GRAPH BASED DEEP LEARNING MODELS FOR ANALYSIS OF RESTING STATE FMRI WITH APPLICATIONS IN LOCALIZATION AND DYNAMIC FUNCTIONAL CONNECTIVITY

by

Naresh Nandakumar

A dissertation submitted to Johns Hopkins University

in conformity with the requirements for the degree of

Doctor of Philosophy

Baltimore, Maryland

September, 2023

Abstract

Resting-state functional magnetic resonance imaging (rs-fMRI) is a noninvasive neuroimaging modality that quantifies the changes in blood flow and oxygenation in the brain at rest. Analyzing connectivity graphs extracted from rs-fMRI can yield insight into the functional organization of the brain.

Neurosurgical resection procedures require immense precision, as the surgeon must remove as much of the lesion while preserving maximum functionality. An incorrect incision could cause severe or even permanent cognitive deficits. Rs-fMRI has emerged as a preoperative mapping modality for eloquent cortex localization for brain tumor removal surgeries as well as epileptogenic zone removal surgeries in patients with epilepsy.

Rs-fMRI connectivity analysis usually begins with applying an existing parcellation to define regions of interest (ROI's) of spatially and temporally homogeneous areas of the brain. However, these existing parcellations do not generalize well to patients with brain tumors or epilepsy due to an atypical rs-fMRI signature.

We present a collection of deep learning methods to analyze rs-fMRI of atypical populations, namely brain tumor and focal epilepsy subjects, to perform localization of regions of interest for neurosurgery. First, we explore

techniques to develop more accurate subject-specific parcellations for downstream analysis using refinement techniques. We develop a Bayesian model with a markov random field prior to refine parcellations on a subject-specific basis. We then present RefineNet, which jointly optimizes parcellation refinement and the downstream tasks.

Then, we present our models on eloquent cortex localization for tumor patients. We leverage graph neural networks to perform localization. We extend our model using both temporal and spatial attention models applied to dynamic connectivity, where our attention mechanisms capture spatiotemporal features that boost localization performance.

Next, we present our models on epileptogenic zone localization for epilepsy patients. We develop a graph convolutional network called DeepEZ which uses anatomical connectivity regularization and a biologically inspired loss function. We extend this work to the dynamic connectivity case as well, using transformer networks to capture temporal nuances. Lastly, we tackle the noisy label problem through the context of epileptogenic zone localization, and develop a framework to perform localization even when trained on a dataset with entirely noisy labels.

Thesis Committee

Primary Readers

Dr. Archana Venkataraman (Primary Advisor)
John C. Malone Assistant Professor
Department of Electrical and Computer Engineering
Mathematical Institute of Data Science
Center for Imaging Science
Johns Hopkins University Whiting School of Engineering

Dr. Vishal Patel
Associate Professor
Department of Electrical and Computer Engineering
Mathematical Institute of Data Science
Johns Hopkins University Whiting School of Engineering

Dissertation committee members

Dr. Carey E. Priebe
Professor, Primary: Applied Mathematics and Statistics
Mathematical Institute for Data Science
Center for Imaging Science
Human Language Technology Center of Excellence,
Secondary: Computer Science, Electrical and Computer Engineering,
and Biomedical Engineering
Johns Hopkins University Whiting School of Engineering

Dr. Haris Sair
Director of Neuroradiology
Associate Professor of Radiology and Radiological Sciences
Malone Center for Engineering in Healthcare
Johns Hopkins University School of Medicine

Acknowledgments

I am very grateful for the support from a few important people throughout this arduous journey.

First and foremost, I dedicate my thesis and PhD to my late maternal grandmother Pushpam Sivagurunathan, whom I called ammammah. My first teacher, ammammah taught me math and english at grade levels above my own to encourage growth and rigorous academics at a young age. My earliest memory is of ammammah strolling to the local library the summer before kindergarten to read picture books with me. I know she is always watching over me, and I always felt her presence pushing me through tough times. Ammammah, there isn't a day that goes by where I don't think about you and the positive impact you've had on my development.

I am extremely grateful for my advisor Professor Archana Venkataraman. Her rigor, depth, and vast knowledge in the subject matters of mathematical modeling, machine learning, deep learning, probability and statistics, and neuroscience has helped me tremendously to grow into the researcher and engineer that I am today. Archana, it is evident to me that you care about your students both professionally and personally, and thank you for exhibiting patience with me along the way. I will always remember and value the lessons

and critiques that you have so graciously given me, and I will do my best to pass them along as my career develops.

I am indebted to my previous and current labmates Niharika D'Souza, Jeff Craley, Ravi Shankar, Sayan Ghosal, Deeksha Shama and Arunkumar Kannan for all of their support and help throughout these six years. You all have been very selfless regarding helping me with any aspects of my work, whether it be through sitting through countless practice talks, co-authoring papers with me, or our routine coffee breaks to get recharged for more work. Niharika, I would like to especially thank you for the extra time and effort you've put into my professional and technical development. I do not think I would have survived this without you.

I would like to acknowledge some key friends and family who have helped and supported me throughout this journey. To my siblings Naveetha and Naren, thank you for understanding and supporting me throughout the past six years. You are always on my mind and I cherish our kinship. To my friends Adam and Anna, thank you for your support throughout this experience. I am glad our best friendship has lasted since the 6th grade and I hope it lasts for the rest of our lives. To the rest of my friends and family, thank you for the continued support and belief in me that I will make it.

Finally, I would like to thank and acknowledge my thesis committee, Dr. Carey Priebe, Dr. Haris Sair and Dr. Vishal Patel. I have learned a tremendous amount from each of you that ties all of the work presented in this thesis together. Carey, thank you for teaching me statistics and how to critically think about pattern recognition patterns from a mathematical point of view.

Haris, thank you for providing me the knowledge of the clinical side of the problems I aimed to solve in this thesis. And Vishal, thank you for your expertise in deep learning and providing technical insight into the models presented in this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The human brain is both widely studied and still not fully understood in it's function and connection from structure to function. It is malleable and able to adapt to trauma or the presence of a lesion. More so, the human brain and functional organization are very different from person to person, especially when pathology is present. The brain is comprised of functionally connected hubs, that coordinate together to perform complicated tasks. Non-invasive imaging techniques, such as magnetic resonance imaging (MRI), positron emission tomography (PET), computed tomography (CT), and functional MRI (fMRI) can be used to provide a glimpse into the structure and function of the brain [1]. Examples of different analyses that are done with imaging range from segmentation of structures of interest from structural MRI [2] to using statistical methods to identify regions of high or abnormal functional activity from fMRI [3].

Task-based fMRI (t-fMRI) is a type of functional imaging modality where the subject is asked to perform a task, such as moving their fingers or speaking a sentence, while in the fMRI scanner. Task fMRI is used primarily to localize

specific functional networks of interest [4]. Resting-state fMRI (rs-fMRI), on the other hand, is a type of functional imaging modality which captures spontaneous fluctuations in the brain while the subject is at rest, from which functional systems in the brain can be identified [5]. Functional connectivity (FC) analysis makes use of steady state co-activation patterns found in rs-fMRI to identify connectivity hubs in the brain and functional networks of interest [6]. FC analysis has gained a lot of popularity in the past decade as a means for answering clinical questions such as diagnosing certain neurological disorders, such as autism spectrum disorder (ASD) [7], or localizing functional regions of interest, such as the language or motor networks [8]. The most common data structure used to summarize FC is a connectivity graph, where nodes represent brain regions and edges represent the functional connectivity between them [9]. The work in this thesis will be primarily applied to rs-fMRI and rs-fMRI connectivity represented by graph structures.

Due to the advent of more compute power and larger datasets, deep learning has made a substantial impact on rs-fMRI analysis and potential for clinical use of rs-fMRI[10, 11]. Deep learning has the capability of successfully performing tasks such as ASD vs. NC classification [12] or Alzheimer's vs NC classification [13] with higher accuracy than traditional machine learning models. The methods used in this thesis are primarily deep learning based.

1.1 Subject-specific differences in rs-fMRI

The use of rs-fMRI to study the brain has been around since the early 1990's, and analysis has typically been done on a group-level, where data is averaged

across multiple participants and then used for subsequent analysis [14]. However, these studies have failed to identify functional characterization at the individual level. It is understood that every individual has a unique functional organization, as no two brains are the same, and that uniqueness should be taken into account during mathematical modeling. More so, subject-specific approaches are necessary for subjects from aytpical cohorts, whose functional connectivity profile will not fit an expected template [15].

1.1.1 Functional parcellations

A functional parcellation seeks to segment the brain into spatially continuous and temporally homogenous regions of interest (ROI's), and is applied to rs-fMRI as a common preprocessing step for improved SNR and dimensionality reduction in analysis pipelines [16]. However, these group-level parcellations are derived from a healthy cohort, which ignore subject-specfic differences. Moreover, these parcellations are especially troubling for atypical cohorts, such as subjects with neurological disorders or lesions (tumors).

The first part of this thesis will aim to develop automated methods to derive subject-specific parcellations using machine learning and deep learning techniques. Specifically, we will introduce and leverage a Bayesian model with a Markov Random Field (MRF) prior to perform subject-specific refinement on existing group level parcellations, which will be validated on a tumor cohort. We then introduce a novel neural network refinement module that leverages learnable weights in a graph structure for parcellation refinement that is capable of improving upon existing rs-fMRI deep learning analysis

tasks. We validate that method using common rs-fMRI analysis tasks such as predicting ASD, fluid intelligence scores, and language localization.

1.1.2 Dynamic connectivity analysis

Not only do individuals have a unique spatial functional organization, but they also have a unique temporal evolution of connectivity patterns. The sub-field of dynamic functional connectivity (dFC) within rs-fMRI analysis aims to track and identify changes in connectivity patterns throughout the rs-fMRI scan [17]. The underlying hypothesis governing dFC analysis is that functional networks "emerge" or are more strongly connected at different time points throughout the scan. Typically dFC is created with a sliding window technique, where a separate connectivity graph is created from segments of the rs-fMRI data. The models present in this thesis will make use of both static FC and dFC.

1.2 Eloquent cortex localization for brain tumor resection

Neurosurgery resection procedures for brain tumor removal requires immense precision, as the surgeon must remove as much of the lesion as possible while preserving maximal neural functionality. An incorrect incision here can cause severe or even permanent deficits [18, 19]. The eloquent cortex includes regions of the brain that are responsible for motor functionality and language generation and comprehension [20, 21]. The eloquent cortex must be preserved during a neurosurgery to avoid severe cognitive deficits. The gold standard

for mapping the eloquent cortex is invasive electrocortical stimulation (ECS) while the patient is awake [22]. While accurate, ECS can be traumatic for the patients, who must remain functioning during the procedure. Noninvasive t-fMRI is a popular alternative preoperative mapping tool [23, 24]. However, the resulting t-fMRI activations may be unreliable for certain cohorts due to cognitive impairments, speech aphasia, or an inability to follow the task protocol [25].

Using rs-fMRI connectivity to localize the language and motor regions in brain tumor patients is a promising alternative to t-fMRI as a means of circumventing the issues associated with t-fMRI [26]. The next part of this thesis utilizes deep learning models applied to rs-fMRI connectivity to perform eloquent cortex localization, which is validated on both an in-house collected dataset as well as an artificially created tumor dataset from healthy rs-fMRI. The models presented in this part of the thesis will sequentially build on eachother, introducing new deep learning techniques applied to graph structures and then dFC to improve localization performance.

1.3 Epileptogenic zone localization in focal epilepsy patients

Epilepsy is one of the most common neurological disorders in the world, and is characterized by the patient having recurrent abnormal neural discharges that lead to seizures [27]. Focal epilepsy is where the seizures originate from the same region in the brain. Medication is one option for treatment, but for those who are medicaton refractory, surgical resection of the area from

which the seizure originates is a viable option [28]. Identifying the EZ using intracranial electrodes provides accurate monitoring, but is highly invasive and can introduce surgical risks [29]. Presurgical planning procedures for identifying the EZ are most often based on the electroencephalogram (EEG) modality [30]. However presurgical evaluation relying on ictal recordings is a time consuming procedure due to the low frequency of seizures during the recording [31].

Using rs-fMRI connectivity to localize the EZ in focal epilepsy has emerged as an alternative modality. The next part of this thesis develops graph based neural networks to localize the EZ using rs-fMRI connectivity. The models presented will build upon the last, where we introduce novel dFC methods and data augmentation techniques to improve EZ localization as well.

1.4 Noisy labels in neuroimaging

With regard to neuroimage analysis, deep learning methods have already achieved impressive and unprecedented performances, partially due to the growing size and availability of datasets. However, it is very difficult to curate large datasets with reliable labels, as this is a very time consuming task. A solution is to use pre-trained models to perform the labeling themselves, but these models often suffer from large amounts of label noise, or incorrect labels [32]. As discussed, using t-fMRI as ground truth for eloquent cortex localization can be troublesome, and not entirely accurate. Furthermore, it is heavily contingent user specified thresholds of t-fMRI activation maps [33].

The last part of this thesis and future work aims to address the noisy label

issue in the context of rs-fMRI analysis, specifically applied to EZ localization for focal epilepsy patients. Regarding EZ localization, models can be trained using the entire resection area as the EZ ground truth, but it is well known that the true EZ will lie within the resection, as the surgeon will be liberal with resection as to make sure all of the EZ is removed. We introduce a novel deep learning technique that can handle noisy labels in the context of EZ localization. We introduce novel data augmentation techniques and sampling procedures to show robust training against noisy or incorrect labels, and show how the proposed model performs on a real dataset.

1.5 Summary and outline

In summary, rs-fMRI has emerged as a powerful tool to provide insight into the functional organization of the brain. Deep learning models applied to Rs-fMRI connectivity have the ability to aid clinicians in challenging tasks such as eloquent cortex localization for tumor removal procedures, or EZ localization for EZ resection procedures. Finally, we will conclude this thesis with a chapter on ongoing work to tackle the noisy label problem that is present within our studies.

Chapter 2 will introduce relevant background and literature for this thesis. Imaging acquisition and details for structural MRI, t-fMRI, rs-fMRI, and diffusion tensor imaging will be presented. A brief overview of connectivity analysis and ICA will be discussed. An overview of existing methods to perform parcellation refinement, automated eloquent cortex localization, and automated EZ localization will be presented. An overview on commonly used

deep learning architectures will be presented. Finally, a description of each of the datasets used in this thesis will be presented.

Chapter 3 will show the work we have done and published on tackling the subject-specific parcellation refinement problem. The first model presented in chapter 3 will be a bayesian model with an MRF prior for refinement, which was published in the connectomics in neuroimaging (CNI) workshop as a part of the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2018 conference [34]. Then we will conclude the chapter with introducing our deep learning based approach RefineNet, which was presented at MICCAI 2022 [35].

