosphorus, carbon, and fluorine MRI [58, 59, 60], and J-resolved MRSI [61].

The power of SPICE comes from the many innovations in its data acquisition and preprocessing techniques. One winning strategy of SPICE is its advanced data processing methods for water and lipid removal [62]. By excluding the water and lipid suppression sequence, SPICE shortens the acquisition time and produces richer information. Sequence-wise, SPICE is unique in utilizing ultrashort echo time (1.6 ms) [58] and short repetition time (160 ms). It has extended MRSI-based readout with a large echo-space (1.76 ms). Sparse sampling further accelerates the process. Because speed, resolution, and SNR are interdependent in imaging, denoising and motion correction innovations also boost the overall performance. In addition, SPICE uses FID instead of spin echoes to encode spatiospectral information to reduce the amount of energy absorbed by the human body, making the process safer for imaged subjects.

SPICE reconstruction uses a union-of-subspaces model [63], which incorporates pre-learned spectral basis functions [64, 65, 66]. An improved LC model-based algorithm employs spatial and spectral priors to accomplish spectral quantification.

3.4.1 Mask Annotation

The neoplastic mass and edema masks were obtained using a level set-based semi-automatic segmentation method. The MPRAGE and SPICE images were co-registered using linear affine transformation using FLIRT.

3.4.2 Metabolite Maps Generation

The metabolite maps are generated using peak integration. It estimates each metabolite by calculating the area under their respective peaks in the frequency domain.

3.4.3 Registration

Due to patient motion in between scans, there might be misalignment between structural scans and MRSI sequence acquisition. Plus, the field of view (FOV) is different in different sequences. And because the ground truth labels can be made more precisely using the high spatial resolution structural data, registration between the high-resolution modalities and the low-resolution modality become crucial. In this experiment, after obtaining high spatial resolution masks from structural scans, we translate them to the same spatial dimension and orientation as the spectroscopic imaging space. The process is carried out by using FLIRT.

FLIRT (FMRIB's Linear Image Registration Tool) is available on the FreeSurfer software application. FLIRT uses a hybrid global-local optimization for registering multiple MRI modalities [67]. In FreeSurfer implementation, the user needs to decide the reference image, degree of freedom (DOF) for the affine transformation, and the cost function for the global optimization. Affine transformation is a type of geometric transformation, which transforms the original data to the same space bijectively. It preserves lines and parallelism, but not distance and angles. For 3D images, the value of DOF can be 6, 7, 9, or 12; each number assumes a different type of transformation. For DOF of 6, a rigid body transformation is assumed. In this type of transformation, the original object is considered a rigid body, and no distortion of it can be made. Instead, only rotation and translation are allowed. This is a valid assumption when two images are of the exact nominal and real resolution. For DOF of 7, a global scale is enabled. Now, the two images can be of different sizes, but no distortion of the field should be present (which distorts the size of each voxel). A DOF of 9 assumes rigid body transformation and independent scaling in each direction, and this accounts for some field inhomogeneity and motion effect that might happen during the acquisition. With full 12 DOF, rigid body, scale, and skews correction are all enabled [68].

3.4.4 nnU-Net

nnU-Net won the second prize on BraTs 2018, and the first prize in BraTS 2020 [69]. It is a framework that automatically adapts itself to any given med-

ical image segmentation task, and it shows top performance in 13 datasets [70, 71]. The self-configuration includes all the design choices in preprocessing, network architecture, training, and post-processing. Preprocessing steps include image and annotation resampling and intensity normalization. There are three network architectures, 2D U-net [72], 3D U-net [73], and cascaded 3D U-net. For each architecture, training parameters include choosing a suitable set of batch size and patch size. The author suggests that although a larger batch size leads to a more accurate gradient estimate, actually to get robust training, any batch size larger than one usually works. Therefore, the author opts for a larger patch size to increase the receptive field while maintaining sufficient depth in the network. The post-processing is tailored to the preprocessing and model design. Besides these dataset-specific parameters, other fixed design choices are essential to building a high-performance model. These are the learning rate, loss function, optimizer, data augmentation workflow, training procedure like the number of epochs and mini-batches, inference procedure, and architecture template, including the choice of instance normalization, leaky ReLU, and deep supervision. The idea of deep supervision [74] is to have multiple segmentation maps generated at different resolution levels, and all participate in the computation of loss function. The decision dependency of these network design choices is thoroughly explained in the paper, and a table is reproduced in Table 3.1 and Table 3.2.

The nnU-Net paper also provides insight into the importance of the training scheme. The author suggests method configuration is more influential on the quality of the model than introducing architectural modifications. According to their observation about the Kidney Tumor Segmentation 2019 challenge hosted by the Medical Image Computing and Computer Assisted Intervention (MICCAI) society, no commonly found variations to u-net can guarantee good performance. Modifications include residual connections [75, 76], dense connections [77], attention mechanisms [78, 79], dilated convolutions [80], and others. The author points out that the same model can perform drastically differently in ranking in the same task. In the same paper, a thorough experiment is carried out, emphasizing the importance of method configuration.

U-net is a fully convolutional neural network that can take input of arbitrary sizes and produce correspondingly-sized output. A fully convolutional neural network avoids using fully connected layers to prevent a rigid struc-

ture. Instead, convolution with kernels that cover the entire input regions can be a good substitute [81].

U-net has a contracting path and a symmetric expanding path with skip connection [72]. Skip connection feeds the output of one layer as the input to another layer and skips the layers in between. A reproduced drawing of the original U-Net architecture can be found in Figure 3.13.

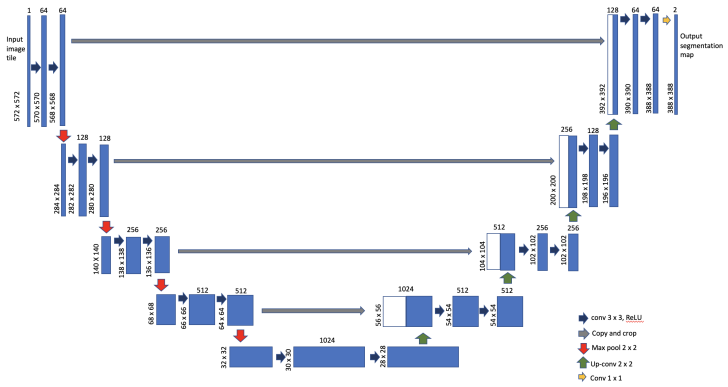

Figure 3.13: The original U-net original architecture. Adapted from source [72].

There is also a 3D counterpart [73]. Figure 3.14 shows its architecture.

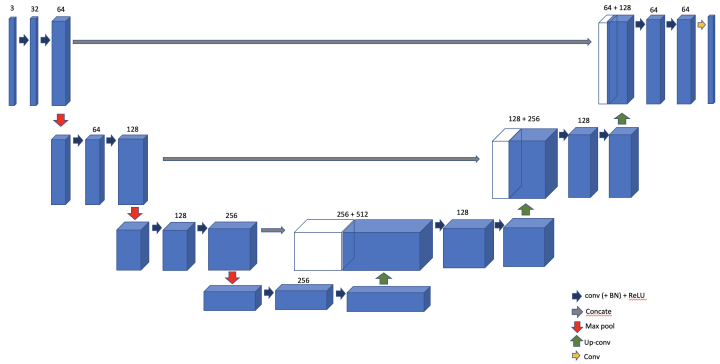

Figure 3.14: The original 3D U-net architecture. Adapted from source [73].

Table 3.1: nnU-Net proposed automated method configuration for deep learning-based biomedical image segmentation (continued in the next table). Adapted from source [70].

Design choice	Required input	Automated (fixed, rule-based or empirical configuration) derived by distilling expert knowledge
Learning rate	-	Poly learning rate schedule (initial, 0.01)
Loss function	-	Dice and cross-entropy
Architecture template	-	Encoder-decoder with skip-connection ('U-Net-like') and instance normalization, leaky ReLU, deep supervision (topology-adapted in inferred parameters)
Optimizer	-	SGD with Nesterov momentum ($\mu = 0.99$)
Data augmentation	-	Rotations, scaling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma correction and mirroring
Training procedure	-	1,000 epochs x 250 minibatches, foreground oversampling
Inference procedure	-	Sliding window with half-patch size overlap, Gaussian patch center weighting
Intensity normalization	Modality, intensity distribution	If CT, global dataset percentile clipping and z score with global foreground mean and s.d. Otherwise, z score with per image mean and s.d.
Image resampling strategy	Distribution of spacings	If anisotropic, in-plane with third-order spline, out-of-plane with nearest neighbor. Otherwise, third-order spline
Annotation resampling strategy	Distribution of spacings	Convert to one-hot encoding $\rightarrow$ If anisotropic, in-plane with linear interpolation, out-of-plane with nearest neighbor. Otherwise, linear interpolation
Image target spacing	Distribution of spacings	If anisotropic, lowest resolution axis tenth percentile, other axes median. Otherwise, median spacing for each axis. (computed based on spacings found in training cases)

Table 3.2: nnU-Net proposed automated method configuration for deep learning-based biomedical image segmentation (continued). Adapted from source [70].

Network topology, patch size, batch size	Median resampled shape, target spacing, GPU memory limit	Initialize the patch size to median image shape and iteratively reduce it while adapting the network topology accordingly until the network can be trained with a batch size of at least 2 given GPU memory constraints. for details see online methods.
Trigger of 3D U-Net cascade	Median resampled image size, patch size	Yes, if patch size of the 3D full resolution U-Net covers less than 12.5% of the median resampled image shape
Configuration of low-resolution 3D U-Net	Median Low-res target spacing or image shapes, GPU memory limit	Iteratively increase target spacing while reconfiguring patch size, network topology and batch size (as described above) until the configured patch size covers 25% of the median image shape
Configuration of post-processing	Full set of training data and annotation	Treating all foreground classes as one; does all-but-largest-component-suppression increase cross-validation performance? Yes, apply; reiterate for individual classes No, do not apply; reiterate for individual foreground classes
Ensemble selection	Full set of training data and annotation	From 2D U-Net, 3D U-Net or 3D cascade, choose the best model (or combination of two) according to cross-validation performance

The developers of nnU-Net retain the general model topology with some modifications. nnU-Net includes instance normalization and uses leaky non-linearities (leaky ReLU) instead of ReLu. They use two computational blocks, each containing conv $\rightarrow$ instance normalization $\rightarrow$ leaky ReLU, per resolution stage in both contracting and expanding paths. Variations in downsampling and upsampling do not play a vital role in performance. For example, max pooling, bi/trilinear upsampling, convolutions transposed, etc. The u-net variant used by nnU-Net is reproduced in Figure 3.15 and Fig 3.16.

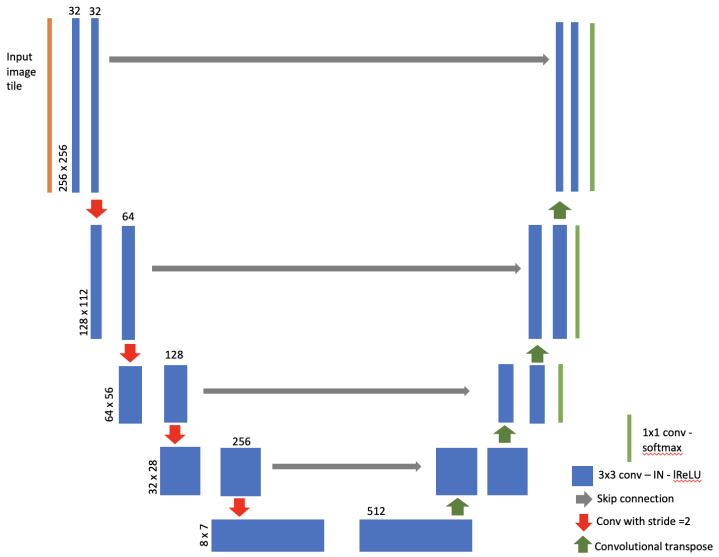

Figure 3.15: The 2D U-net variant used by nnU-Net. Adapted from source [70].

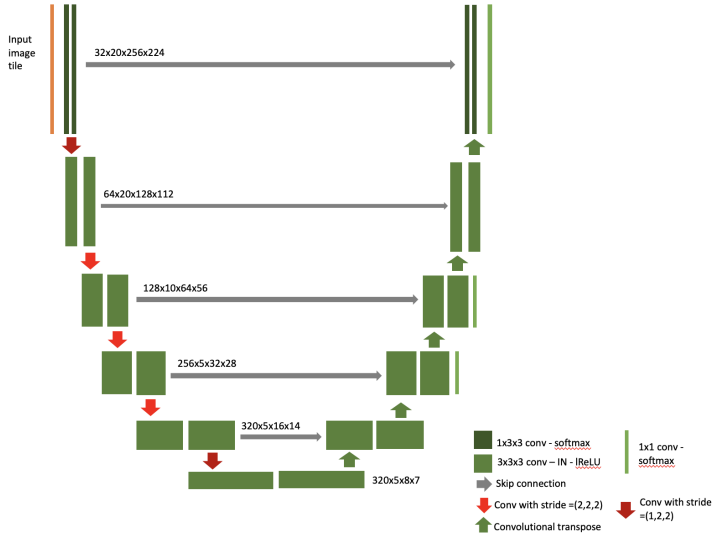

Figure 3.16: The 3D U-net variant used by nnU-Net. Adapted from source [70].

As for the training scheme, an empirical number of 1000 epochs, with each epoch comprising 250 training iterations, is recommended. The best-performing optimizer is stochastic gradient descent with a high initial learning rate and a large Nesterov momentum. They use the 'polyLR' schedule to reduce it. Data augmentation is also shown to be necessary for guaranteeing state-of-the-art performance. Training efficiency is improved by having

data augmented simultaneously as the forward and backward propagation happens, such that GPU and CPU can work concurrently.

Oversampling of rare positive classes approach is taken to mitigate the class imbalance issue. The loss function combines Dice loss and cross-entropy loss because Dice loss alone is a bad approximation of the real Dice value due to patch-based training and over-sampling.

3.5 Evaluation

To show the potential benefits of including MRSI for brain tumor segmentation, we experiment and compare both with and without it. Both settings make use of structural MRI scans and use the same machine learning pipeline, nnU-Net. The first experiment contains structural MRI scans only, including MPRAGE, FLAIR, T1-weighted imaging, and T2-weighted imaging. The second one includes the structural MRI scans mentioned above and the metabolites density maps derived from the MRSI, including creatine, choline, and NAA. Among the eight subjects, one is the testing data. Five-fold cross-validation is performed on the seven training subjects.

During inference, each patch results from an ensemble of five-fold validation. Gaussian importance weighting for softmax aggregation with the patch distance being half of the patch size is selected as recommended by nnU-Net original paper and implementation.

To get a quantitative evaluation of the results, researchers in this field often use Dice, sensitivity (true positive rate), specificity (true negative rate), and accuracy. They are calculated as below

$$DICE = \frac{2TP}{2TP + FP + FN} \tag{3.4}$$

$$sensitivity = \frac{TP}{TP + FN} \tag{3.5}$$

$$specificity = \frac{TN}{TN + FP} \tag{3.6}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.7}$$

where

Table 3.3: TP, TN, FP, FN Table.

	Predict as True	Predict as False
Ground truth is true	True Positive (TP)	False Negative (FN)
Ground truth is false	False Positive (FP)	True Negative (TN)

During the evaluation, subject 118 was excluded because the ground truth mask mismatches with the actual edema region. Subject 108 is excluded because the field of view for structural scans and metabolite maps are different. The five cross-validations are split according to the following table:

Table 3.4: Cross Validation Splits.

Fold	Train Set	Validation Set
0	8, 18, 103, 105, 108	104, 118
1	18, 103, 104, 108, 1118,	8, 105
2	8, 103, 104, 105, 108, 118	18
3	8, 18, 103, 104, 105, 118	108
4	8, 18, 104, 105, 108, 118	103

To get a better visual understanding of the model performance, MPRAGE, FLAIR, T2-weighted, along with ground truth and 4 experiment output overlaying T2-weighted scans, are drawn and compared for each subject. The next chapter covers the discussion about each patient. Network predictions are also overlaid to creatine, choline, and NAA maps for each subject. These visualizations help draw takeaways on whether the model finds metabolite maps useful.

3.6 Comparison to Other Works

Most of the works on brain tumor segmentation focus on the architectural search of the best-performing model on the BraTS dataset. Therefore, only structural scans are concerned. There are some works on using MRSI for brain tumor segmentation.

Simi et al. [82] suggest MRSI and MRS can show the spectral changes of different tissue types. But structural scans can fail in providing sufficient evidence for differentiating white matter and edema, necrosis and gray matter, edema and glioblastoma (GBM).

Different nosologic imaging methods have been proposed to identify brain tumors using MRI and MRSI. Luts et al. [83] explored the combination of registered brain atlas, a subject-specific abnormal tissue prior, and supervised pattern recognition methods. It uses a cascade of data-driven and non-data-driven optimization tasks. For example, k-means clustering for edema detection, LS-SVM for creating the nosologic image, canonical correlation analysis for exploiting spatial information, and kernel logistic regression for providing class probabilities.

Unsupervised methods like convex non-negative matrix factorization (Convex-NMF) is popular in delimitating brain tumor from MRSI data. Ortega-Martorell et al. [84] showcase a successful application of it on seven brain tumor-bearing mice. Li et al. [2] uses NMF on MRSI and fused with MRSI using a wavelet-based approach. This data fusion scheme works well on seven patients with low-grade glioma.