Chapter 4 will describe the work we have published on eloquent cortex localization using a static connectivity graph input. We start with our base graph neural network (GNN) presented at CNI MICCAI 2019 [36], which was extended to a multi-task learning setting and validated on more comprehensive experiments and an augmented dataset in our medical image analysis journal [37].

In chapter 5, we extend our analysis to the dFC case, and present our LSTM based model from the machine learning in clinical neuroimaging (MLCN) workshop from MICCAI 2020 to improve localization using temporal attention [38]. Finally, we conclude this chapter with our work presented at the Information Processing in Medical Imaging (IPMI) 2021 conference, which extends on prior work by adding convolutional based spatial attention models to improve localization performance [39].

Chapter 6 presents the work we have published on EZ localization, which

primarily focuses on using graph convolutional networks. First, we will present our DeepEZ model (published in IEEE transactions in biomedical engineering), which uses anatomically regularized graph convolution networks to perform EZ localization on a small dataset [40]. We extend our journal work to the dFC case, and employ transformer networks alongside data augmentation techniques in the next model we present, which was recently published and presented at the International Symposium for Biomedical Imaging (ISBI) 2023 conference [41].

Chapter 7 presents ongoing and future work within the realm of noisy label issues in our classification (localization) models, specifically regarding epileptogenic zone localization. We have been developing a data-driven approach to identifying label uncertainty which makes use of a shared representation learned from connectivity data. We will present the mathematical model, the deep networks and optimization scheme, and the accompanying experiments , which include localization performance on real focal epilepsy subjects. We plan on submitting this as a new conference paper publication.

Chapter 2

Background

In this chapter, we will go over a comprehensive background for this thesis. First, we will present the acquistion and details on the different neuroimaging modalities that are present in this thesis. Next, we will give a brief overview of functional parcellations and existing methods for creating parcellations. Then we will give an overview of functional connectivity analysis with an emphasis on prior machine learning work in the eloquent cortex and EZ localization fields. Then we will go over the foundations of deep learning and common network architectures as a preliminary to discuss the models present in this thesis. Finally, we give the acquisition, preprocessing, and relevant details for each of the datasets used in this thesis.

2.1 Neuroimaging modalities

Both structural and functional neuroimaging modalities can give complementary and useful information for brain connectivity analysis. Each of the modalities presented in this section will be used in the models presented later

in this thesis, with rs-fMRI connectivity as the model input. Strutural MRI is used to delineate the EZ resection for the focal epilepsy subjects and the tumor boundaries for the tumor subjects. Task fMRI was taken in the tumor cohort to act as ground truth for localizing the motor and language networks. Finally, diffusion tensor imaging from healthy subjects was used in regularizing the EZ localization models.

2.1.1 Structural MRI

Structural MRI is a noninvasive imaging tool that provides information on the physical structures of the brain, where different tissues exhibit different contrasts in the resulting 3D image. The acquisition starts with using a magnetic field to align the water molecules to the same orientation before applying a pulse sequence. Then the scanner applies an excitation pulse (using radio frequency), which tilts these nuclei from their alignment. The nuclei then precess, or return, back to the alignment. Essentially, the time it takes for the nuclei of a certain tissue to precess, or relaxation time, is directly proportional to the intensity for which that tissue shows up in the resulting structural MRI. The final image contrast depends on design parameters such as echo time (TE) and repetition time (TR).

The sub-field of automated segmentation within neuroimage analysis has made large strides with the advent of deep learning [42]. This thesis makes use of manual segmentation of structural MRI to provide important label information for both the brain tumor models and the EZ localization models.

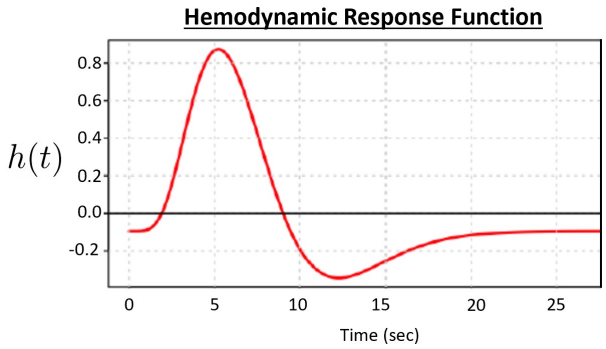

Figure 2.1: The hemodynamic response function.

2.1.2 Functional MRI

As opposed to structural MRI, functional MRI is a noninvasive neuroimaging tool that tracks blood flow throughout the brain over time. Specifically, the protein hemoglobin is tasked with bringing oxygen to neurons during cognitive functioning. When blood is moving towards neurons, the hemoglobin is in a diamagnetic state and when blood is moving away from neurons (after oxygenation) the hemoglobin is in a paramagnetic state. Blood oxygen level-dependent (BOLD) fMRI uses the a T2-weighted protocol to provide contrast in the fMRI image produced. Increases in blood oxygenation levels result in an increased T2 value and therefore a higher intensity in the image [43]. Representation wise-, fMRI is a 4D image, where the first three dimensions represent the voxel spatial coordinates and the last dimension represents time. Typically, there are hundreds of thousands of voxels per scan, which makes this a high-dimensional data structure.

Studies and experiments involving fMRI rely on the fact that blood flow and neuronal activation are coupled. We use the hemodynamic response function (HRF) as a transfer function linking neural activity with the fMRI

signal [44]. Fig.2.1 shows an example of the function. Studies show that the shape of the HRF varies across brain regions and individuals [45]. It is generally difficult to identify the separate contributions of neural and non-neural factors regarding the HRF shape. While fMRI is widely used in research studies, the relationship between the HRF and actual neural activity is not entirely understood.

2.1.2.1 Task-based fMRI

Traditional fMRI experiments use a block design in which the subject is instructed to perform experimental and control tasks in an alternating sequence or 20-40 sec blocks [46]. Signal from hundreds of thousands of voxels are recorded during the t-fMRI scan and the goal of t-fMRI is to localize the specific regions or cognitive systems which are responsible for the experimental protocol administered. Changes in the fMRI related to the experimental and control tasks to identify these regions have been analyzed in different ways such as subtraction, correlation, and time frequency analysis [46]. However, the general linear model (GLM) has been the de-facto standard for identifying regions of high-activity from t-fMRI.

The GLM assumes that each experimental block has a linear contribution to the overall fMRI response. Let $\mathbf{x}_i \in \mathbb{R}^{T\times 1}$ be the time series associated with voxel i. There is a design matrix $\mathbf{D} \in \mathbb{R}^{T\times S}$ for which the experimental protocol is encoded in. As discussed, the HRF is a transfer function that links neural activity with the fMRI signal. This HRF gets convolved with the protocol to obtain the columns of $\mathbf{D}$. The goal of GLM is to solve for the

following linear regression

$$\mathbf{x}_i = \mathbf{D}\beta_i + \epsilon_i \tag{2.1}$$

where ϵ represents a noise term and we are primarily interested in solving for β_i, the activation coefficients assigned to voxel i. We can solve for $\hat{\beta}$ using the least-squares solution $\hat{\beta} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{x}$. The larger the β_i value, the more active that voxel i is in response to the task stimulus. A common technique to identify functional systems from t-fMRI is to overlay the activation coefficients β on the brain and threshold the values at a certain level to only retain the high activations. Later in this thesis, we will make use of this technique to obtain ground truth labels for language and motor networks in the brain tumor dataset.

2.1.2.2 Resting-state fMRI

Unlike t-fMRI, resting-state fMRI (rs-fMRI) is taken with the subject fully and completely at rest, i.e. without an administered task paradigm to follow. Spontaneous fluctuations in the BOLD signal during the rs-fMRI scan are known to be correlated within regions that work together to perform a cognitive task [47, 48]. This effect was first identified in the seminal 1995 paper by Biswal [5], and underpins the core assumptions of many research studies today [49].

Studying regional co-activation patterns found in rs-fMRI is a sub-field called functional connectivity (FC) analysis, which uses the aforementioned assumptions to draw inference for research studies. For example, correlations between the time series values of rs-fMRI can be used to identify brain differences in patients with Autism Spectrum Disorder (ASD) [50] or can used for

classification of the different stages of Alzheimer's Disease [51]. Furthermore, rs-fMRI provides a useful alternative to t-fMRI in identifying functional systems due to the potential downfalls of t-fMRI, such as inability to perform the task. Rs-fMRI is an option for pediatric populations as well.

Each model in this thesis will use rs-fMRI as the primary input. First, we will explore subject-specific differences in functional organization using rs-fMRI connectivity. Then, we will show predictive power rs-fMRI has in identifying the eloquent cortex in brain tumor patients. Then we will discuss rs-fMRI as a modality for which EZ localization is possible, and finally conclude with a study on noisy labels that can likely occur in these studies.

2.1.3 Diffusion tensor imaging

White matter in the brain is composed of bundles of nerve fibers that connect neurons in different brain regions. Diffusion tensor imaging (DTI) is a neuroimaging technique that uses MRI to identify white matter structures in the brain. The diffusion of water molecules goes in the direction along white matter bundles as opposed to across them. DTI uses MRI to leverage this property in its imaging acquisition. It is common to use fiber tracking algorithms to construct maps of the structural pathways associated in the brain. From the clinical point of view, DTI has shown regional differences in certain neurological disorders such as Multiple Sclerosis (MS) [52] and Alzhemier's Disease [53].

Complementary to functional connectivity extracted from rs-fMRI data, DTI provides a structural connectivity profile of the brain, highlighting the

density of white matter bundles that connect different ROIs in the brain. Recent advances in neuroimaging techniques have been making use of a multi-modal approach using both structural and functional connectivity to improve the downstream task. Some examples include using structural and functional connectivity to improve ASD characterization [54] or Alzheimer's Disease [55]. We will use DTI of healthy subjects in the EZ localization models in this thesis to act as an anatomical regularization to our network.

2.2 Functional parcellations

Rs-fMRI suffers from the curse of high-dimensionality, where scans can be on the order of hundreds of thousands of voxels. A parcellation seeks to segment the brain into spatially continuous and functionally synchronous regions of interest (ROI). Due to voxel-level variability, rs-fMRI data is often analyzed at the region level based on a predefined brain parcellation, where each ROI can contain hundreds of voxels [56]. Parcellation construction can occur on the group-level, using many subjects' data, or the individual level. There exists both anatomical parcellations, which segment anatomical structures, and functional parcellations, which seek to use functional imaging to segment functionally cohesive regions [57]. There exist many parcellations in the literature, and there is no one de-facto standard or way of evaluating which parcellation is better than the rest [58].

2.2.1 Group-level approaches

Most parcellation schemes are constructed on group-level averages; here we will describe a few of the many existing parcellations. The commonly used Craddocks functional parcellation is constructed by employing spectral clustering on 41 healthy subjects' rs-fMRI data [59]. The method varies the variable k to obtain parcellations with $100 - 1000$ ROIs. The widely used Brainnetome atlas uses multi-modal data (rs-fMRI and DTI) and employs clustering algorithms on seed based correlation matrices in conjunction with tractography connectivity information on 40 healthy subjects to obtain a parcellation with 246 ROIs [60]. Parcellations can be coarse as well, with fewer ROIs. The Yeo atlas uses a surface registration technique followed by clustering of 1000 subjects' data to obtain both a 7 ROI and 17 ROI atlas [61].

While applying a group-level parcellation to the rs-fMRI as a preprocessing step is the de-facto standard for connectivity analysis, it is well understood that functional landmarks vary from person to person, and therefore these parcellations might not capture the correct boundaries for the entire cohort [62]. Furthermore, group-level derived parcellations are usually construced from healthy subjects' data, which is especially troublesome with lesional (tumor) or other neuroatypical cohorts [34]. For the problems addressed in this thesis, developing subject-specific approaches for parcellation construction would be more beneficial for analysis because we are interested in clinical populations such as brain tumor cohorts and epilepsy cohorts.

2.2.2 Subject-specific approaches

In contrast to group-level approaches, subject-specific approaches aim to segment one subjects' rs-fMRI data at a time. The work in [63] uses a combination of seed region growing alongside spatially constrained hierarchical clustering to obtain a single subject parcellation. The authors validated their method on test-retest reliability of recovered parcellations and comparision with task-based obtained clusters. The work of [64] describes a method to obtain single-subject morphological parcellations. Their method constructs similarity graphs based on morphological indexes and they use test-retest as a validation metric. While subject-specific approaches for parcellation construction might work better for our tasks in theory, one can't draw ROI-level correspondence using a subject-specific approach. Therefore, inference drawn from training models on data using multiple single-subject parcellations isn't feasible.

2.2.2.1 Parcellation refinement

To circumvent the issues assosciated with both group-level and single-subject parcellation construction techniques, parcellation refinement techniques have gained popularity in recent years. A parcellation refinement technique makes use of a group-level derived parcellation, such as the Craddocks atlas, and then reassigns voxel parcel membership on a subject-specific basis. Therefore, there is a group-level concordance between subjects and the boundaries vary from subject to subject to capture individual functional differences. For example, the work of [65] iteratively reassigns voxel membership based on the pearson's

correlation coefficient between the voxel and the mean time series defined by each parcel. However, this method does not ensure spatially continuity in the resulting parcels and requires the user to specify parameters. In the next main chapter of this thesis we develop two different models to perform parcellation refinement.

2.3 Connectivity analysis

Functional communication between brain regions, as highlighted through rs-fMRI, plays a key role in complicated cognitive processes. Therefore, it is important to understand and examine functional connectivity to learn more about individual brain organization. Functional connectivity (FC) is defined as the temporal dependency between spatially remote regions in the brain [66]. Each model in this thesis will make use as functional connectivity as summarized in the form of a connectivity graph from rs-fMRI as the input. More specifically, the models presented will identify key connectivity similarity and differences between individuals to maximize performance on the downstream task.

2.3.1 Functional concordance: seed based analysis

Rs-fMRI studies have shown the presence of spontaneous fluctuations within regions of the brain , usually between the frequency band of $0.01 - 0.1$ Hz [67]. Despite the lack of an experimental task present, these signals have been found to be strongly correlated between certain structures in the brain across different subjects. The de-facto standard for defining connectivity between

brain regions is to take the pearson's correlation coefficient between the time series associated with different voxels in the rs-fMRI scan. Given voxels i and j, with associated time series $\mathbf{x}_i$ and $\mathbf{x}_j$, the correlation coefficient $\rho_{i,j}$ is defined as

$$\rho_{i,j} = \frac{\mathbf{x}_i^T \mathbf{x}_j}{||\mathbf{x}_i|| ||\mathbf{x}_j||} \tag{2.2}$$

Seed based analysis (SBA) is a common rs-fMRI analysis technique in which the connectivity of a different seed, or small region in the brain, with the rest of the brain is assessed [8]. The resulting seed based maps are then thresholded at a user specified threshold before analysis. This method of analysis has been used to compare connectivity across different cohorts [68]. The seed is defined as a spatially continuous group of voxels that is chosen *a priori* and is usually determined from pre-existing domain knowledge. Seed based analysis has been useful for identifying brain systems reliably across subjects or cohorts, specifically for healthy subjects and large commonly found networks such as the visual network [69] or the motor network [70].