Also using non-negative matrix factorization, Ortega-Martorell et al. [85] propose a Semi-Supervised Source Extraction (SSSE), which outperforms unsupervised methods. The experiment is conducted on mice, and the technique requires a user-defined rough delineation of the tumor region. As the first of the three steps, SSSE uses fisher information and a multi-layer perceptron (MLP) classifier to learn the conditional probabilities of class membership. Next, multi-dimensional scaling approximates the empirical data distribution. Lastly, Convex-NFM constructs the nosologic image.

Works reviewed above mainly use NMF, neural networks that are not convolutional, and optimization of novel cost functions. To the best of my knowledge, there are not many works on using end-to-end convolutional neural networks on MRSI metabolite maps to segment brain tumor subregions. In this work, a state-of-the-art deep convolutional neural network is applied on both the structural scans and the metabolite maps derived from the MRSI to compare their performance.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Metrics

The following tables show the DICE value and accuracy of the model prediction for each subject.

Table 4.1: Edema DICE.

Data	U-Net	Sub 8	Sub 018	Sub 103	Sub 104	Sub 105
Struct	2D	0.620	0.663	0.644	0.331	0.593
Both	2D	0.600	0.625	0.670	0.433	0.606
Struct	3D	0.632	0.668	0.694	0.607	0.541
Both	3D	0.584	0.628	0.745	0.563	0.500

Table 4.2: Edema Accuracy.

Data	U-Net	Sub 8	Sub 018	Sub 103	Sub 104	Sub 105
Struct	2D	0.992	0.992	0.996	0.997	0.981
Both	2D	0.992	0.990	0.997	0.996	0.980
Struct	3D	0.991	0.991	0.996	0.998	0.979
Both	3D	0.992	0.990	0.997	0.997	0.978

Table 4.3: Tumor DICE.

Data	U-Net	Sub 8	Sub 018	Sub 103	Sub 104	Sub 105
Struct	2D	0.142	0.790	0.512	0.318	0.576
Both	2D	0.356	0.678	0.669	0.147	0.649
Struct	3D	0.160	0.706	0.639	0.593	0.534
Both	3D	0.453	0.638	0.598	0.618	0.573

Table 4.4: Tumor accuracy.

Data	U-Net	Sub 8	Sub 018	Sub 103	Sub 104	Sub 105
Struct	2D	0.995	0.996	1.000	0.996	0.990
Both	2D	0.993	0.995	1.000	0.995	0.990
Struct	3D	0.994	0.995	0.999	0.997	0.991
Both	3D	0.995	0.994	1.000	0.997	0.991

For each subject, the experiment with the highest DICE value is highlighted. For edema prediction, 3 out of 5 best methods only use structural scans. For subjects 8, 18, 104, and 105, the two approaches, using structural scans only and using both structural scans and metabolite maps, do not differ much. The difference in performance is within 0.05. For subject 103, using metabolite maps significantly increase the performance.

For tumor prediction, 4 out of 5 best performances are using a combination of structural scans and metabolite maps. Both approaches yield much better results than the structural only approach when making predictions on subjects 8 and 105. These observations indicate the potential benefits of incorporating MRSI scans in the brain tumor segmentation pipeline. It gives a better view of tumor boundary and therefore be important in tumor treatment and prognosis.

Note that the evaluation here favors structural scans. All ground truth maps are obtained from structural scans, but the definition of tumor and edema boundary is still under investigation. With the ever-advancing MR techniques, researchers are gradually learning about the brain pathological pathways. A better definition of tumor and edema might be raised and accepted in the future.

4.2 Visualizing Each Subject

Subject 8

Row 4 represents the ground truth, where dark red represents the edema region and bright red represents the tumor region. Because the gold standard mask is labeled in a higher nominal spatial resolution space, the mask becomes ragged and discontinued after registration and re-sampling to a lower-dimensional space. For the four predictions shown, metabolite results tend to be smoother. The 2D structural result has holes that are not very biologically plausible. 3D structural scans even show edema prediction on slice 28. This is probably because, in some structural scans, these voxels have high intensity due to noise in the scan. The model is biased, and it will misclassify in such cases. Metabolite pipeline, on the other hand, shows a more consistent performance. It is less affected by such noises because the metabolite maps are quantitative.

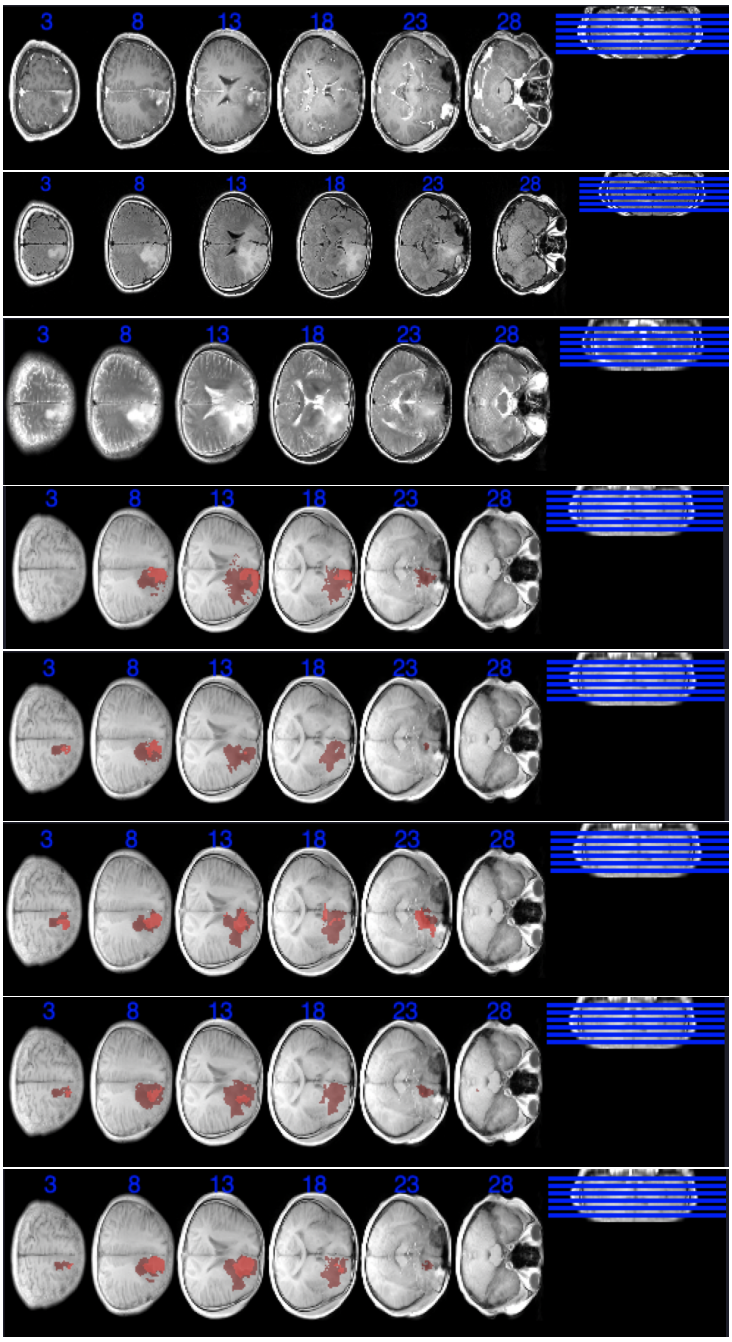

Figure 4.1: Axial view of subject 8. From top to bottom: MPRAGE, FLAIR, T2-weighted, Ground Truth, 2D structural pipeline prediction, 2D metabolic pipeline prediction, 3D structural pipeline prediction, 3D metabolic pipeline prediction.

Subject 18

Again, row 4 is the ground truth, which is incorrect due to the loss of information during registration and compression. A tumor almost constitutes the entire abnormal mask. It is very distinctive on FLAIR and T2-weighted images in human eyes. All four pipelines are doing an excellent job in identifying the tumor region. I think the 2D metabolite pipeline does a pretty solid job. Note that on slice 8, 2D metabolite prediction matches the bright area in FLAIR. Again, the 2D structural prediction has an unnatural cavity in slice 18.

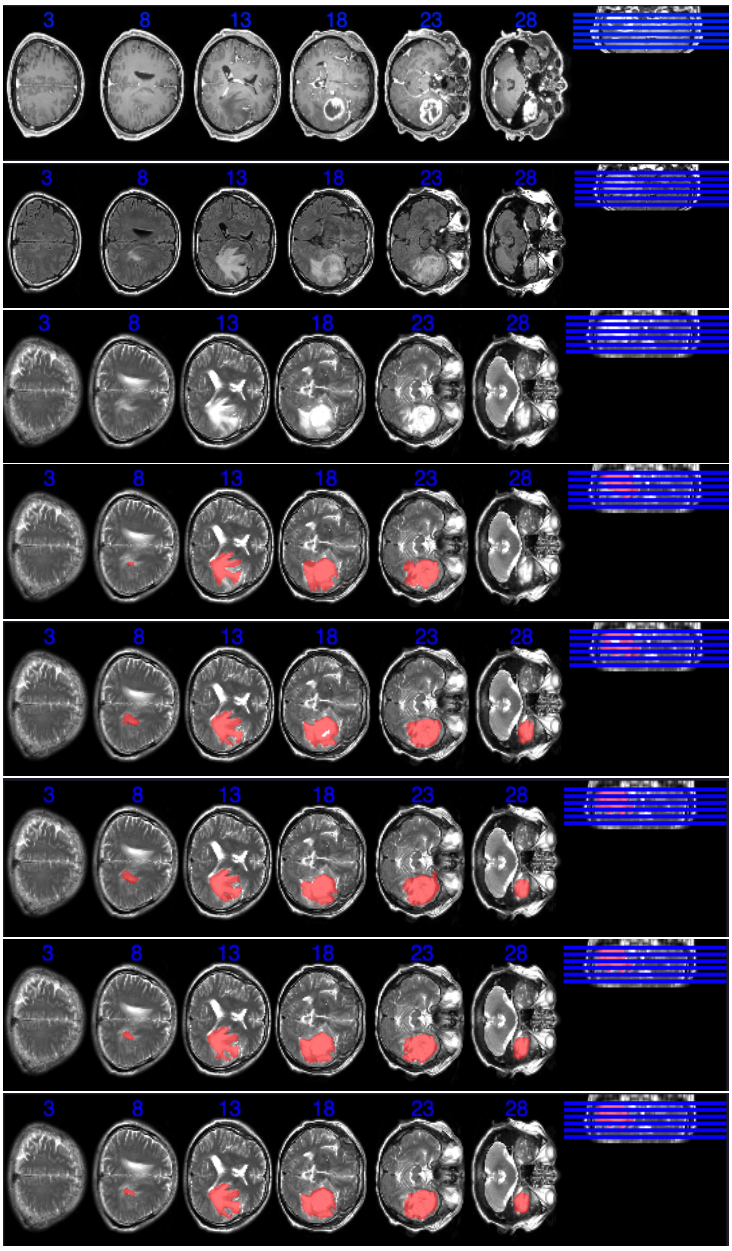

Figure 4.2: Axial view of subject 18. From top to bottom: MPRAGE, FLAIR, T2-weighted, Ground Truth, 2D structural pipeline prediction, 2D metabolic pipeline prediction, 3D structural pipeline prediction, 3D metabolic pipeline prediction.

Subject 103

This subject has a significant lesion with a small purely edema region. The damaged part matches with the high-intensity area in FLAIR. Both 3D pipeline results compare with the ground truth well in both shape and location. 2D metabolic pipeline produces a low score for all voxels. 2D structural scans mistakenly predict them as edema regions instead of a tumor. 3D metabolic predictions are smoother than the structural one, which is desirable in this task. In slice 8, the structural pipeline produces a false positive prediction in the CSF region in the middle of the brain.

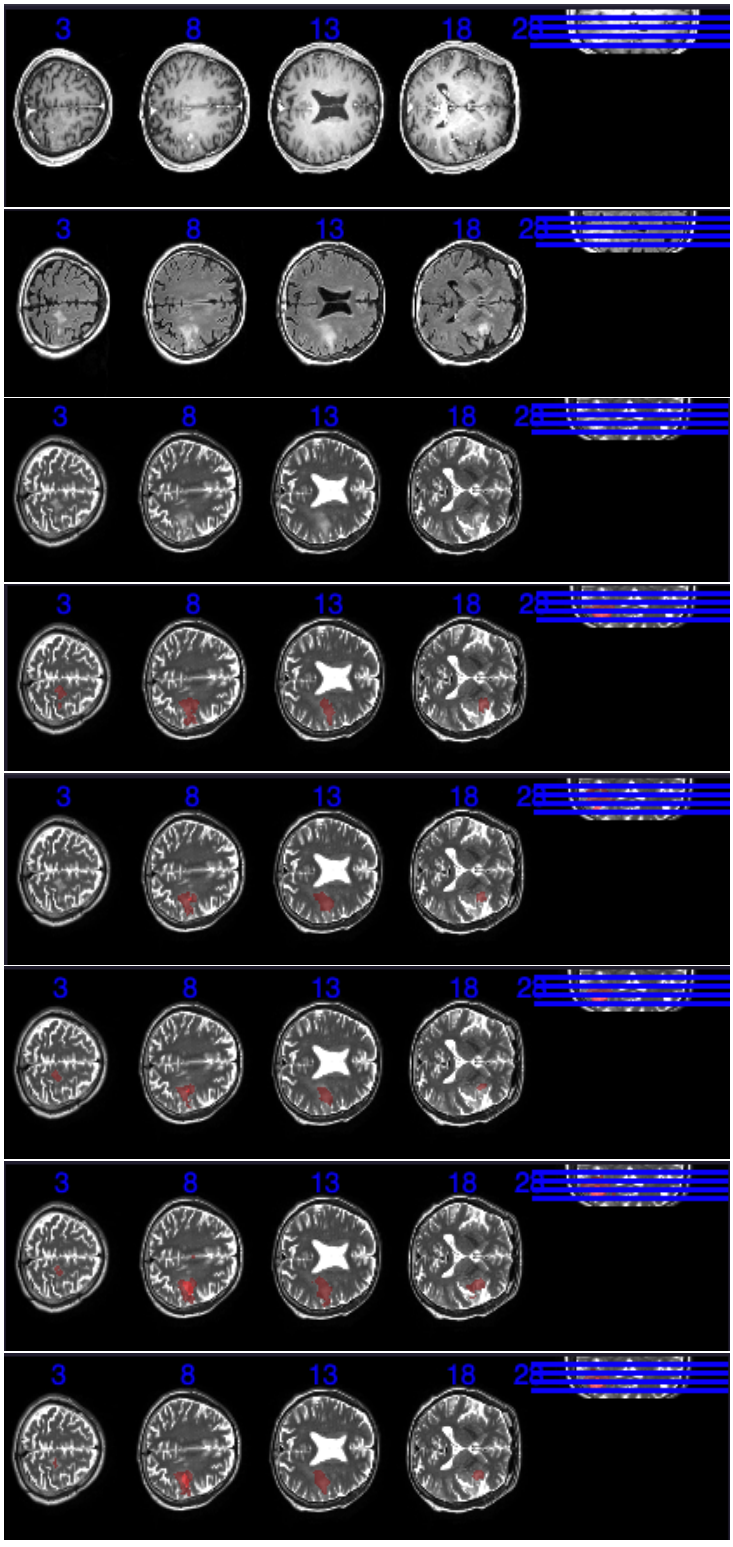

Figure 4.3: Axial view of subject 103. From top to bottom: MPRAGE, FLAIR, T2-weighted, Ground Truth, 2D structural pipeline prediction, 2D metabolic pipeline prediction, 3D structural pipeline prediction, 3D metabolic pipeline prediction.

Subject 104

Subject 104 has a large lesion area and tiny pure edema region. 2D networks are doing better. Especially the metabolic pipeline.