Despite its success, seed based analysis is not suited for the goals of the models presented in this thesis. First, SBA heavily relies on accurate *a priori* knowledge of seed placement, which is very troublesome for a brain tumor cohort, where the size and location of the tumor will disrupt the neuronal connections in those regions. In general, SBA will work better for healthy cohorts. Furthermore, the user defined threshold proves to be troublesome and also makes these methods not fully automated [71]

2.3.2 ICA for rs-fMRI analysis

The independent component analysis (ICA) method for defining functional connectivity emerged as a very popular and useful tool in rs-fMRI analysis [72]. ICA is a method for separating a multivariate signal into additive sub components, specifically by assuming statistical independence between the extracted sub components. Both subject-level and group-level ICA have shown success in network identification for rs-fMI studies [73].

Fig. 2.2 shows a visual representation of spatial components extracted from rs-fMRI data using both individual and group ICA [72]. The spatial components extracted from ICA applied to rs-fMRI are assumed to be either functional sub-systems or noise components. For example, the spatial map on the bottom right of Fig. 2.2 represents the motor network. Group ICA applied to rs-fMRI has made strides in various sub-fields of rs-fMRI analysis, such as characterizing ASD [74] or Alzheimer's disease [75].

2.3.2.1 Eloquent cortex localization prior work

Here we will go over non-deep learning techniques that have been applied to rs-fMRI of tumor patients to identify the language and motor networks. In the SBA domain, the work of [76] uses lateralized anatomical seeds to localize bilateral activations on the supplementary motor area in tumor patients. The drawback from this method is that it requires an expert to manually select the seed location, which is time consuming and can vary greatly from patient to patient due to tumor size and location.

Using ICA on rs-fMRI has emerged as a potential way of identifying

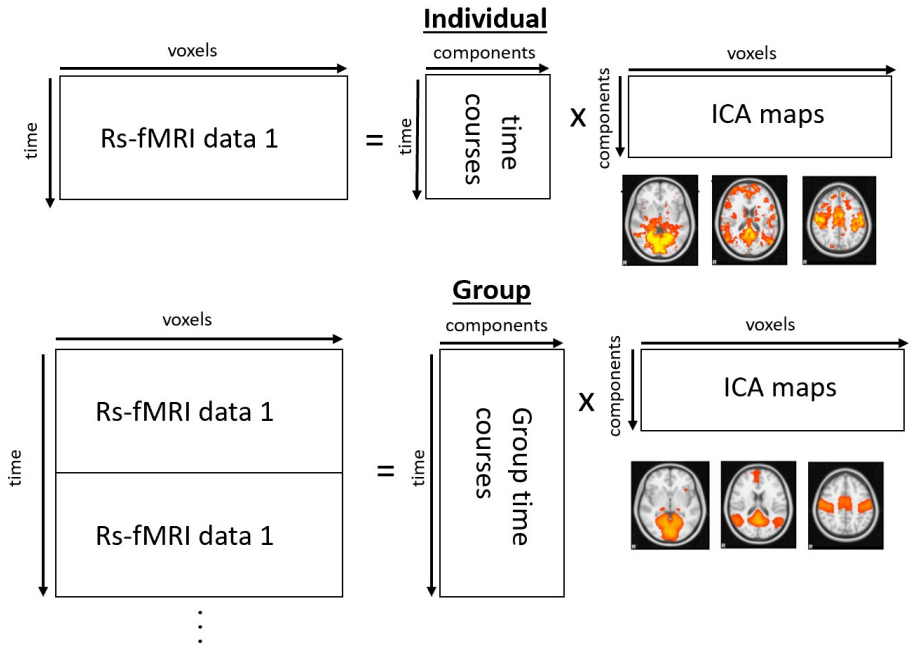

Figure 2.2: Visual representation of individual (top) and group (bottom) ICA applied to rs-fMRI data.

language and motor regions in brain tumor patients [77]. The work of [33] describes a method that uses group ICA alongside manual thresholding and selection on tumor patients to identify language network components. The drawback to these methods are that they are not fully automated, and require an expert to manually set a threshold for the ICA and select the components.

2.3.2.2 EZ localization prior work

Here we will go over machine learning techniques applied to rs-fMRI of epilepsy patients to perform EZ localization. The earliest work of [78] computed network theoretic measures from rs-fMRI data. The authors used outlying values in these network measures, as compared to a healthy cohort,

to define the EZ for each epilepsy patient. While promising on a small validation dataset, similar to the eloquent cortex prior work, the performance requires careful tuning of different threshold values. The follow-up work of [79] proposed a hierarchical Bayesian model that inferred patient-specific hubs of abnormal connectivity. Both of these studies rely on comparison with a normative cohort and the [79] paper only validated with six subjects.

The work of [80] takes a different approach by first running independent component analysis (ICA) and then constructing a set of rules (e.g., asymmetry, power spectrum) to select the components associated with the EZ. While the authors demonstrate highly promising localization performance, they rely on visual inspection to select between candidate EZ components. Thus, a careful read of the method suggests that it is not fully automated. Specifically, the authors rely on visual inspection to select between candidate EZ components. Finally, the work of [31] also runs ICA on the rs-fMRI data and extracts a set of hand-crafted features from the components. In this case, the authors employ a support vector machine classifier to automatically learn which components are associated with the EZ. While the authors demonstrate good localization performance on some patients, the reported sensitivity is low for others.

2.3.3 Connectivity graphs

A connectivity graph summarizes the functional co-activation patterns found in an rs-fMRI scan. Fig. 2.3 shows the workflow to obtaining a connectivity graph. A parcellation is applied to the rs-fMRI data and typically the time series values defined by each ROI is averaged to get one mean time course

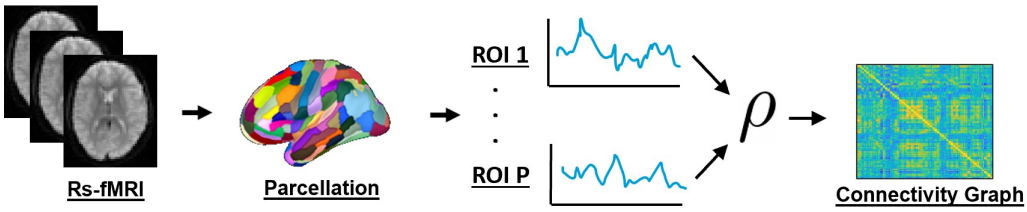

Figure 2.3: The pipeline for obtaining a functional connecivity graph starts with a parcellation with P ROIs. The average time series for each ROI is taken and typically the Pearson's correlation coefficient between regions is used to populate the $P \times P$ connectivity graph.

that summarizes the fMRI at that location. Typically the Pearson's correlation coefficient between each pair of mean time series is used as the entries of the graph. However, a similarity graph containing just entries between $0 - 1$ is also typically used by taking a positive transformation of the Pearson's correlation coefficient [36].

Many rs-fMRI analysis works use the connectivity graph as an input data structure. From here, one can extract local and global network properties associated with each node within the brain, as defined by the parcellation chosen. For example, the work of [81] uses graph theoretic features such as node degree, betweeness centrality and eigenvector centrality to characterize brain networks. Features defined by the graph can also be input to common machine learning techniques. For example, the work of [82] uses a support vector machine on graph theoretic features from rs-fMRI to classify Parkinson's Disease. In this thesis, we will use the entire connectivity graph as the input to our models.

2.3.4 Dynamic functional connectivity

There exists growing evidence that whole-brain functional connectivity changes over time, and that the brain goes between different intrinsic states of connectivity throughout the rs-fMRI scan. The field of dynamic functional connectivity (dFC) aims to quantify and track these changes and improve upon common rs-fMRI analysis tasks using dFC. For example, the work in [83] uses k-means clustering on dFC matrices to identify differences among schizophrenic, healthy, and bipolar subjects. The work in [84] analyzes multiple networks using different time scales combined with a support vector machine to perform ASD classification. In this thesis, we will leverage dFC to improve localization for both the eloquent cortex and EZ work.

One way to obtain dFC input graphs is to use the sliding window technique. Here, multiple connectivity graphs are computed using the Pearson's correlation coefficient on windowed sub-segments of the rs-fMRI scan. While Fig. 2.3 shows the process of getting one static connectivity graph, the sliding window technique obtains multiple connectivity graphs for one rs-fMRI scan. Another way of identifying dFC is to use dynamic conditional correlation, which introduces a time-varying matrix estimation problem to model the evolution of dynamics within the scan [85]. The work in [86] introduces an approach called dynamic sparse connectivity patterns, which leverages matrix factorization and graphical lasso to obtain dFC. While many methods for obtaining dFC exist, the sliding window is the most commonly used in the literature.

2.4 Deep learning

The development of the perceptron model and artificial neural networks (ANNs) has been around since 1959 [87], but deep learning has only taken a front and center presence in machine learning in the past two decades, likely due to the increases in compute power and size and availability of datasets. Deep learning has shown large performance gains in the fields of image classification [88] and text classification [89]. As opposed to traditional machine learning classifiers, deep learning models seek to learn the optimal classification function in a completely data-driven manner. The underlying idea is that deep learning models do not need hand-crafted feature selection, as opposed to traditional machine learning models.

Deep learning models extract higher-order representations of the data via different layers, where each layer is parameterized by a set of weights. Ultimately, the goal of deep learning is to approximate a function $y = f^*(x;\theta)$ by learning the optimal θ^* using a technique called backpropagation. The different layers represent different sub-functions within the overall function. A nonlinear activation function is applied in between each layer so that the network can learn nonlinear relationships in the training data, thus expanding their representational power. There are three main categories of deep learning layers that are all used to process different types of data efficiently. These are fully-connected layers, convolutional layers and recurrent layers. In this section, we will mathematically describe these types of layers and conclude with a description of optimization and training of neural networks.

2.4.1 Fully-connected neural networks

The fully-connected neural network (also called the multi-layer perceptron, linear layer, or artificial neural network) was first introduced in [87] and is the most simple type of neural network layer. The perceptron model was first introduced as a way to model how neurons fire, and thus the term neural network was introduced. The fully-connected layer connects every input neuron to every output neuron in a linear fashion. The layer calculates the sum of products of the inputs and their corresponding weights. Mathematically, let $\mathbf{h}^{(l-1)}$ be the previous layer's activations, the forward propagation equation for a fully-connected layer is

$$\mathbf{h}^{(l)} = \sigma(\mathbf{h}^{(l-1)}\mathbf{W}_l + b_l) \tag{2.3}$$

where $\mathbf{W}_l$ is a matrix of learned weights, b_l is a learned bias, and σ is a non-linear activation function. Typically, there is one input layer, one output layer, and multiple intermediate, or hidden layers, in ANNs. Fully-connected layers are also usually the final layer of a classification model, even if the rest of the network uses different types of layers. Furthermore, the number of neurons in an ANN is an important hyper-parameter of choice because it defines the complexity of the function that the ANN can approximate [90].

2.4.1.1 Activation functions

The power of neural networks is in the fact that they act as a universal function approximator, capable of approximating any function if the network has enough depth [91]. The key to this finding is the application of an activation

function after the operation of a neural network layer. Specifically, the activation function is nonlinear, which extends the learning capabilities of neural networks from just linear transformations of the data to any arbitrary function. Therefore, the application of an activation function is crucial in understanding why neural networks are such powerful representation learning tools.

Here we will go over four commonly used activation functions. The first popular activation function is the sigmoid function, which restricts the output to be between $0 - 1$. The sigmoid function is not widely used in the current literature due to the fact that it has shown to slow down training because of its non-zero mean [92]. The hyperbolic tangent (tanh) activation function aims to mitigate this by outputting values between -1 and 1. However, both of these activation functions suffer from potential saturation during training. Saturation, or the phenomenon of a neuron predominantly outputting values close to the boundary values of an activation function, makes gradient descent slow and training inefficient. Specifically, the derivative values are very small near the asymptotes of the activation functions [93]. The rectified linear unit (ReLU) activation function emerged as an alternative to mitigate these ill-effects associated with sigmoid and tanh. $\text{ReLU}(x)$ simply outputs x if $x > 0$ and outputs 0 otherwise. Therefore, there is no clipping at the upper bound of the activation function. Another advantage of the ReLU function is representational sparsity, or the fact that it can output a true 0 value. However, there is a potential disadvantage for having gradients exactly equal to 0 during training and that is the unit may never activate and not get out of this inactive regime. The leaky ReLU activation aims to mitigate issues associate with

ReLU by allowing for a small, non-zero gradient when $x < 0$. The leaky ReLU activation sacrifices hard-zero sparsity for more robust training.

$$\begin{aligned} \text{sigmoid}(x) &= \frac{1}{1+e^{-x}} \\ \tanh(x) &= \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \\ \text{ReLU}(x) &= \max(0,x) \end{aligned} \tag{2.4}$$

$$\text{LeakyReLU}(x) = \max(\alpha x, x) \quad 0 < \alpha < 1$$

2.4.2 Convolutional neural networks

One large limitation of ANNs is they tend to struggle with the complexity associated with processing color images or videos. The next main branch of neural network architectures that we will discuss is the convolutional neural network, which uses filters (or kernels) of varying sizes to process data and perform feature extraction. The most common CNN layer is 2D and is used to process images, but it should be noted that both 1D and 3D CNN layers exist as well, which are capable of processing time series data or videos. For simplicity, we will just include the 2D filter equation. Suppose we have a $N \times N$ square neuron layer which is followed by the convolutional layer. Let $\sigma()$ be the activation function and we have filter $\mathbf{W} \in \mathbb{R}^{m \times m}$ and bias b, the resulting output of the layer will be of size $(N-m+1) \times (N-m+1)$ with the following forward propagation equation.

$$x_{ij}^{(l)} = \sigma\Big(\sum_{c=0}^{(m-1)} \sum_{d=0}^{(m-1)} \mathbf{W}_{cd} x_{(i+c)(j+d)}^{(l-1)} + b \Big) \tag{2.5}$$

It is important to note that the equation shown was just for one filter when usually convolutional networks have multiple filters associated with one layer. As shown, the filter takes direct spatial neighborhood information to perform feature extraction while simultaneously performing downsampling or dimensionality reduction between layers. For this reason, CNNs typically work well on data that have spatially continuous features, such as images or videos. CNNs learn representations which are symmetric to spatial and temporal transformations of the data. CNNs aim to mimic the visual system in humans, and have made vast improvements in image classification and computer vision.