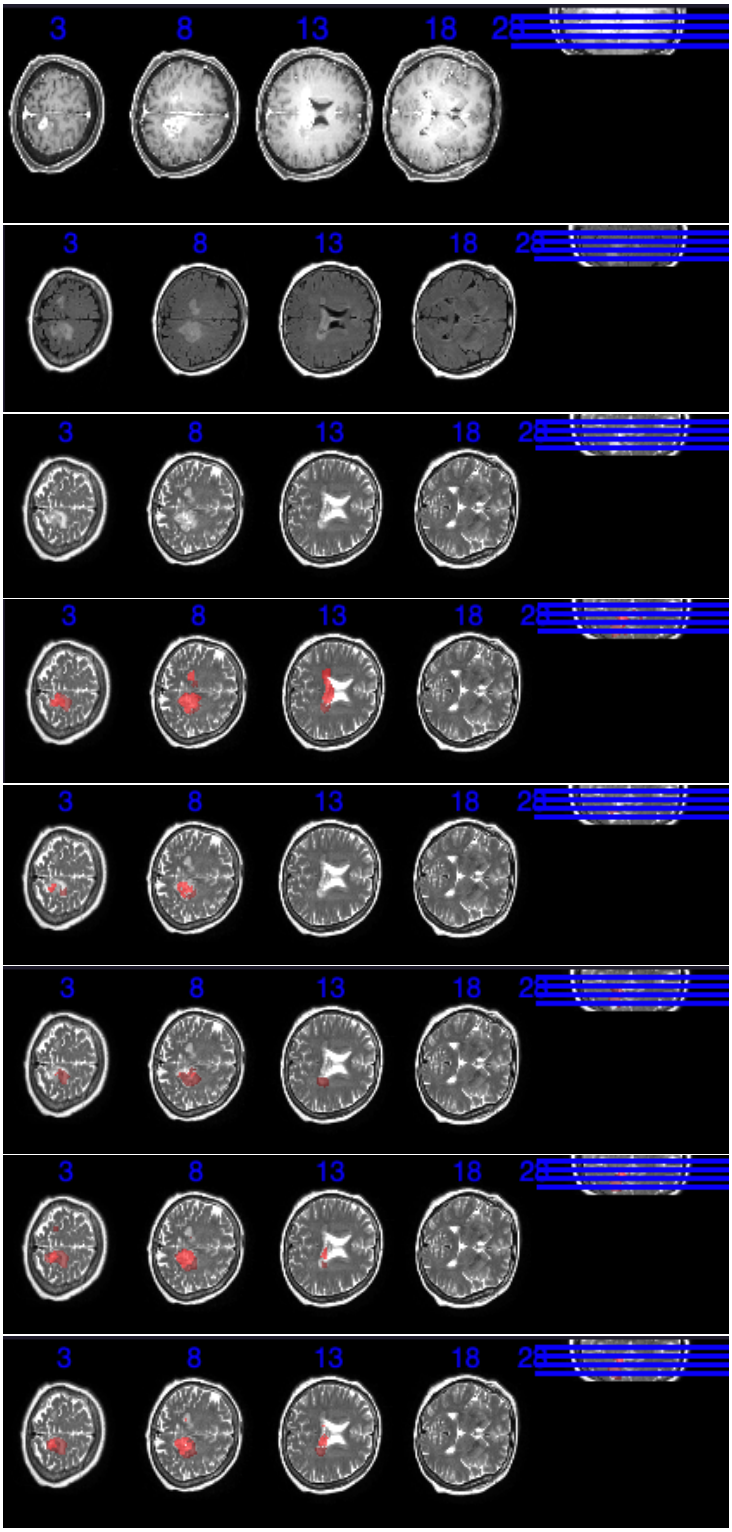

Figure 4.4: Axial view of subject 104. From top to bottom: MPRAGE, FLAIR, T2-weighted, Ground Truth, 2d structural pipeline prediction, 2d metabolic pipeline prediction, 3d structural pipeline prediction, 3d metabolic pipeline prediction.

Subject 105

We have a large tumor mask with a tiny pure edema region identified for this subject. 3D networks seem to underestimate the affected area. There might be fewer data in the training data set to have such a large tumor region. 2D metabolic prediction makes a more accurate prediction on slices 3 and 18.

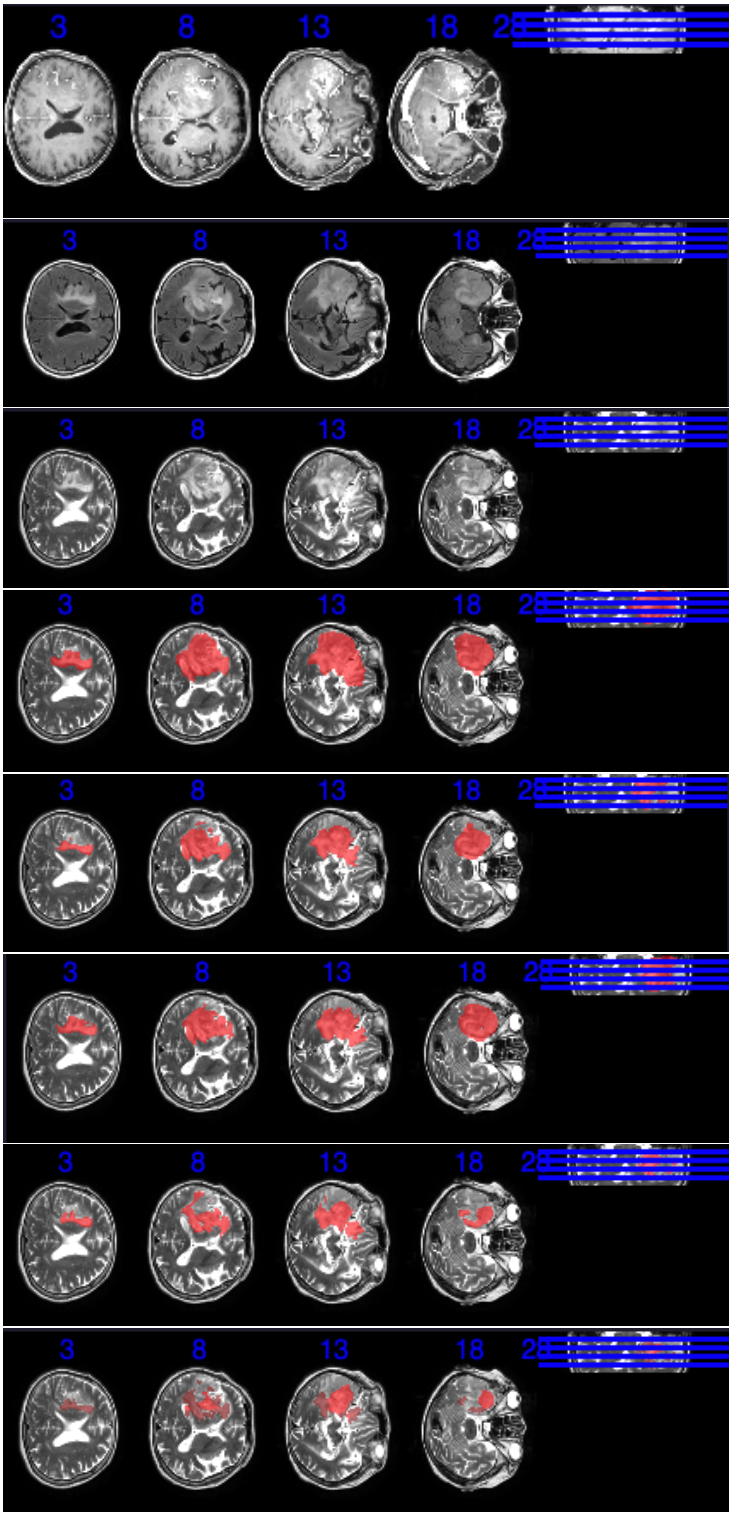

Figure 4.5: Axial view of subject 105. From top to bottom: MPRAGE, FLAIR, T2-weighted, Ground Truth, 2D structural pipeline prediction, 2D metabolic pipeline prediction, 3D structural pipeline prediction, 3D metabolic pipeline prediction.

Subject 128

Subject 128 is the testing data that does not have ground truth.

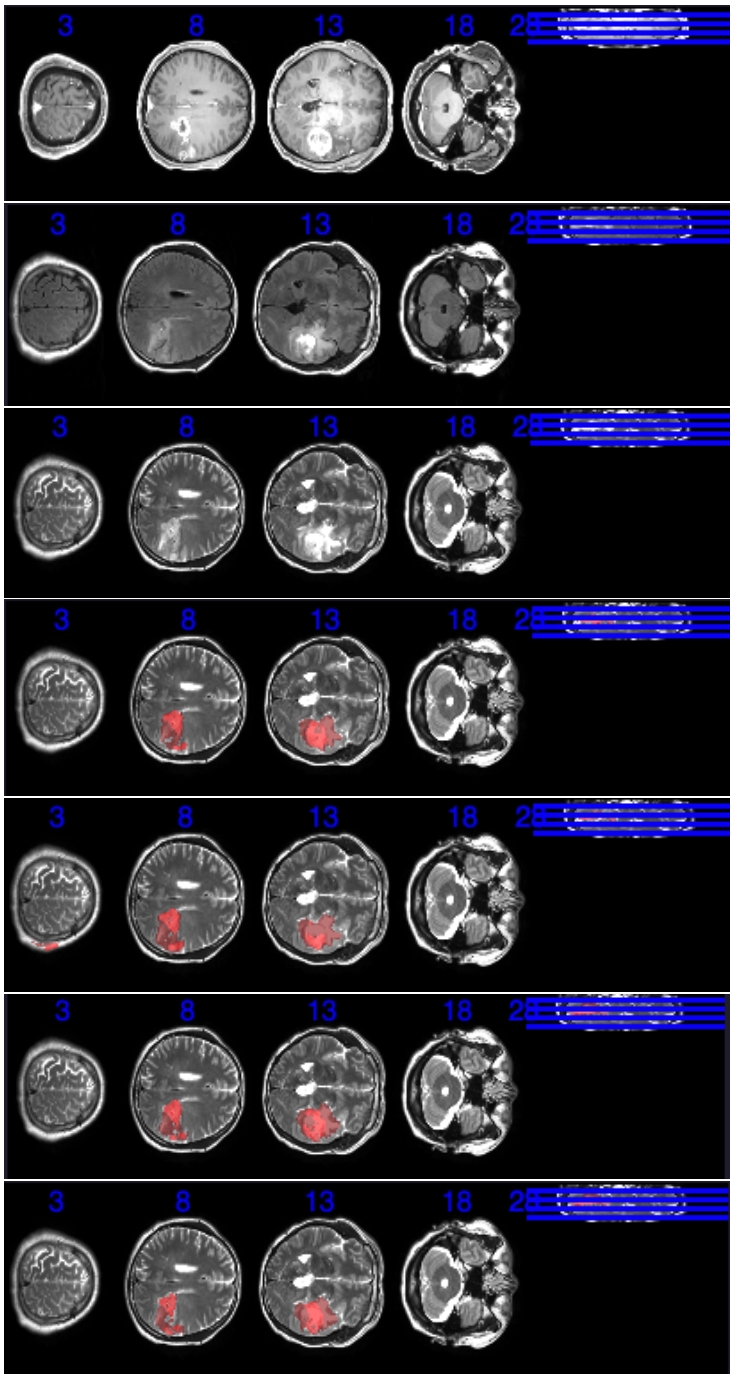

Figure 4.6: Axial view of subject 128. From top to bottom: MPRAGE, FLAIR, T2-weighted, 2d structural pipeline prediction, 2d metabolic pipeline prediction, 3d structural pipeline prediction, 3d metabolic pipeline prediction.

4.2.1 Visualizing prediction and compare with metabolite maps

This section shows the network overlaying metabolite maps. These visualizations are intended to provide intuition on how network prediction correlates with changes in metabolite concentration.

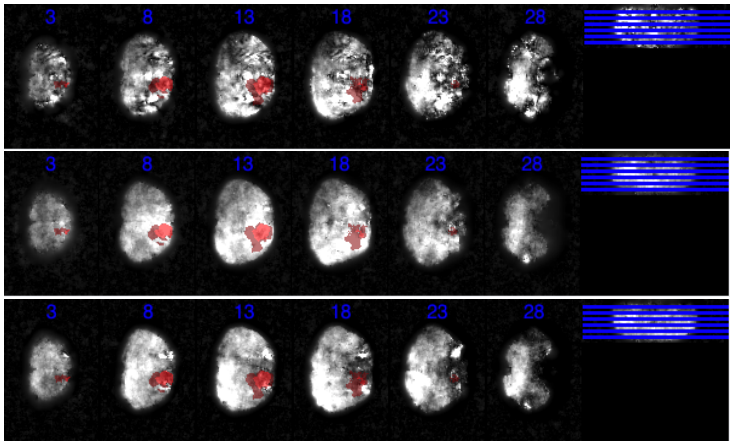

Figure 4.7: 3D metabolic prediction overlays metabolite maps for subject 8. From top to bottom: Creatine, Choline, NAA.

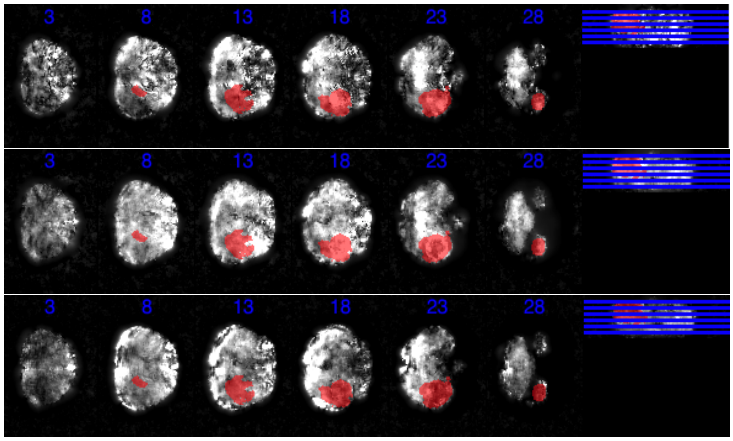

Figure 4.8: 2D metabolic prediction overlays metabolite maps for subject 18. From top to bottom: Creatine, Choline, NAA.

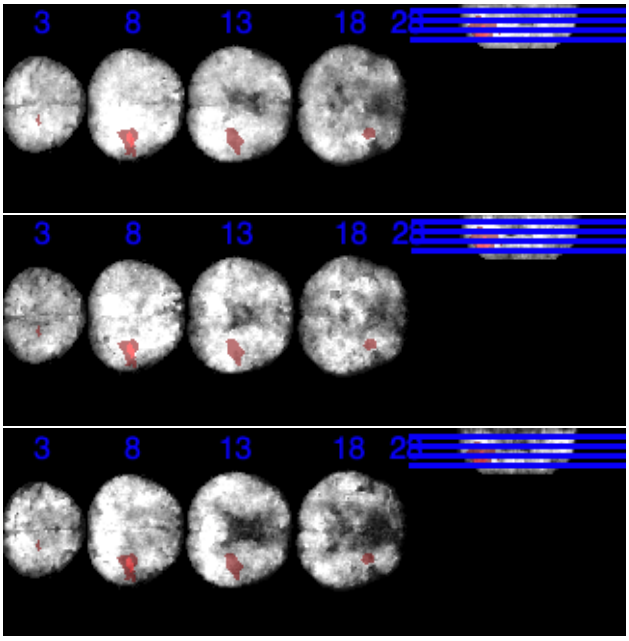

Figure 4.9: 3D metabolic prediction overlays metabolite maps for subject 103. From top to bottom: Creatine, Choline, NAA.

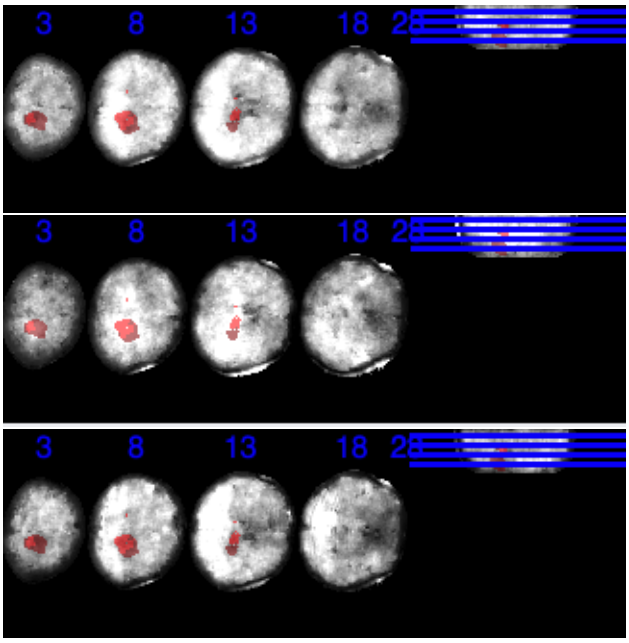

Figure 4.10: 3D metabolic prediction overlays metabolite maps for subject 104. From top to bottom: Creatine, Choline, NAA.

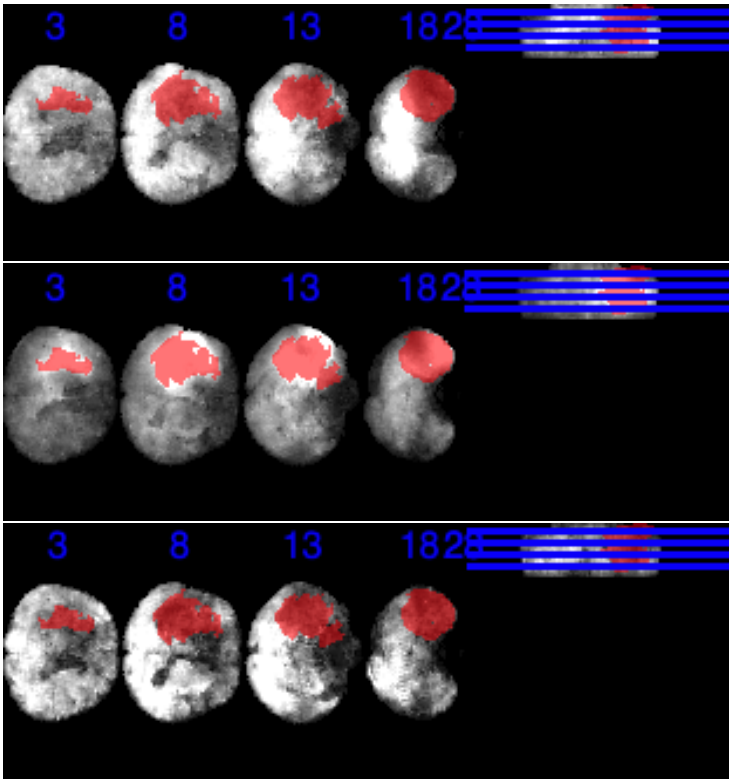

Figure 4.11: 2D metabolic prediction overlays metabolite maps for subject 105. From top to bottom: Creatine, Choline, NAA.

For all predictions, choline level elevates, and NAA concentration decreases in the tumor region. It is most evident in subject 105, as the prediction shape closely follows the metabolite maps.