2.4.3 Graph convolutional networks

Graphs arise in many real-world applications, such as social network analysis, traffic prediction, and brain connectivity analysis [94]. While traditional CNNs work well for data with spatially continuous features, such as images, they fail to generalize well to graph structures due to the non-euclidean and complex nature of graph data. In this section, we present the spatial graph convolutional network, which work on a local neighborhood of nodes. Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ be a binary adjacency graph which has value $\mathbf{A}_{ij} = 1$ when nodes i and j are connected and 0 otherwise. The layer propagation rule for spatial graph convolution is as follows, where $\mathbf{W}$ and b represents the learnable weights and bias:

$$\mathbf{H}^{(l)} = \sigma\big(\mathbf{A}\mathbf{H}^{(l-1)}\mathbf{W}^{l} + b_l\big). \tag{2.6}$$

The pre-multiplication of the graph structure acts as a filtering method to use the fact that connected nodes should be similar in representation for downstream tasks. It is common to replace the adjacency matrix $\mathbf{A}$ with the graph laplacian $\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ where $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$ as a normalized filter. We will use these formulations in the models presented later in this thesis.

2.4.4 Recurrent neural networks

We have discussed the ANN, CNN, and GCN, which all are powerful tools for extracting useful information from data that assume a independent identically distributed (IID) nature with no sequential relationship. However, there exist many data structures that do have dependency in a sequential manner, such as time series data from fMRI or stock market data. Recurrent neural networks (RNNs) are deep learning architectures that are capable of processing sequential data. While an ANN can only map input to output vectors, an RNN can map from the entire history of inputs to an output [95]. Furthermore, it has been shown that RNNs can learn universally arbitrary mappings from input to output sequences [96].

A commonly used RNN known as the Long Short Term Memory (LSTM) network processes sequential data and has been widely used in the fields of rs-fMRI analysis [97], speech [98], and stock market analysis [99]. The LSTM network has input, output, and forget gates alongside a cell state which tracks information throughout the entire network [100]. Let x_t be the input data at time t, h_{t-1} be the previous hidden state, c_{t-1} be the previous cell state. Let i_t, f_t, and o_t be the input, forget, and output values, which are parameterized by

the weights $\mathbf{W}_i, \mathbf{U}_i, \mathbf{W}_f, \mathbf{U}_f, \mathbf{W}_o, \mathbf{U}_o$, each with bias b_i, b_f, b_o. Let $\mathbf{W}_c, b_c$ be the weights for the cell gate, and let $\sigma()$ be the sigmoid function. The governing equations for the LSTM are

$$\begin{aligned} f_t &= \sigma(\mathbf{W}_f x_t + \mathbf{U}_f h_{t-1} + b_f) \\ i_t &= \sigma(\mathbf{W}_i x_t + \mathbf{U}_i h_{t-1} + b_i) \\ o_t &= \sigma(\mathbf{W}_o x_t + \mathbf{U}_o h_{t-1} + b_o) \\ c'_t &= \tanh(\mathbf{W}_c x_t + \mathbf{U}_c h_{t-1} + b_c) \\ c_t &= f_t c_{t-1} + i_t c'_t \\ h_t &= o_t \tanh(c_t). \end{aligned} \tag{2.7}$$

We can see that the input and forget values determine how much of the last time point's cell state we keep. The cell state gets passed through the hyperbolic tangent activation function and then multiplied by the output value to get the new hidden state h_t. The LSTM alleviates vanishing gradient issues by propagating c_t throughout the network. However, the LSTM still does suffer from very long-term dependencies in the input sequence.

2.4.5 Attention networks in DL

Attention models have gained popularity in deep learning as a way to improve sequence modelling for various tasks, allowing for modelling of dependencies without regard to their distance or magnitude in the input or output sequences [101]. Attention networks have contributed to impressive results in

neural machine translation [102], image captioning [103], and speech recognition [104]. A standard neural network consists of a series of non-linear transformation layers, where each layer outputs a fixed-dimensional hidden representation. An attention network maintains a set of hidden representations that scale with the size of the input / source signal.

Formally, let $\mathbf{x} = [\mathbf{x_1}, \cdots, \mathbf{x_n}]$ represent a sequence of inputs, let q be a query, and z be a categorical latent random variable with space $\{1, \cdots, n\}$. Our goal is to produce a refined context c based on x, q. We define the attention distribution as $z \sim p(z|x, q)$ where the context c over a sequence x is defined as

$$\mathbf{c} = \mathbb{E}_{z \sim p(z|x,q)}[f(x, z)] \tag{2.8}$$

where $f(x, z)$ is an annotation function. The vanilla attention distribution p is simply $p(z = i|x, q) = \text{Softmax}(\theta_i)$, where θ_i is a parameterization which is typically from a neural network, i.e. $\theta_i = \text{MLP}([\mathbf{x}_i; q]$. The context is then computed with the simple sum

$$\mathbf{c} = \mathbb{E}_{z \sim p(z|x,q)}[f(x, z)] = \sum_{i=1}^{N} p(z = i|x, q)\mathbf{x}_i \tag{2.9}$$

We will leverage the principles of attention networks to improve our localization throughout the models presented in this thesis. We will focus on attention that works on the temporal scale as well as spatial scale.

2.4.6 Transformer networks

One issue with the LSTM is increased training times depending on the length of the sequence. The last network we will discuss is the transformer network,

which has recently been developed as a model that can encode sequential relationships with just a cascade of feedforward networks [101]. Therefore, transformers are less computationally complex during forward and backward propagation than the LSTM model. The transformer processes an entire sequence (for example, a sentence) at once as opposed to the LSTM (which processes word by word). Therefore, the transformer does not suffer from long-term dependencies in the same way the LSTM does.

There have been many published works using the transformer network. ChatGPT has emerged as a powerful NLP tool that is capable of auto-completing sentences, writing paragraphs, and even code [105]. The image transformer uses transformers to generate images or perform image in painting [106]. The work in [107] shows a graph transformer network applied to rs-fMRI data to improve Alzheimer's Disease.

The key behind the transformer's success is the use of multi-headed self-attention (MHA) mechanisms. Scaled dot-product attention computes weights between a set of queries Q and keys K. To obtain attention, a softmax operation is used such that they sum to 1 and multiplied with a set of values V. The attention equation is as follows

$$\text{Attn}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{2.10}$$

where d is the dimension of K. The outer product QK^T represents the similarity, and therefore the output of an attention layer is a linear combination of V. Multi-headed attention extends the attention layer to h different heads. For each head i, Q , K, and V are multiplied by learnable weight matrices W_i^Q, W_i^K,

and W_i^V. So $\text{head}_i = \text{Attn}(QW_i^Q, KW_i^K, VW_i^V)$. Then, a final weight matrix W^O is computed to combine the individual heads.

$$\text{MultiHead}(Q, K, V) = \text{Concatenate}(\text{head}_1, \cdots, \text{head}_h)W^O \tag{2.11}$$

The overall architecture of the transformer uses an encoder-decoder structure alongside MHA and feedforward (FF) networks. The encoder layers have sublayers that use residual connections followed by a MHA layer and FF network. The decoder follows a similar structure but adds an intermediate sub-layer that compares the output to the encoder output. Let $\mathbf{X}$ be the input, and FF denote feedforward network, the encoder formulation is as follows

$$\mathbf{o} = \text{FF}\Big(\text{MHA}(\mathbf{X}) + \mathbf{X}\Big) + \text{MHA}(\mathbf{X}) + \mathbf{X} \tag{2.12}$$

where we can see the residual connections alongside FF networks. We will use the encoder part of the transformer to process dynamic functional connectivity data later in this thesis.

2.4.7 Optimization and training

Developing robust and accurate optimization techniques for deep networks is a challenging sub-field within the literature. Back propagation is the standard for training neural networks, as the gradients of the parameters can be computed and updated in one backward pass of the network [108]. Gradient descent was the first optimization technique used, which evolved into using stochastic gradient descent (gradient descent using minibatches of the data)

to ease training and prevent overfitting [109]. The de-facto standard for optimizer choice is to use the Adam optimization algorithm, which uses first and second order gradients alongside correction terms during backpropagation [110]. Other hyperparameters involved in training deep networks include the learning rate, number of epochs (or time steps) to perform training for, and batch size.

Many regularization techniques have been proposed to improve training, prevent overfitting to a training set, and improve efficiency of deep networks. For example, dropout is a method where a random subset of nodes are not considered during backpropagation per epoch [111]. Hyperparameter tuning is a commonly faced problem in deep learning optimization. It is common to use a separate validation set of the data (separate from train or test) to tune the hyperparameters to prevent overfitting. With increasing numbers of parameters, one issue deep networks have is lack of sufficient training data to learn from. The use of data augmentation techniques to provide a richer, more comprehensive set of training data is a large sub-field within deep learning [112]. We will make use of data augmentation techniques in this thesis to improve the generalization power of the networks presented.

2.4.7.1 Loss functions

The network first calculates a loss during the forward pass before backpropagation occurs during gradient updates. The choice of loss function will be essential in how the overall network learns from the data. The network engineer will have many options for loss functions for networks that perform the

same task. For example, regression tasks can use either the mean square error (MSE) or the mean absolute error (MAE) loss function. The dice loss function is commonly used for medical image segmentation tasks [113]. The cross entropy loss function is the de-facto standard for classification tasks [114]. However, some classification tasks might have a large class imbalance, such as localization tasks [36], for which the network might not identify the class of interest while just assigning every prediction to the larger class. One technique to mitigate this is to use a weighted cross-entropy term, or the recently developed focal loss, which weights predictions different based on class size [115].

One can regularize deep networks, or encourage a desired property to be learned, by adding terms to the loss function. For example, one can add an L-2 penalty on the weights to encourage stability during training [116]. In this thesis, we will present novel loss functions which capture biological nuances in the problems that we are trying to solve.

2.4.8 Deep learning for rs-fMRI analysis

With its capacity for generalization and learning, deep learning has recently dominated the field of rs-fMRI analysis, especially due to data-sharing and larger publicly available neuroimaging datasets. To date, nearly all types of deep learning architectures have been used in rs-fMRI analysis [117]. Here we will go over a brief review of how each different network (ANN, CNN, GCN, LSTM, transformer) has been applied to rs-fMRI for certain application domains.

The work in [118] uses fully-connected ANN layers to learn mappings

from rs-fMRI connectivity to ASD labels. Their network uses a simple 3 layer ANN to map whole-brain rs-fMRI correlation values to disease label. Similarly, the work in [119] employs weight sparsity alongside ANNs to classify Schizophrenic patients from healthy controls. Their network includes an autoencoder (deep network which reconstructs its input) comprised of only fully-connected layers.

CNNs have shown promising results for rs-fMRI analysis, due to the spatially continuous features present in neuroimaging data. The work of [120] uses 3D CNNs alongside ensemble prediction with a stochastic parcellation scheme (data augmentation) to improve upon ASD vs healthy classification and age prediction. Combining relevant machine learning techniques alongside CNNs have been explored for rs-fMRI connectivity analysis as well. The work in [121] used a modified version of VGGnet (3DCNNs) and ICA derived features to perform schizophrenia vs healthy classification. The work we present on eloquent cortex localization will build off a base CNN model applied to connectivity graphs.

Since using connectivity features summarized in the form of a connectivity graph is the de-facto standard for rs-fMRI connectivity analysis, it is natural to use a graph convolution network for feature extraction and downstream task analysis. The work in [122] uses the K-nearest neighbor algorithm on the FC data to define a graph structure and is combined with a recurrent network structure to identify ASD vs. controls. The work in [123] develops a spatio-temporal graph convolution operation, which operates on spatio-temporal neighborhoods within the graph, taking the entire 4D structure of

the rs-fMRI into account. They validate their method on identifying sex and age differences in the HCP and NCANDA cohort. The work we present on EZ localization will involve graph convolutional networks.

As a 4D modality that contains time-series data, rs-fMRI is a natural candidate for sequential neural networks such as the LSTM or the transformer network. The work from [97] is one of the first networks to use an LSTM in processing rs-fMRI data, where the authors use the time series data as an input to the LSTM followed by mean pooling from each time step's prediction to classify ASD vs controls. The work in [124] leverages both spatial and temporal properites in rs-fMRI by using 3DCNNs and an LSTM network to perform ASD vs controls classification. The work in [125] uses pre-training techniques with transformer networks for brain network classification. Their method uses the encoder part of a transformer applied to connectivity features alongside an ANN for classification. We will discuss how both LSTM and transformer networks can be used as a temporal attention mechanism for dynamic connectivity analysis later in this thesis.

2.5 Datasets

In this section, we will report the datasets that we used for our experiments in this thesis. For each dataset, we will discuss the modalities present, the number of subjects, acquisition protocols, and preprocessing. It is important to note that different works in this thesis may use different subsets of the data that we have access to, due to different subjects having different modalities present and updates to the publicly available dataset used. We will present the

datasets in the order in which we will discuss them in the thesis, starting with the datasets we used for our parcellation refinement validation, following with the eloquent cortex work, the EZ localization work, and then ending with the publicly available human connectome project (HCP) dataset and how we derived various augmented datasets from HCP for certain experiments and training procedures.

2.5.1 ABIDE II

One of the datasets we use to validate our parcellation refinement model is the Autism Brain Imaging Data Exchange (ABIDE) II dataset. The dataset contains 233 subjects (131 ASD, 102 NC) from the Autism Brain Imaging Data Exchange (ABIDE) II dataset [126]. The rs-fMRI data was acquired across six different sites and preprocessed using the Configurable Pipeline for Analysis of Connectomes (CPAC) toolbox [127]. More processing details on ABIDE II can be found in [126].

2.5.2 JHH brain tumor dataset

Our tumor cohort consists of 62 patients who underwent presurgical fMRI at the Johns Hopkins Hospital (JHH). The data was obtained using a 3.0 T Siemens Trio Tim system. Structural images were acquired via an MPRAGE sequence (TR = 2300 ms, TI = 900 ms, TE = 3.5 ms, flip angle = 9°, FOV = 24 cm, acquisition matrix = 256 $\times$ 256 $\times$ 176, slice thickness = 1 mm). Functional BOLD images were acquired using 2D gradient echo-planar imaging (TR = 2000 ms, TE = 30 ms, flip angle = 9°, FOV = 24 cm, acquisition matrix = 64 $\times$

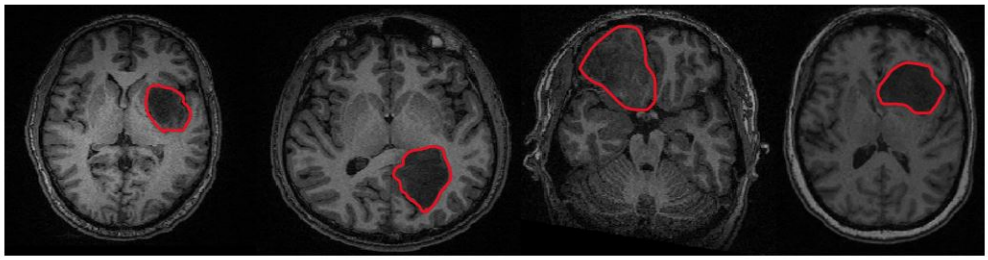

Figure 2.4: Structural T1 MRI of four separate tumor subjects with tumor outlined in red.

64 $\times$ 33, slice thickness = 4 mm, slice gap = 1 mm, interleaved acquisition). A more detailed description of the participants, the task paradigms, and acquisition protocol can be found in [33].

The models present in this thesis make use of rs-fMRI, structural MRI, and t-fMRI of brain tumor patients. The structural MRI was used for manual tumor segmentation via the MIPAV package [128]. The segmentations were performed by a medical fellow and confirmed with an expert neuroradiologist. Fig. 2.4 illustrates structural the T1 MRI of four patients to motivate the heterogeniety in tumor size and location. T-fMRI data was acquired for all patients as part of the presurgical workup. In this work, t-fMRI is used to derive the ground truth eloquent class labels using the General Linear Model (GLM) implemented in SPM-8 [129]. The resulting activation maps were manually thresholded on a patient-specific basis and confirmed by an expert neuroradiologist. The t-fMRI is only used during the training phase of the model. Three motor task paradigms (finger tapping, tongue moving, foot tapping) were used to target specific locations of the motor homonculus [130]. Fig. 2.5 (**L**) shows the various sub-networks of interest for a single patient. Likewise, two language paradigms, sentence completion (SC) and silent word

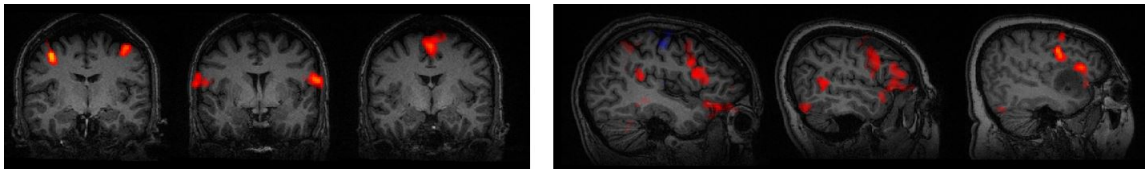

Figure 2.5: **L**: The tongue, finger, and foot sub-networks for one patient. **R**: The language network for three separate patients. The language network boundaries are very variable from patient to patient.

generation (SWG), were performed. These language tasks are designed to target both primary and secondary regions in the brain responsible for language generation [131, 132]. For each patient, instructions and practice sessions were provided. During acquisition, real-time fMRI maps for each task were monitored by the neuroradiologist to assess for global data quality; any task performance deemed suboptimal due to motion-related or other artifact was repeated. Since the t-fMRI was acquired as part of routine clinical care, not all patients performed each task. Finally, our cohort has 57 patients with left-hemisphere language networks and 5 patients with bilateral language networks. Fig. 2.5 (**R**) illustrates the high anatomical variability in language regions, especially due to tumor presence.

Rs-fMRI was acquired while subjects were awake but passive in the scanner. The rs-fMRI data was preprocessed using SPM-8. The steps include slice timing correction, motion correction and registration to the MNI-152 template. The data was linearly detrended and physiological nuisance regression was performed using the CompCorr method [133]. The data was bandpass filtered from 0.01 to 0.1 Hz, and spatially smoothed with a 6 mm FWHM Gaussian kernel. Finally, images found to exceed the default noise threshold by the ArtRepair toolbox [134] were removed (scrubbed) from the rs-fMRI volumes.

Table 2.1: Patient, tumor and t-fMRI information for the JHH cohort.

Age			38 ± 6.3
Sex (M,F)			$37, 25$
Tumor location (lobe)		**Hemisphere**	
Frontal	21	Left	35
Parietal	18	Right	20
Temporal	17	Both	7
Occipital	6		
Volume (x1000)mm^3		**WHO grade**	
<35	21	**1**	14
35-70	28	**2**	27
70-100	8	**3**	13
>100	5	**4**	8
Task protocol	**Subjects**		
Language	62		
Finger	38		
Tongue	41		
Foot	18		

Confounders such as tumor size and handedness are intrinsically tied within the models presented in this thesis, as handedness relates to laterality of language (e.g., we have 57 unilateral and 5 bilateral language subjects), and the tumor is explicitly modelled within our similarity graph. Table 2.1 presents information for the JHH cohort, where we report the number of patients that performed each task, the tumor grade and size, and demographics.

2.5.3 UW pediatric focal epilepsy dataset

Our EZ dataset consists of preoperative functional and postoperative structural MRI scans from 14 pediatric subjects with focal epilepsy that underwent a EZ resection procedure at UW Madison. The MRI data was acquired as a

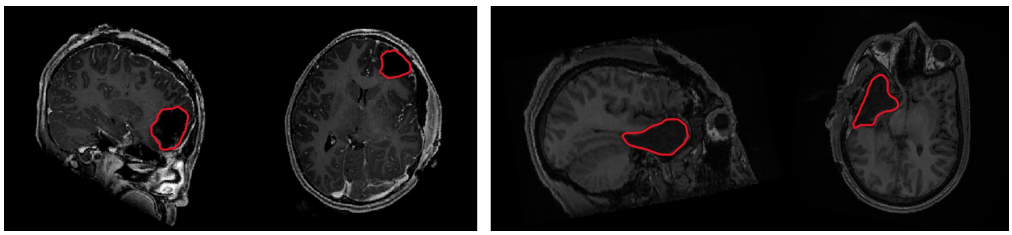

Figure 2.6: From (**L-R**), post resection structural MRI scans of two separate patients. We use the resection boundary, delineated via the red lines above, to derive pseudo ground truth labels for the EZ during training and testing.

part of standard care on either a GE 1.5T or a GE 3T Signa scanner. This study was approved by the University of Wisconsin-Madison Institutional Review Board under protocol 2019-1265 (approved Feb 2020).

Preoperative Rs-MRI (rs-fMRI) data was acquired using an echo planar imaging sequence (EPI, TR $=$ 802 ms, TE $=$ 33.5 ms, flip angle $=$ 50°, FOV $=$ 20.8 cm, res $=$ 2 mm isotropic). The data was preprocessed using the CPAC pipeline [127], which includes slice time correction, motion correction, nuisance signal regression, band-pass filtering $(0.01 - 0.1Hz)$, and registration to the MNI template. For each EZ localization work, we use the Brainnetomme parcellation [60] to define $N = 246$ cortical and subcortical regions for our analysis. We chose this atlas due to its fine spatial resolution and symmetric region definitions.

Diffusion MRI (d-MRI) was also acquired for each patient and integrated into one of the baseline algorithms we compare against for evaluation. D-MRI was collected on a 3T GE scanner (TR$=$ 7000ms, TE$=$ 82.4ms,res$=$ 1.5 mm isotropic, b-value $=$ 1000).

Postoperative T1-weighted structural images were acquired using a three-dimensional gradient-echo pulse sequence (MPRAGE, TR $=$ 604 ms, TE

$= 2.516$ ms, flip angle $= 8°$, FOV $= 25.6$ cm, res $= 0.8$ mm isotropic). After skull stripping, we use affine registration to align the T1 data for each patient to the MNI space. All registrations were visually inspected for quality assurance. We manually delineate the resection zone and use this boundary to define pseudo ground truth EZ labels for training and evaluation after applying the Brainnetomme atlas. Fig. 2.6 depicts two examples of the post-operative T1 images, where the resection is marked by red arrows. Finally, Table 2.2 reports the age, gender, EZ location, scanner used to acquire the rs-fMRI data for each patient and patient outcome using the Engel [135] and ILAE scale [136]. As seen, our epilepsy cohort is highly heterogeneous and every patient experienced reduced seizures after surgery, with the majority being completely seizure free.

2.5.4 The Human Connectome Project

The Human Connectome Project (HCP) is a comprehensive publicly available dataset that contains a substantial amount of neuroimaging data from hundreds of healthy participants. Rs-fMRI, various language and motor t-fMRI, and DTI for healthy subjects are all available. We use different subsets of the HCP dataset for different validation experiments in this thesis. First we present the HCP dataset used as part of our refinement models validation. Then we present a synthetic dataset where we simulate fake tumors in healthy connectomes to provide as an additional eloquent cortex localization dataset. Then we present a synthetic dataset where we simulate the EZ in healthy connectomes to provide a larger EZ cohort, which we use in the noisy label work

Table 2.2: Demographic information, EZ location, and scanner type, and outcome for each patient in the UW Madison dataset.

Subject	Age	Location	Scanner	Outcome
1	14	left anterior and medial parietal	3T	Engel IA, ILAE 1
2	17	left anterior temporal lobe	3T	Engel IA, ILAE 1
3	10	left frontal pole	3T	Engel IA, ILAE 1
4	11	right temporal lobe	3T	Engel IA, ILAE 1
5	15	right anterior frontal/amygdala	3T	Engel IIIA, ILAE 4
6	11	left inferior frontal operculum and insula	3T	Engel IA, ILAE 1
7	13	right middle frontal region	3T	Engel IIIA, ILAE 2
8	14	right posterior temporal and parietal region	1.5T	Engel IA, ILAE 1
9	15	right anterior precentral gyrus	1.5T	Engel IIIA, ILAE 4
10	18	right temporal	3T	Engel IIB ILAE 2
11	9	right middle frontal lobe	1.5T	Engel IA, ILAE 1
12	15	right middle parietal lobe	1.5T	Engel IVA, ILAE 5
13	16	Left inferior frontal	3T	Engel IIIA, ILAE 4
14	11	left inferior frontal gyrus	3T	Engel IA, ILAE 1

as well. Finally, we discuss how we used DTI data from HCP to construct a structural connectome graph used for graph convolutions. All preprocessing details can be found in [137].

2.5.4.1 Fluid intelligence prediction

As part of our parcellation refinement work, we use a subset of HCP for validation. The dataset contains 300 healthy subjects from the HCP S1200 release [137]. The goal of this portion of the thesis is a regression task to predict cognitive fluid intelligence scores (CFIS) from rs-fMRI connectivity data. Standard rs-fMRI preprocessing was done according to [138], which handles motion, physiological artifacts, and registration to the MNI template. For simplicity, the CFIS values are scaled between $(0 - 10)$ based on the training data of each fold.

2.5.4.2 Synthetic tumor dataset

We conduct a proof-of-concept simulation study by applying our method to 100 subjects drawn from the Human Connectome Project (HCP1) dataset [137], in which we simulate "fake tumors". We limit the analysis to 100 subjects, so that the dataset is of comparable size to our JHH cohort. Details on the acquisition paramters, sequencing, and preprocessing for both rs-fMRI and t-fMRI can be found in [137].

The language task for HCP was developed in [139] to map the anterior temporal lobe for presurgical planning. The task consisted of alternating between story comprehension and performing basic arithmetic operations (addition, subtraction etc.). In both blocks, the participants received questions

in the form of text-to-speech, to activate their language processing networks. For the motor task, participants were instructed to tap their left or right fingers, squeeze their left or right toes, or move their tongue to map motor areas (block design [140]). We used the FEAT software from FSL [141] to obtain GLM activation maps of the HCP t-fMRI.

The "fake tumors" overlayed onto the HCP1 connectomes are randomly created, and ensured to be spatially continuous, akin to a real tumor. We include this augmented dataset to simulate various issues the tumor introduces to our classification task and ultimately show robustness of our method. Our motivation for including the HCP simulation study is to evaluate our MT-GNN performance on real-world data with similar characteristics (i.e., resting-state functional connectivity inputs and labels derived from t-fMRI). Though we cannot model neural reorganization due to the tumor, our HCP simulation study provides a baseline of how removing functionality from these regions affects the overall performance.

Finally, we have downloaded a second dataset of HCP subjects (HCP2) to use solely for hyperparameter tuning of our model and baseline approaches. Once tuned, these hyperparameters are fixed for all experiments. This second HCP dataset ensures that there is no bias from our hyperparameter selection that enters the training and testing procedures for the JHH and HCP1 datasets.

2.5.4.3 Synthetic focal epilepsy dataset

It is well understood that augmenting the dataset during training can improve the generalizability and test performance of a deep learning model [112].

Furthermore, rs-fMRI studies of focal epilepsy patients are often limited in size. Therefore, following [142], we develop an EZ localization model trained entirely on augmented data derived from a neurotypical control dataset. Following the procedure outlined in [142], we create an augmented dataset from HCP for the EZ localization work.

Our dataset consists of 300 HCP subjects [137]. We generate training three samples per subject ($S = 900$ total) by varying the EZ location and/or noise model used for data augmentation. For each training sample, we augment the healthy rs-fMRI data by first randomly selecting a spatially continuous neighborhood of voxels to form the EZ and then modifying the time series at those voxels via one of six noise models: (1) adding normally distributed noise, (2) adding uniformly distributed noise, (3) adding power-law noise, (4) adding Brownian noise, (5) adding noise generated by a Levy walk process, and (6) randomly permuting the time series. Since there is no established ground truth for how the EZ affects rs-fMRI, the combination of these six noise models exposes our network to a broad range of data abnormalities during training [142]. We use this dataset for EZ localization and the noisy label experiments as well, as we have control over the ground truth labels and noise signals in the data.

2.5.4.4 Structural connectivity template from DTI

Similar to obtaining a functional connectivity graph using rs-fMRI and a parcellation, one can obtain a structural connectivity graph using a DTI fiber tracking algorithm with a parcellation as well. In our EZ localization work,

we use the structural connectivity of a healthy template to guide our graph convolutions. We derive the graph from d-MRI tractography of 50 subjects from the Human Connectome Project (HCP) dataset [137]. The d-MRI data for each subject was preprocessed using the pipeline of [143] to obtain individual structural connectivity matrices based on the Brainnetome atlas. The steps of the pipeline include linear registration, tensor estimation, tractography, and graph estimation. The graphs are then averaged and thresholded to obtain the template matrix.