4.3 Discussion

One of the challenges that conventional MRI scan faces in identifying brain tumor are the contrast. Although MRI is preferred over other imaging techniques due to its fine contrast among different brain tissues, different tissues can still have voxels with the same intensity. A human can separate them because we can identify boundaries quickly, and we have a large receptive field. However, due to the limited view of a neural network, it can get confused. On the other hand, Metabolite maps are helpful because they are quantitative. It is less affected by field inhomogeneity and therefore behaves more consistently. They are more sensitive and specific to brain tumors.

This proof-of-concept study strikes a fair comparison between having and

not having metabolite maps in segmenting tumor sub-region in the brain. nnU-Net has a provenly highly optimized way of identifying hyper-parameters and deciding data processing steps. I find that with metabolite maps, the model can learn a smoother and more accurate way of identifying abnormal regions. Metric-wise, both having and not having metabolite maps produce the same level of performance on the edema region. Having metabolite maps helps tumor region segmentation.

However, this is still an underestimate of the use of metabolic maps. The metabolite maps are not denoised yet. There are evident background noises, which might confuse the model. Secondly, all ground truth labels are derived from the structural scans. Some ongoing research challenges this definition of edema and tumor. Diffusion imaging, perfusion imaging, and spectroscopic imaging can reflect key indicators about these regions, and they disagree with the conventional structural scans like MPRAGE, FLAIR, and T2-weighted images. The differences in the range of abnormal tissue reflect the different pathological states in tumor growth. Therefore, the gold standard in the current study favors the structural approach.

Lastly, the research community has identified more metabolites than choline, creatine, and NAA being useful. More metabolites or even the original spectra should become network input in the future.

To conclude, MRSI is a promising technique for better understanding tumors and developing better automated clinical tools.

CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 Summary of Findings

This thesis has given an overview of brain tumor segmentation and proposed using MRSI to improve performance. Brain tumor has become an increasingly important health issue worldwide. MR imaging is the first choice of its neuroimaging. Segmentation of MR images becomes an essential step in the therapy planning and prognosis of brain tumor treatment. The common paradigms include manual segmentation, semi-automatic and fully automated methods, unsupervised methods, and supervised methods. After 2014-2015, the focus of brain tumor research has shifted to deep convolutional neural network methods. The second chapter also reviews research works in each category.

This work also explains the reasoning and shows a proof-of-concept experiment on how MRSI's metabolite density maps can help segment brain tumors better. By including creatine, choline, and NAA metabolite density maps derived from the MRSI data, the model of choice can increase its DICE value in identifying tumors. The close examination of results implies that the neural network relies heavily on metabolites density information to identify the tumor region.

5.2 Future Work

More metabolite maps should be incorporated in future research because metabolites are essential in determining brain tumor regions. According to the literature, lactate, lipid and MI are related to brain tumor pathology and are detectable by MRSI. More data is desirable to maximize the performance

so that the model can generalize better.

On the model side, other network architecture and modifications should be tested with robust hyper-parameters tuning.

REFERENCES

[1] A. Stadlbauer, E. Moser, S. Gruber, R. Buslei, C. Nimsky, R. Fahlbusch, and O. Ganslandt, "Improved delineation of brain tumors: an automated method for segmentation based on pathologic changes of 1h-mrsi metabolites in gliomas," *Neuroimage*, vol. 23, no. 2, pp. 454–461, 2004.

[2] Y. Li, X. Liu, F. Wei, D. M. Sima, S. Van Cauter, U. Himmelreich, Y. Pi, G. Hu, Y. Yao, and S. Van Huffel, "An advanced mri and mrsi data fusion scheme for enhancing unsupervised brain tumor differentiation," *Computers in biology and medicine*, vol. 81, pp. 121–129, 2017.

[3] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2021." *CA: a cancer journal for clinicians*, vol. 71, no. 1, pp. 7–33, 2021.

[4] K. D. Miller, M. Fidler-Benaoudia, T. H. Keegan, H. S. Hipp, A. Jemal, and R. L. Siegel, "Cancer statistics for adolescents and young adults, 2020," *CA: a cancer journal for clinicians*, vol. 70, no. 6, pp. 443–459, 2020.

[5] M. Zhou, J. Scott, B. Chaudhury, L. Hall, D. Goldgof, K. W. Yeom, M. Iv, Y. Ou, J. Kalpathy-Cramer, S. Napel et al., "Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches," *American Journal of Neuroradiology*, vol. 39, no. 2, pp. 208–216, 2018.

[6] J. R. McFaline-Figueroa and E. Q. Lee, "Brain tumors," *The American journal of medicine*, vol. 131, no. 8, pp. 874–882, 2018.

[7] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art survey on mri brain tumor segmentation," *Magnetic resonance imaging*, vol. 31, no. 8, pp. 1426–1438, 2013.

[8] G. Kaur and J. Rani, "Mri brain tumor segmentation methods-a review," 2016.

[9] A. Wadhwa, A. Bhardwaj, and V. S. Verma, "A review on brain tumor segmentation of mri images," *Magnetic resonance imaging*, vol. 61, pp. 247–259, 2019.

[10] J. Liu, M. Li, J. Wang, F. Wu, T. Liu, and Y. Pan, "A survey of mri-based brain tumor segmentation methods," *Tsinghua science and technology*, vol. 19, no. 6, pp. 578–595, 2014.

[11] I. Diaz, P. Boulanger, R. Greiner, and A. Murtha, "A critical review of the effects of de-noising algorithms on mri brain tumor segmentation," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 3934–3937.

[12] J. Nalepa, M. Marcinkiewicz, and M. Kawulok, "Data augmentation for brain-tumor segmentation: a review," *Frontiers in computational neuroscience*, vol. 13, p. 83, 2019.

[13] A. Işın, C. Direkoğlu, and M. Şah, "Review of mri-based brain tumor image segmentation using deep learning methods," *Procedia Computer Science*, vol. 102, pp. 317–324, 2016.

[14] K. Y. Lim and R. Mandava, "A multi-phase semi-automatic approach for multisequence brain tumor image segmentation," *Expert Systems with Applications*, vol. 112, pp. 288–300, 2018.

[15] 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*, vol. 34, no. 10, pp. 1993–2024, 2014.

[16] E. Vansteenkiste, "Quantitative analysis of ultrasound images of the preterm brain," Ph.D. dissertation, Ghent University, 2007.

[17] D. R. White, A. S. Houston, W. F. Sampson, and G. P. Wilkins, "Intra- and interoperator variations in region-of-interest drawing and their effect on the measurement of glomerular filtration rates," *Clinical nuclear medicine*, vol. 24, no. 3, pp. 177–181, 1999.

[18] D. C. Collier, S. S. Burnett, M. Amin, S. Bilton, C. Brooks, A. Ryan, D. Roniger, D. Tran, and G. Starkschall, "Assessment of consistency in contouring of normal-tissue anatomic structures," *Journal of applied clinical medical physics*, vol. 4, no. 1, pp. 17–24, 2003.

[19] X. Zhang, X. Li, H. Li, and Y. Feng, "A semi-automatic brain tumor segmentation algorithm," in *2016 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2016, pp. 1–6.

[20] A. Yezzi, S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum, "A geometric snake model for segmentation of medical imagery," *IEEE Transactions on medical imaging*, vol. 16, no. 2, pp. 199–209, 1997.

[21] Y. Zhu, G. S. Young, Z. Xue, R. Y. Huang, H. You, K. Setayesh, H. Hatabu, F. Cao, and S. T. Wong, "Semi-automatic segmentation software for quantitative clinical brain glioblastoma evaluation," *Academic radiology*, vol. 19, no. 8, pp. 977–985, 2012.

[22] 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*, vol. 23, pp. S208–S219, 2004.

[23] 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*, vol. 45, no. 1, pp. S173–S186, 2009.

[24] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[25] M. Prastawa, E. Bullitt, N. Moon, K. Van Leemput, and G. Gerig, "Automatic brain tumor segmentation by subject specific modification of atlas priors1," *Academic radiology*, vol. 10, no. 12, pp. 1341–1348, 2003.

[26] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," Stanford, Tech. Rep., 2006.

[27] B. Srinivas and G. S. Rao, "Unsupervised learning algorithms for mri brain tumor segmentation," in *2018 Conference on Signal Processing And Communication Engineering Systems (SPACES)*. IEEE, 2018, pp. 181–184.

[28] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters," 1973.

[29] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

[30] N. Sauwen, M. Acou, S. Van Cauter, D. Sima, J. Veraart, F. Maes, U. Himmelreich, E. Achten, and S. Van Huffel, "Comparison of unsupervised classification methods for brain tumor segmentation using multiparametric mri," *NeuroImage: Clinical*, vol. 12, pp. 753–764, 2016.

[31] M. Ghaffari, A. Sowmya, and R. Oliver, "Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the brats 2012–2018 challenges," *IEEE reviews in biomedical engineering*, vol. 13, pp. 156–168, 2019.

[32] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[33] J. R. Quinlan, *C4. 5: programs for machine learning*. Elsevier, 2014.