2.6 Summary

To summarize, we discussed different neuroimaging modalities, specifically fMRI, and how they are used to make inference about the brain. We discussed different connectivity methods used for rs-fMRI analysis. We introduced parcellation construction methods and discussed how group-level and subject-specific approaches may not be ideal for analysis of neuroatypical cohorts. We then introduced the notion of parcellation refinement techniques, which will be topic of the next part of this thesis.

We went over prior non-deep learning based methods for functional connectivity analysis and potential pitfalls to applying these methods for our problems of eloquent cortex localization in tumor patients and EZ localization in focal epilepsy patients. We discussed a preliminary overview of deep learning techniques, and went over common network architectures that this thesis will build off of. Chapter 2 concludes with a summary of each of the datasets that will be used in this thesis for validation, which include both in-house

collected and publicly available datasets.

Chapter 3

Parcellation refinement techniques

3.1 Introduction

While group-level parcellation construction techniques are the de-facto standard for rs-fMRI preprocessing during analysis pipelines, they may not generalize well to neuroatypical subjects, because they are usually derived on healthy subjects' data. Subject specific parcellation approaches, on the other hand, are unable to reliably draw group-level concordances, and are usually validated using test-retest metrics, which do not provide clinical use. Parcellation refinement techniques have emerged as an alternative to obtain subject-specific parcellations that have a group-level correspondance for analysis. In this section of the thesis, we will first summarize / present the work we published in the Connectomics in NeuroImaging (CNI) 2018 workshop as a part of the Medical Image Computing and Computer Assisted Intervention (MICCAI) conference [34]. Then we will present a neural-network inspired refinement module, RefineNet, which we published in the MICCAI 2022 main conference [35].

3.1.1 Contributions

Prior work in this field includes the method developed in [65], which uses the Pearson's correlation between voxel and mean time series and does not consider spatial continuity in reassignment, while needing the user to specify various parameters. We develop a Bayesian model that uses both spatial and temporal information to iteratively refine an initial functional parcellation on a patient-specific basis. Our model uses a Markov Random Field (MRF) prior to encourage spatial contiguity within the functional parcels. We employ a maximum *a posteriori* (MAP) inference strategy for voxel-wise network assignment until a predefined convergence criteria is met. Our method builds on prior work in Bayesian network modeling [144] and MRF priors for rs-fMRI data [145]. We validate our method on rs-fMRI data from 67 tumor patients from the JHH dataset. Our initial atlas is the Yeo 17 network functional parcellation [61], which is one of the most widely cited functional atlases in the literature. We compare the performance of our method with the original parcellation (no reassignment) and with reassignment according to [65], whose paper references the same parcellation. Our validation metrics include the intra-network cohesion amongst the final parcels and motor network identification via task based fMRI concordance using three distinct task paradigms.

While traditional rs-fMRI analysis pipelines (classification, regression, etc) begin with applying a parcellation, there does not exist a refinement method that incorporates the downstream task performance in its refinement scheme. Our next work we will present is called RefineNet, the first deep learning

approach for subject-specific and task-aware parcellation refinement using rs-fMRI data. RefineNet encodes both spatial and temporal information via a weight matrix that learns relationships between neighboring voxels and a coherence module that compares the voxel- and region-level time series. Importantly, RefineNet is designed as an all-purpose module that can be attached to existing neural networks to optimize task performance. We validate RefineNet on rs-fMRI data from three different datasets, each one designed to perform a different task: (1) cognitive fluid intelligence prediction (regression) on HCP [137], (2) autism spectrum disorder (ASD) versus neurotypical control (NC) classification on ABIDE [126], and (3) language localization using an rs-fMRI dataset of brain tumor patients. In each case, we attach RefineNet to an existing deep network from the literature designed for the given task. Overall, RefineNet improves the temporal cohesion of the learned region boundaries and the downstream task performance.

3.2 A Bayesian Model for parcellation refinement using Markov Random Fields

3.2.1 Model

3.2.1.1 Prior and likelihood models

Our model infers an underlying (i.e latent) network architecture that integrates both spatial contiguity and temporal synchrony across the brain. At each voxel, we leverage the time series data, the neighborhood network membership, and a binary tumor label, indicating if the voxel lies within the lesion or not. Let X_v be the network assignment for voxel v. In our framework, X_v gets

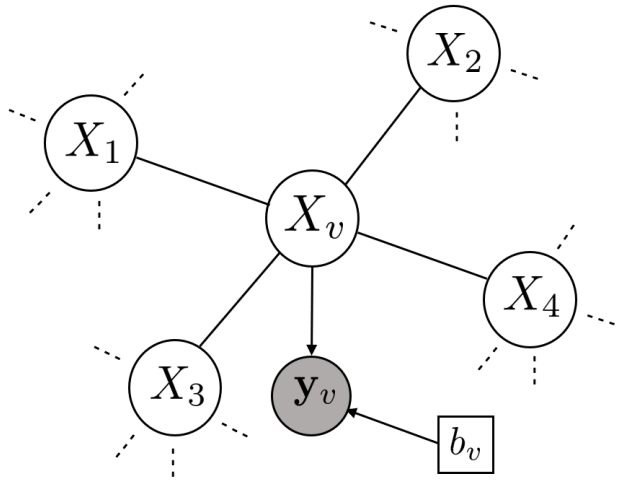

Figure 3.1: A graphical model of our framework where shaded nodes represent observed random variables.

assigned to one of $K+1$ values, where K is the number of networks (or region parcels) defined in the initial atlas. An assignment of $X_v = 0$ indicates no network membership for voxel v if it belongs to the glioma. Mathematically, let $\boldsymbol{X}_{-v}$ be the current network assignments of all other voxels in the brain. Likewise, $\boldsymbol{y}_v$ is the time series data at voxel v, $\boldsymbol{\mu}_k$ is the current reference signal for network $k \in \{1 \dots K\}$, and $b_v \in \{0, 1\}$ is the binary tumor label such that $b_v = 0$ implies that voxel v is tumorous. Our setup is illustrated in Fig. 3.1. As seen, the assignment for X_v depends on its immediate spatial neighbors. The relationship between X_v and $\boldsymbol{X}_{-v}$ is captured by an MRF prior while the relationship between X_v and $\boldsymbol{y}_v, b_v$ is captured by the data likelihood. For visualization purposes the 2D representation in Fig. 3.1 shows four neighbors per pixel. However, we have implemented a 3D model, which has six neighbors per voxel.

We encourage spatial contiguity in our latent network assignments by stipulating that voxel v will be more likely to assume the state of its neighboring voxels. We model the MRF prior after the Potts model [146]:

$$P(X_v = k|\boldsymbol{X}_{-v}) = \frac{1}{Z_x}\Psi(X_v, \boldsymbol{X}_{-v}, k) \propto \left\{1 + \exp\left[-(\beta + \sum_{i \in ne(v)} \mathbb{1}_{X_i = k})\right]\right\}^{-1} \tag{3.1}$$

where β controls the influence of the neighbor voxel network memberships on voxel v. Here, $ne(v)$ denotes the neighbors of voxel v, and the sum $\sum_{i \in ne(v)} \mathbb{1}_{X_i = k}$ captures how often these neighbors are assigned to network k. Notice that this sum will be zero for network k when $X_i \neq k$ for all $i \in ne(v)$.

The likelihood $P(\boldsymbol{y}_v|X_v = k; \boldsymbol{\mu}_k, b_v)$ is modeled after a rescaled version of the Pearson correlation coefficient between the reference $\boldsymbol{\mu}_k$ and data $\boldsymbol{y}_v$:

$$\rho = \frac{\text{cov}(\boldsymbol{y}_v, \boldsymbol{\mu}_k)}{\sigma_{\boldsymbol{y}_v}\sigma_{\boldsymbol{\mu}_k}} \tag{3.2}$$

where ρ is subsequently shifted and scaled to be between $[0, 1]$ to allow for a normalizable density. The final likelihood model is given by

$$P(\boldsymbol{y}_v|X_v = k; \boldsymbol{\mu}_k, b_v) = \frac{1}{Z_y}\Psi(\boldsymbol{y}_v, \boldsymbol{\mu}_k, b_v) = \frac{1}{Z_y}\left\{\frac{(\rho_{\boldsymbol{y}_v, \boldsymbol{\mu}_k} + 1)}{2} \times b_v\right\}. \tag{3.3}$$

The rescaled Pearson correlation coefficient goes to one for strong positive correlations and zero for strong negative correlations. The label b_v sets the likelihood to zero for tumorous voxels, which corresponds to no network membership.

3.2.1.2 Approximate inference and implementation details

The observed rs-fMRI time series $\boldsymbol{y}_v$ are conditionally independent given $\{X_v\}$. Based on the model factorization, our posterior distribution can be written as

$$P(X_v = k|\boldsymbol{X}_{-v}, \boldsymbol{y}_v; \theta) = \frac{1}{Z}\Psi(X_v, \boldsymbol{X}_{-v}, k)\Psi(\boldsymbol{y}_v, \boldsymbol{\mu}_k, b_v) \tag{3.4}$$

where $\Psi(X_v, \boldsymbol{X}_{-v}, k)$ models the prior, $\Psi(\boldsymbol{y}_v, \boldsymbol{\mu}_k, b_v)$ models the likelihood under the belief that $X_v = k$, and Z is a normalization constant that combines both Z_x and Z_y. We have derived an update procedure based on maximizing the following log-posterior over all possible network assignments:

$$X_v^* = \underset{k}{\operatorname{argmax}} \Big\{ -\log Z + \log \Psi(X_v, \boldsymbol{X}_{-v}, k) + \log \Psi(\boldsymbol{y}_v, \boldsymbol{\mu}_k, b_v) \Big\}. \tag{3.5}$$

We have derived an algorithm based on max product message passing to ensure atlas stability [147]. Our algorithm iterates between two main steps: updating the network assignments $\{\boldsymbol{X}_v\}$ and updating the reference signals $\{\boldsymbol{\mu}_k\}$. Let $\boldsymbol{Y} \in \mathbb{R}^{x \times y \times z \times T}$ be the aggregated rs-fMRI data across all (x, y, z) spatial coordinates and let $\boldsymbol{X}^{(t)} \in \mathbb{R}^{x \times y \times z}$ be the assignment information at iteration t. Let $\boldsymbol{B} \in \mathbb{R}^{x \times y \times z}$ be the binary tumor matrix. The values stored in $\boldsymbol{B}$ are 0 at tumorous voxels and 1 elsewhere. We initialize our algorithm with the Yeo atlas and then the Hadamard product $\boldsymbol{X}^{(0)} = \boldsymbol{X} \odot \boldsymbol{B}$, which defaults all tumorous voxel assignments to 0 due to unreliable signal at these locations. We then parcellate $\boldsymbol{Y}$ by the assignments in $\boldsymbol{X}^{(0)}$ and calculate the initial reference signals $\boldsymbol{\mu}_k^{(0)}$ for $k \in \{1 \dots K\}$ with

$$\boldsymbol{\mu}_k^{(t)} = \frac{\sum_{v=1}^{V} \boldsymbol{Y}_v \cdot \mathbb{1}(\boldsymbol{X}_v^{(t)} = k)}{\sum_{v=1}^{V} \mathbb{1}(\boldsymbol{X}_v^{(t)} = k)}. \tag{3.6}$$

At each main iteration t, we determine the voxel assignments I times according to our MAP rule in Eq. 3.5 initializing with $\boldsymbol{X}^{(t-1)}$. Using the assignments in $i-1$, we employ a flooding schedule to simultaneously determine the network values $\widehat{\boldsymbol{X}}^{(i)}$ at iteration i. The updated assignments $\boldsymbol{X}^{(t)}$ is given by majority vote over the determined network values $\{\widehat{\boldsymbol{X}}^{(i)}\}_{i=1}^{I}$. Given the new assignments, we update the network signals by Eq. 3.6 for $k \in \{1 \dots K\}$ and check if the convergence criteria has been met by calculating the fraction of non-zero assigned voxels retaining the same network membership between iterations $t-1$ and t. If the membership consistency between iterations is less than a specified stopping criteria, we repeat the procedure. Each voxel of interest has six neighbors as determined by adjacency in each of the three coordinate directions. Algorithm 1 presents our pseudo-code where the subject-specifc inputs are $\boldsymbol{B}$ and $\boldsymbol{Y}$.

[t!] [1] MRFrefinement$\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{Y}$ $\boldsymbol{X}^{(0)} \leftarrow \boldsymbol{X} \odot \boldsymbol{B}$ $\{\boldsymbol{\mu}_1^{(0)} \cdots \boldsymbol{\mu}_K^{(0)}\} \leftarrow \boldsymbol{Y}, \boldsymbol{X}^{(0)}$ Eq.(3.6) $t \leftarrow 1$ $M^{(0)} \leftarrow 0$ Membership retention $M \in [0,1]$ $M < c$ Convergence threshold $c \in [0,1]$ $\widehat{\boldsymbol{X}}^{(1)} \leftarrow \boldsymbol{X}^{(t-1)}$ $i = 2 : I$ $v \in V$ $\widehat{\boldsymbol{X}}_v^{(i)} \leftarrow$ argmax$_k \left\{ \log \Psi(X_v, \widehat{\boldsymbol{X}}_{-v}^{(i-1)}, k) + \log \Psi(\boldsymbol{y}_v, \boldsymbol{\mu}_k^{(t-1)}, b_v) \right\}$ $\boldsymbol{X}^{(t)} \leftarrow mode(\{\widehat{\boldsymbol{X}}^{(i)}\}_{i=1}^{I})$ $\{\boldsymbol{\mu}_1^{(t)} \cdots \boldsymbol{\mu}_K^{(t)}\} \leftarrow \boldsymbol{Y}, \boldsymbol{X}^{(t)}$ Eq.(3.6) $M \leftarrow \frac{\sum_v 1(\boldsymbol{X}_v^{(t-1)} = \boldsymbol{X}_v^{(t)}) \cdot \boldsymbol{B}_v}{\sum_v \boldsymbol{B}_v}$ Fraction of retained voxel memberships $t \leftarrow t+1$ **return** $\boldsymbol{X}^t$

3.2.2 Experimental results

3.2.2.1 Baseline comparisons and data

We compare our Bayesian approach with the original parcellation and with the voxel reassignment method described by Liu et al. [65]. We use the Yeo 17

network atlas due to its strong reproducibility and the large sample size used for construction. We confine our experiments to the more conservative cortical ribbon version of the Yeo atlas to get a more detailed parcellation. The method of Liu initializes the reference signals to the average time series defined by the original parcellation. From here, Liu reassigns voxel v by considering the maximum correlation between its time series and all K reference signals. A confidence value for each voxel is also computed as the ratio of the maximum correlation over the second highest correlation. The reference signal updates are only taken from voxels that have confidence values which exceed a predetermined threshold. They are computed as weighted combinations of the previous iteration's reference signals with the updated reference signals. The corresponding weights are nonlinear functions of the signal-to-noise ratio, the inter-subject variability, and the iteration number. We applied the Liu baseline with the parameter suggestions provided in [65], which were optimized for the 17 network Yeo atlas.

Our dataset includes task and rs-fMRI for 67 glioma patients who underwent preoperative mapping as part of their clinical workup. The preprocessing details are described in chapter 2. Our dataset includes three different motor paradigms that were designed to target distinct parts of the motor homonculus [130]: finger tapping, tongue moving, and foot tapping. Since the task-fMRI data was acquired for clinical purposes, only 42 patients performed the finger task, 35 patients performed the tongue task, and 20 patients performed the foot task. The population-based atlas contains 17 distinct functional networks confined to the cortical ribbon [61]. For both methods, a

network retention convergence criteria of 0.98 was used. We chose $\beta = -0.5$ and $I = 100$ iterations for our model and a confidence value of 1.5 for the Liu baseline. Different combinations of reference signal calculation between updates for both our method and the Liu baseline were explored; we have reported the optimal results in each case.

3.2.2.2 Evaluating Resting State Network Cohesion

Our intra-network cohesion metric quantifies the temporal synchrony between voxels that belong to the same network [148]. Let V_k be the voxels assigned to network k, we define the Network Cohesion (NC) as the average correlation between voxels assigned to network k with the network signal $\boldsymbol{\mu}_k$.

$$NC_k = \frac{\sum_{j \in V_k} \rho \boldsymbol{y}_j, \boldsymbol{\mu}_k}{|V_k|} \tag{3.7}$$

Fig. 3.2 illustrates the difference in NC between our proposed method and both the original parcellation (left) and the Liu baseline (right). A value greater than zero is considered to be more temporally synchronous while a value less than zero is considered to be less temporally synchronous. In all 17 networks, our method outperforms the original atlas with significance $p < 0.005$. This highlights the importance of our subject-specific approach for glioma patients, whose functional networks are substantially reorganized due to tumor presence.

Naturally, the Liu baseline achieves higher NC due to its correlation-based voxel reassignment procedure. Fig. 3.3 shows the original parcellation, and the final network assignment using our method and Liu's method in a single

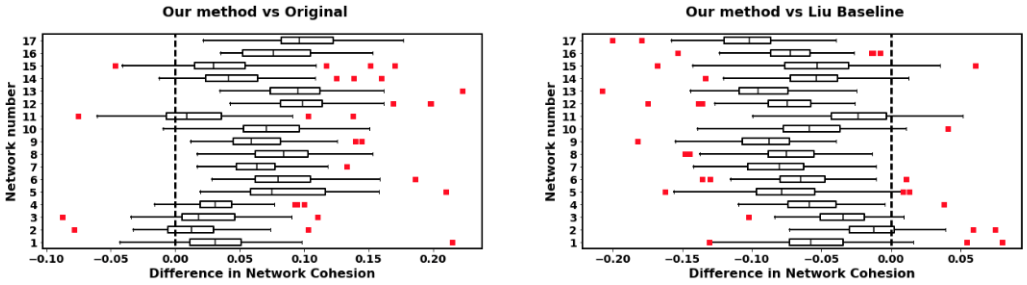

Figure 3.2: Difference in intra-network cohesion between our method and the original parcellation (**left**) and the Liu baseline (**right**).

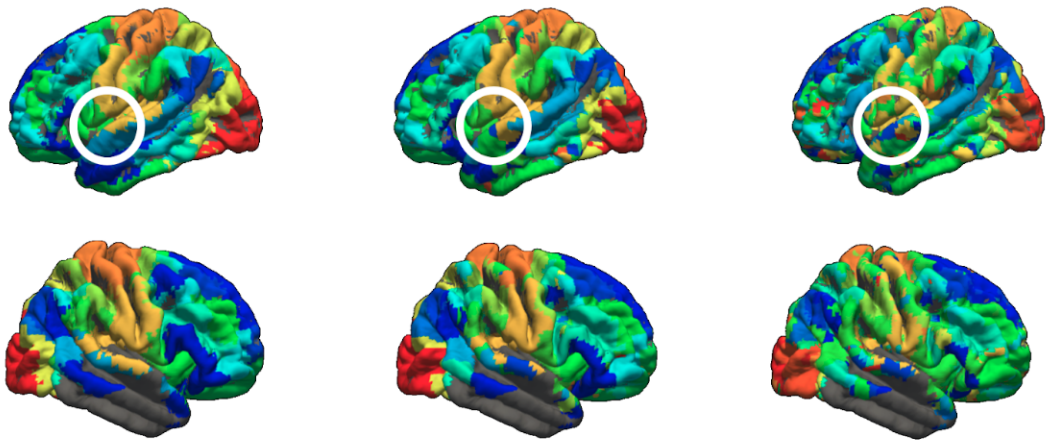

Figure 3.3: Left: Original network assignment. **Middle**: Our final network assignment. **Right**: Liu's final network assignment. For visualization, we have dilated the networks according to the liberal Yeo mask.

patient. Each distinct color represents one of the 17 networks. We observe an overall lack of spatial contiguity in the Liu baseline, as highlighted in the white circle. This might be due to spurious noise within rs-fMRI signal at the voxel level, resulting in some spatially discontiguous reassignment. The large grey area in the right hemisphere is the excluded tumorous region for this subject.

Fig. 3.4 shows the proportion of voxels retained in the original network membership between our method (left) and the Liu baseline (right). We

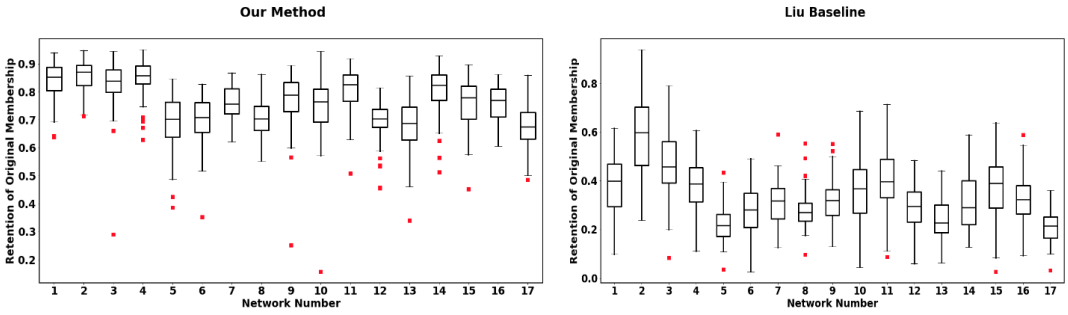

Figure 3.4: Network retention for our method (**left**) and Liu's method (**right**).

observe substantial reorganization in the networks defined from our method. Along with higher NC, this further motivates our approach, showing that many voxels in the original parcellation may not belong to the proper RSN for this cohort. We observe an even larger reorganization in the Liu networks. In the following section we conjecture that the displacement in the Liu networks may be too large, because while the Liu baseline provides more temporally cohesive RSNs, it fails to identify functionally consistent motor networks.

3.2.2.3 Motor Network Concordance, as Validated by Task-fMRI

Our second experiment quantifies the rs-fMRI concordance betweeen the pseudo-ground truth motor network in each patient and the motor RSN identified by each of the methods. Specifically, we will use the GLM activation map across three distinct motor tasks to define seed locations for motor functionality. The seed is defined as a group of highly activated voxels within the activation map. The Yeo atlas separates the motor network into two different parcels [61]. Our measure of task concordance will be the maximum correlation between the reference signals of these two RSNs and the average time series associated with the GLM activation seed. We determine that a

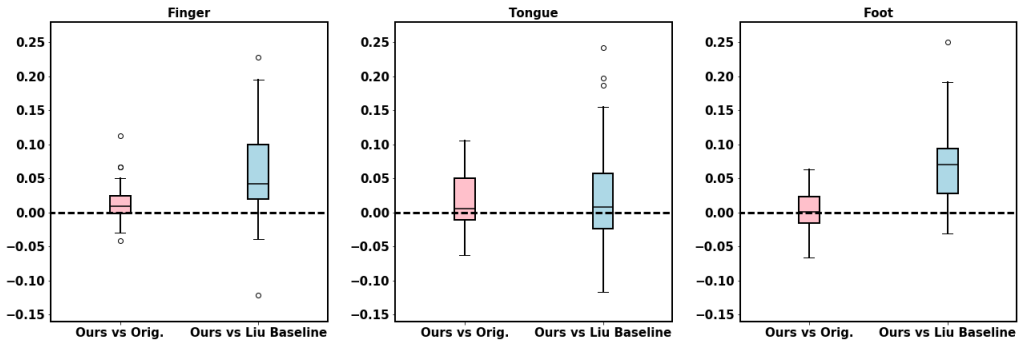

Figure 3.5: Difference in task concordance between our method and both the original atlas (pink) and the Liu baseline (blue). Our method achieves significantly better performance in five out of the six comparisons.

Table 3.1: P-values for our method vs. the original atlas and the Liu baseline.

Task	Sample size	Ours vs. Original	Ours vs. Liu
Finger	42	**3.6e-3**	**1.2e-5**
Tongue	35	**7.0e-3**	**3.7e-2**
Foot	20	0.45	**9.8e-5**

method is better at motor network identification by having a higher positive correlation with significance $p < 0.05$.

Fig. 3.5 illustrates the performance gain of our method. The pink boxplots show the difference in task concordance between our method and the original atlas, while the blue shows the difference in task concordance between our method and the Liu baseline. The tasks are ordered as finger, tongue, and foot from left to right. Table 3.1 summarizes the results and corresponding p-values for this experiment. The values in bold show when our method outperforms other methods with a student t-test with significance threshold $\alpha = 0.05$.

Our method outperforms the Liu method in each of the three tasks. In addition, our method performs better than the original atlas in the finger and

tongue task, but not the foot task. This latter result can be due to the local area of the motor homonculus that foot activaton lies in [130] or the smaller sample size. By observing p-values reported for the finger and foot task, we conclude that no reassignment would be preferrable to the Liu baseline in this experiment. However, the Liu method RSNs were the most temporally cohesive. Though network cohesion is a desirable property for RSNs [148], we have demonstrated that higher cohesion does not always lead to a functionally consistent motor network. We conjecture that (1) Liu is too liberal in the voxel reassignment, and (2) both spatial and temporal consistency are required for RSN identification.

In summary, our method balances both spatial contiguity with temporal synchrony to help describe functional networks in patients who have undergone localized neural plasticity. We observe that our method shows more cohesive RSNs for tumor patients than a population-based functional atlas. We also determine that the motor network refined by our method is a closer representation to the actual motor network in these patients. This combination of results give us confidence in our method for characterizing RSNs in a lesional population. Next, we will present our neural network module for performing parcellation refinement, which can be attached to existing networks to simultaneously optimize parcellation refinement and a downstream task.

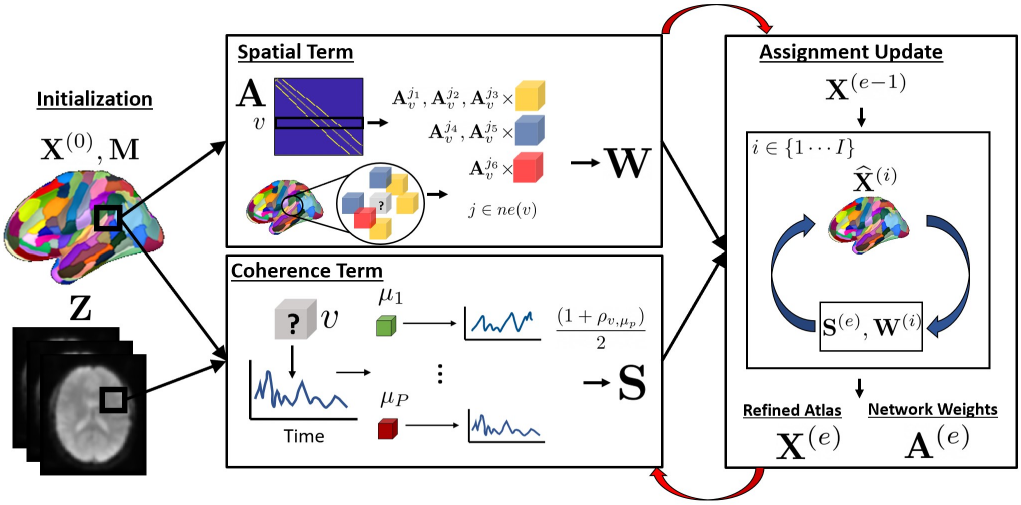

Figure 3.6: Inputs: rs-fMRI $\mathbf{Z}$, existing parcellation $\mathbf{X}^{(0)}$ and neighbor mask $\mathbf{M}$. **Top:** We show a six neighbor model for clarity. Our network parameter $\mathbf{A}$ learns voxel neighbor weights. **Bottom:** The coherence term $\mathbf{S}$ uses the pearson correlation coefficient with each mean time series μ_p. **Right:** We obtain intermediate labels $\widehat{\mathbf{X}}^{(i)}$ I times before taking the mode and producing the next epoch's parcellation $\mathbf{X}^{(e)}$, which is used during backpropagation to obtain $\mathbf{A}^{(e)}$.

3.3 RefineNet: a task-aware neural network refinement module

3.3.1 Model

Fig. 3.6 illustrates our RefineNet strategy. The inputs to RefineNet are the 4D rs-fMRI data $\mathbf{Z}$ and the original brain parcellation $\mathbf{X}^{(0)}$. We formulate a pseudo-prior, pseudo-likelihood and MAP style inference model to obtain the refined parcellation $\mathbf{X}^{(e)}$. Following this procedure, RefineNet can be attached to an existing deep network to fine-tune $\mathbf{X}^{(e)}$ for downstream task performance.

3.3.1.1 Spatial and Temporal Coherence Terms

Let V be the number of voxels in the rs-fMRI scan, and P be the number of regions in the original parcellation. We define $\mathbf{X} \in \mathbb{R}^{V \times P}$ to RefineNet as a one-hot encoded label matrix, where $\mathbf{X}_{v,p} = 1$ when voxel v is assigned to region p and $\mathbf{X}_{v,p} = 0$ otherwise. The core assumption of RefineNet is that voxels in close spatial proximity to each other are likely to belong to the same region [65, 34]. We encode this information via the intermediate activation $\mathbf{W} \in \mathbb{R}^{V \times P}$

$$\mathbf{W} = \mathrm{ReLU}\big(\mathbf{A}\mathbf{X}\big), \tag{3.8}$$

where the matrix $\mathbf{A} \in \mathbb{R}^{V \times V}$ enforces the local structure of the data. Formally, we obtain $\mathbf{A}$ as the Hadamard product of a sparse binary adjacency matrix $\mathbf{M} \in \mathbb{R}^{V \times V}$ that is nonzero only when the voxels are spatial neighbors and a learnable weight matrix $\hat{\mathbf{A}} \in \mathbb{R}^{V \times V}$ to encode spatially varying dependencies. Fig. 3.6 shows the nonzero weights in $\mathbf{A}_v$ being multiplied by the current labels of the neighbors of voxel v, where $ne(v)$ denotes neighbors of voxel v.