[34] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. Routledge, 2017.

[35] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[36] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[37] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[38] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological physics and technology*, vol. 10, no. 3, pp. 257–273, 2017.

[39] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," *Classification in BioApps*, pp. 323–350, 2018.

[40] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[41] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," *arXiv preprint arXiv:1603.07285*, 2016.

[42] A. L. Maas, A. Y. Hannun, A. Y. Ng et al., "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30, no. 1. Citeseer, 2013, p. 3.

[43] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

[44] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.

[45] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.

[46] M. C. Preul, Z. Caramanos, D. L. Collins, J.-G. Villemure, R. Leblanc, A. Olivier, R. Pokrupa, and D. L. Arnold, "Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy," *Nature medicine*, vol. 2, no. 3, pp. 323–325, 1996.

[47] F. S. De Edelenyi, C. Rubin, F. Esteve, S. Grand, M. Decorps, V. Lefournier, J.-F. Le Bas, and C. Remy, "A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images," *Nature medicine*, vol. 6, no. 11, pp. 1287–1289, 2000.

[48] C. Choi, S. K. Ganji, R. J. DeBerardinis, K. J. Hatanpaa, D. Rakheja, Z. Kovacs, X.-L. Yang, T. Mashimo, J. M. Raisanen, I. Marin-Valencia et al., "2-hydroxyglutarate detection by magnetic resonance spectroscopy in idh-mutated patients with gliomas," *Nature medicine*, vol. 18, no. 4, pp. 624–629, 2012.

[49] O. Keunen and S. P. Niclou, "Is there a prominent role for mr spectroscopy in the clinical management of brain tumors?" 2020.

[50] F. Lam, C. Ma, B. Clifford, C. L. Johnson, and Z.-P. Liang, "High-resolution 1h-mrsi of the brain using spice: data acquisition and image reconstruction," *Magnetic resonance in medicine*, vol. 76, no. 4, pp. 1059–1070, 2016.

[51] C. Ma, F. Lam, and Z.-P. Liang, "Encoding and decoding with prior knowledge: From slim to spice," *Handbook of Magnetic Resonance Spectroscopy In Vivo: MRS Theory, Practice and Applications*, p. 141, 2016.

[52] R. Guo, Y. Zhao, Y. Li, Y. Li, and Z.-P. Liang, "Simultaneous metabolic and functional imaging of the brain using spice," *Magnetic resonance in medicine*, vol. 82, no. 6, pp. 1993–2002, 2019.

[53] X. Peng, F. Lam, Y. Li, B. Clifford, and Z.-P. Liang, "Simultaneous qsm and metabolic imaging of the brain using spice," *Magnetic resonance in medicine*, vol. 79, no. 1, pp. 13–21, 2018.

[54] R. Guo, Y. Zhao, Y. Li, T. Wang, Y. Li, B. Sutton, and Z.-P. Liang, "Simultaneous qsm and metabolic imaging of the brain using spice: Further improvements in data acquisition and processing," *Magnetic Resonance in Medicine*, vol. 85, no. 2, pp. 970–977, 2021.

[55] T. Wang, D. Wang, Y. Hu, R. Guo, Y. Li, Y. Zhao, J. Liu, Z.-P. Liang, and Y. Li, "Simultaneous myelin water imaging and 3d h-mrsi relates myelin degradation to neurometabolic changes in mild traumatic brain injury patients."

[56] H. Lee, J. E. Song, J. Shin, E. Joe, C. G. Joo, Y.-S. Choi, H.-T. Song, and D.-H. Kim, "High resolution hyperpolarized 13c mrsi using spice at 9.4 t," *Magnetic resonance in medicine*, vol. 80, no. 2, pp. 703–710, 2018.

[57] R. Guo, Y. Zhao, Y. Li, P. Bhattacharyya, M. Lowe, H. M. Wiesner, Y. Li, X.-H. Zhu, W. Chen, and Z.-P. Liang, "Fast high-resolution h-mrsi of the human brain at 7t," in *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 29. Virtual Meeting, 2021.

[58] C. Ma, F. Lam, Q. Ning, C. L. Johnson, and Z.-P. Liang, "High-resolution 1h-mrsi of the brain using short-te spice," *Magnetic resonance in medicine*, vol. 77, no. 2, pp. 467–479, 2017.

[59] J. E. Song, J. Shin, H. Lee, Y.-S. Choi, H.-T. Song, and D.-H. Kim, "Dynamic hyperpolarized 13c mr spectroscopic imaging using spice in mouse kidney at 9.4 t," *NMR in Biomedicine*, vol. 33, no. 3, p. e4230, 2020.

[60] Y. Zhao, T. K. Hitchens, M. Herneisey, J. M. Janjic, R. Guo, Y. Li, and Z.-P. Liang, "Fast high-resolution 19f-mrsi of perfluorocarbon nanoemulsions for mri cell tracking using spice with learned subspaces."

[61] L. Tang, Y. Zhao, Y. Li, R. Guo, B. Clifford, G. El Fakhri, C. Ma, Z.-P. Liang, and J. Luo, "Accelerated j-resolved 1h-mrsi with limited and sparse sampling of (-space," *Magnetic Resonance in Medicine*, vol. 85, no. 1, pp. 30–41, 2021.

[62] Y. Li, F. Lam, R. Guo, B. Clifford, X. Peng, and Z. Liang, "Removal of water sidebands from 1h-mrsi data acquired without water suppression," in *Proceedings of the Joint Annual Meeting of ISMRM-ESMRMB, Paris, France*, 2018, p. 288.

[63] C. Ma, F. Lam, C. L. Johnson, and Z.-P. Liang, "Removal of nuisance signals from limited and sparse 1h mrsi data using a union-of-subspaces model," *Magnetic resonance in medicine*, vol. 75, no. 2, pp. 488–497, 2016.

[64] F. Lam, Y. Li, R. Guo, B. Clifford, and Z.-P. Liang, "Ultrafast magnetic resonance spectroscopic imaging using spice with learned subspaces," *Magnetic resonance in medicine*, vol. 83, no. 2, pp. 377–390, 2020.

[65] Z. Wang, Y. Li, and F. Lam, "High-resolution, 3d multi-te 1h mrsi using fast spatiospectral encoding and subspace imaging," *Magnetic Resonance in Medicine*, 2021.

[66] F. Lam, Y. Li, R. Guo, B. Clifford, X. Peng, and Z.-P. Liang, "Further accelerating spice for ultrafast mrsi using learned spectral features."

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

[68] 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*, vol. 17, no. 2, pp. 825–841, 2002.

[69] F. Isensee, P. F. Jäger, P. M. Full, P. Vollmuth, and K. H. Maier-Hein, "nnu-net for brain tumor segmentation," in *International MICCAI Brainlesion Workshop*. Springer, 2020, pp. 118–132.

[70] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.

[71] F. Isensee, J. Petersen, S. A. Kohl, P. F. Jäger, and K. H. Maier-Hein, "nnu-net: Breaking the spell on successful medical image segmentation," *arXiv preprint arXiv:1904.08128*, vol. 1, pp. 1–8, 2019.

[72] 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*. Springer, 2015, pp. 234–241.

[73] 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*. Springer, 2016, pp. 424–432.

[74] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Artificial intelligence and statistics*. PMLR, 2015, pp. 562–570.

[75] 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*, 2016, pp. 770–778.

[76] M. U. Rehman, S. Cho, J. H. Kim, and K. T. Chong, "Bu-net: Brain tumor segmentation using modified u-net architecture," *Electronics*, vol. 9, no. 12, p. 2203, 2020.

[77] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[78] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[79] J. Zhang, X. Lv, H. Zhang, and B. Liu, "Aresu-net: Attention residual u-net for brain tumor segmentation," *Symmetry*, vol. 12, no. 5, p. 721, 2020.

[80] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[81] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[82] V. Simi and J. Joseph, "Segmentation of glioblastoma multiforme from mr images–a comprehensive review," *The Egyptian Journal of Radiology and Nuclear Medicine*, vol. 46, no. 4, pp. 1105–1110, 2015.

[83] J. Luts, T. Laudadio, A. J. Idema, A. W. Simonetti, A. Heerschap, D. Vandermeulen, J. A. Suykens, and S. Van Huffel, "Nosologic imaging of the brain: segmentation and classification using mri and mrsi," *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In vivo*, vol. 22, no. 4, pp. 374–390, 2009.

[84] S. Ortega-Martorell, P. J. Lisboa, A. Vellido, R. V. Simões, M. Pumarola, M. Julià-Sapé, and C. Arús, "Convex non-negative matrix factorization for brain tumor delimitation from mrsi data," 2012.

[85] S. Ortega-Martorell, A. P. Candiota, R. Thomson, P. Riley, M. Julia-Sape, and I. Olier, "Embedding mri information into mrsi data source extraction improves brain tumour delineation in animal models," *PloS one*, vol. 14, no. 8, p. e0220809, 2019.