At a high level, Eq. 3.8 acts as a proxy for the prior probability that voxel v belongs to region p based on the contribution of its neighbors currently assigned to region p, as governed by the spatially varying weights in $\mathbf{A}$. Thus, our pseudo-prior term is designed to identify which neighbors are more important for voxel reassignment, which is important for boundary areas. Note that $\mathbf{A}$ is sparse by construction, which reduces both memory and computational overhead.

It is generally accepted that highly correlated voxels are more likely to be involved in similar functional processes, and if near each other, should be

grouped into the same region [65, 62]. Let $\mathbf{Z} \in \mathbb{R}^{V \times T}$ denote the voxel-wise time series, where T is the duration of rs-fMRI scan. Thus, our pseudo-likelihood matrix $\mathbf{S} \in \mathbb{R}^{V \times P}$ that captures the un-normalized probability of voxel v being assigned to region p is simply a shifted and scaled version of the Pearson's correlation coefficient between the voxel and mean region-wise time series, i.e., $\mathbf{S}_{v,p} = \frac{\rho_{\mathbf{z}_v,\mu_p}+1}{2}$.

Mathematically, given the voxel-to-region membership captured in $\mathbf{X}$, we can compute the region-wise mean time series μ_p as follows:

$$\mu_p = \frac{\sum_v^V \mathbf{Z} \cdot \mathbf{X}_{v,p}}{\sum_{v=1}^V \mathbf{X}_{v,p}}. \tag{3.9}$$

The correlation coefficient $\rho_{\mathbf{z}_v,\mu_p}$ can also be obtained via matrix operations, allowing us to integrate the pseudo-likelihood term directly into a deep network.

3.3.1.2 RefineNet Training and Optimization

We adopt an iterative max product approach to derive our assignment updates. For convenience, let the index e denote the main epochs and the index i denote the refinement iterate. For each epoch e, we initialize the intermediate variable $\widehat{\mathbf{X}}^{(1)}$ with the assignment matrix $\mathbf{X}^{(e-1)}$ from the previous iterate and compute the pseudo-likelihood matrix $\mathbf{S}^{(e)}$ via the mean time series defined in Eq. (3.9). We then iteratively update $\widehat{\mathbf{X}}^{(i)}$ based on neighborhood information as follows:

$$\widehat{\mathbf{X}}_{v,p}^{(i+1)} = \begin{cases} 1 & \text{argmax}_p \left\{ \mathbf{S}_{v,:}^{(e)} \odot \mathbf{W}_{v,:}^{(i)} \right\} \\ 0 & \text{else,} \end{cases} \tag{3.10}$$

where $\mathbf{W}^{(i)} = \text{ReLU}\left(\mathbf{A}\widehat{\mathbf{X}}^{(i)}\right)$ as defined in Eq. (3.8), and the operator $\odot$ is the Hadamard product. The term $\mathbf{S}^{(e)}$ remains constant throughout this iterative process from $i = \{1 \cdots I\}$ to act as the previous stationary point. The refined parcellation $\mathbf{X}^{(e)}$ for epoch e is given by the majority vote over the intermediate region assignments $\{\widehat{\mathbf{X}}^{(i)}\}_{i=1}^{I}$. We employ this iterative approach over the pseudo-prior term to leverage the space of intermediate label distributions for a robust re-assignment. We fix $I = 20$ in this work, as we empirically observed that this was large enough to provide robust reassignment.

We optimize the weights $\mathbf{A}$ in RefineNet via stochastic gradient descent to maximize the average temporal coherence with the newly assigned regions. Let V_p be the set of voxels assigned to region p. Our loss for backpropagation is

$$\mathcal{L}_{RN} = -\frac{1}{P}\sum_{v=1}^{V}\frac{1}{|V_p|}\sum_{v\in V_p}\frac{(1+\rho_{\mathbf{z}_v,\mu_p})}{2} \tag{3.11}$$

For clarity, our full training procedure is described in Algorithm 2. [t!] [1] RefineNet$\mathbf{X}, \mathbf{Z}, \mathbf{M}$,$E$,$I = 20$ $\mathbf{X}^{(0)} \leftarrow \mathbf{X}$ $\mathbf{A}^{(0)} \leftarrow \hat{\mathbf{A}}, \mathbf{M}$ Random initialization of weights in nonzero entries $\{\mu_1^{(0)} \cdots \mu_P^{(0)}\}, \mathbf{S}^{(0)} \leftarrow \mathbf{Z}, \mathbf{X}^{(0)}$ Eq.(3.9) $e = 1 : E$ $\widehat{\mathbf{X}}^{(1)} \leftarrow \mathbf{X}^{(e-1)}$ $i = 1 : I$ $\mathbf{W}^{(i)} \leftarrow \mathbf{A}^{(e-1)}, \widehat{\mathbf{X}}^{(i)}$ Eq.(3.8) $\widehat{\mathbf{X}}^{(i+1)} \leftarrow \mathbf{S}^{(e-1)}, \mathbf{W}^{(i)}$ Eq.(3.10) $\mathbf{X}^{(e)} \leftarrow mode(\{\widehat{\mathbf{X}}\}_{i=1}^{I})$ $\{\mu_1^{(e)} \cdots \mu_P^{(e)}\}, \mathbf{S}^{(e)} \leftarrow \mathbf{Z}, \mathbf{X}^{(e)}$ Eq.(3.9) $\mathcal{L}_{RN} \leftarrow \mathbf{X}^{(e)}, \{\mu_1^{(e)} \cdots \mu_P^{(e)}\}$ Eq.(3.11) $\mathbf{A}^{(e)} \leftarrow \mathcal{L}_{RN}$, SGD Backpropagation and gradient update **return** $\mathbf{X}^{(E)}$

3.3.1.3 Creating Task-Aware Parcellations with RefineNet

Once pretrained using Eq. (3.11), RefineNet can be attached to existing deep neural networks and re-optimized for performance on the downstream task.

Our strategy is to pre-train RefineNet for 50 epochs using a learning rate of 0.001 before jointly training RefineNet with the network of interest. Here, we alternate between training just the network of interest for task performance and training both RefineNet and the network of interest in an end-to-end fashion. Empirically, we observed this strategy provides a good balance of task-optimization and preserving functional cohesion. Our second-stage loss function is a weighted sum of the downstream task and the RefineNet loss in Eq. (3.11):

$$\mathcal{L}_{total} = \mathcal{L}_{net} + \lambda \mathcal{L}_{RN}, \tag{3.12}$$

where the hyperparameter λ can be chosen via a grid search or cross validation.

3.3.2 Experimental Results

We validate RefineNet on three different rs-fMRI datasets and prediction tasks. In each case, we select an existing deep network architecture from the literature to be combined with RefineNet. These networks take as input a $P \times P$ rs-fMRI correlation matrix. Fig. 3.7 illustrates the combined network architectures for each prediction task. We implement each network in Pytorch and use the hyperparameters and training strategy specified in the respective paper.

Our task-aware optimization alternates between by training the network of interest for e_a epochs while keeping RefineNet (and the input correlation matrices) fixed. We then jointly train both networks for e_a epochs while refining the parcellation, and thus, the correlation inputs between epochs.

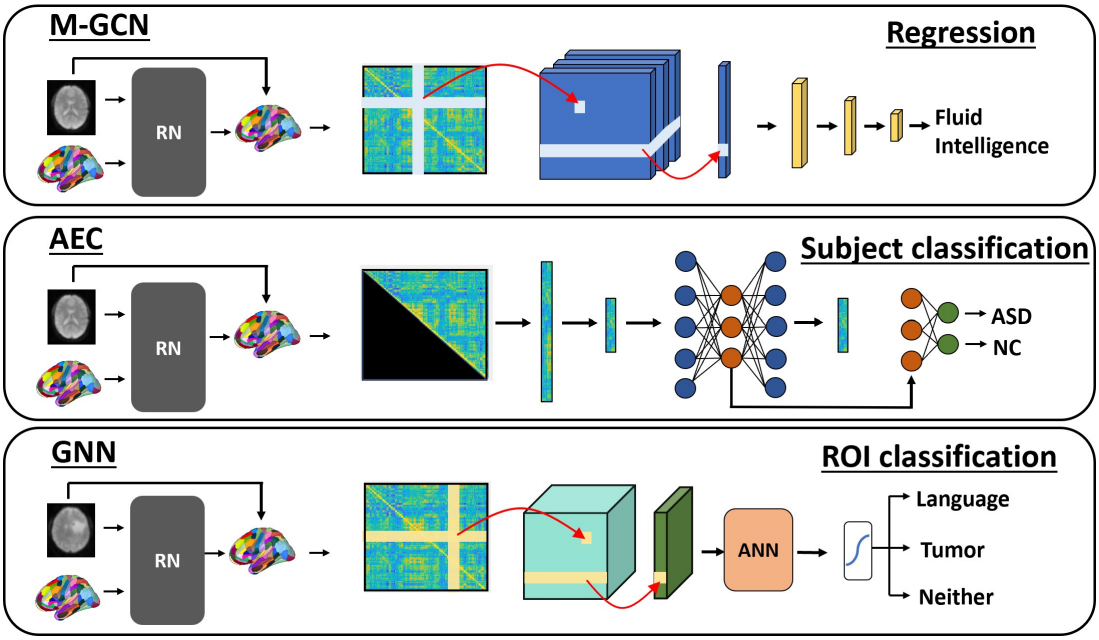

Figure 3.7: Top: The M-GCN uses a graph convolution network applied to the connectivity matrix to predict fluid intelligence in HCP subjects. **Middle:** The AEC couples an autoencoder and a single layer perceptron to classify ASD vs. NC on ABIDE data. **Bottom:** The GNN uses graph convolutions to segment the language areas of eloquent cortex on a tumor dataset.

3.3.2.1 Description of Networks and Data

M-GCN for Regression using HCP: We use the M-GCN model (rs-fMRI only) from [149] to predict the cognitive fluid intelligence score (CFIS). The dataset contains 300 healthy subjects from the publicly available Human Connectome Project (HCP) S1200 release [137]. Standard rs-fMRI preprocessing was done according to [138], which handles motion, physiological artifacts, and registration to the MNI template. For simplicity, the CFIS values are scaled between $(0-10)$ based on the training data of each fold. We report the mean absolute error (MAE) and correlation coefficient between the predicted and true scores.

AEC for Classification using ABIDE: We use the autoencoder/classifier (AEC) framework from [150] to predict subject diagnosis. The dataset contains 233 subjects (131 ASD, 102 NC) from the Autism Brain Imaging Data Exchange (ABIDE) II dataset [126]. Preprocessing details can be found in chapter 2. As per [150], the AEC network performs ASD vs. NC (neurotypical control) classification using the upper triangle portion of the rs-fMRI correlation matrix. We report the accuracy and area under the curve (AUC).

GNN for Language Localization in Tumor Patients: We use the GNN that we proposed in [36] (in Chapter 4 of this thesis) to localize language areas of the brain in a lesional cohort. The dataset contains rs-fMRI and task fMRI data from 60 brain tumor patients. Preprocessing details can be found in chapter 2. The task fMRI was used to derive “ground-truth" language labels for training and evaluation [129]. The tumor boundaries were obtained via expert segmentation.The GNN outputs a label (language, tumor, or neither)

for each region. We report the overall accuracy and AUC for detecting the language class.

3.3.2.2 Quantitative Task Performance

We compare three model configurations: (1) no refinement (original), (2) using just RefineNet to maximize temporal coherence (RefineNet only), and integrating RefineNet into an auxiliary network, as described in Section 2.3 (combined). We also apply three parcellations to each task: the Brainnetome atlas (BNA246) [60], the Craddocks 200 atlas (CC200) [59], and the Automated Anatomical Labelling (AAL90) atlas [21]. To prevent data leakage, we tune the hyperparameters λ in Eq. (3.12) and alternating training epoch for regression on 100 additional HCP subjects, yielding $\lambda = 0.2$ and $e_a = 5$.

Table 3.2 reports the quantitative performance for each model/atlas configuration. Metrics 1/2 refer to MAE/correlation for the regression task and AUC/accurary for the classification and localization tasks, respectively. We employ a ten repeated 10-fold cross validation (CV) evaluation strategy to quantify performance variability. We report mean$\pm$ standard deviation for each metric along with the FDR corrected p-value to indicate statistically improved performance in Metric 1 over the original model using the same parcellation [151]. As seen, the combined model provides statistically significant performance gains in eight out of nine experiments. In contrast, using RefineNet alone to strengthen functional coherence does not necessarily improve performance. Thus, our task-aware optimization procedure is crucial when considering downstream applications. Finally, we note that the AAL90

Table 3.2: Results across all experiments considered. Metric 1 represents MAE for regression and AUC for classification and localization while metric 2 represents correlation for regression and overall accuracy for classification and localization.

Task	Model	Atlas	Metric 1	Metric 2	P-value
CFIS Prediction	Original	BNA246	2.20 ± 0.13	0.24 ± 0.029	
		CC200	2.24 ± 0.14	0.27 ± 0.045	
		AAL90	2.22 ± 0.16	0.23 ± 0.048	
	RefineNet Only	BNA246	2.22 ± 0.13	0.19 ± 0.026	0.64
		CC200	2.22 ± 0.18	0.22 ± 0.036	0.293
		AAL90	2.15 ± 0.14	0.25 ± 0.032	0.121
	Combined	BNA246	$\mathbf{1.73 \pm 0.14}$	$\mathbf{0.3 \pm 0.039}$	$\mathbf{0.016}^{**}$
		CC200	$\mathbf{1.84 \pm 0.12}$	$\mathbf{0.34 \pm 0.046}$	$\mathbf{0.045}^{**}$
		AAL90	$\mathbf{1.91 \pm 0.11}$	$\mathbf{0.36 \pm 0.04}$	$\mathbf{0.078}^{*}$
ASD vs. NC	Original	BNA246	0.65 ± 0.017	65.5 ± 1.57	
		CC200	0.66 ± 0.024	64.9 ± 2.12	
		AAL90	0.66 ± 0.029	64.5 ± 2.49	
	RefineNet Only	BNA246	0.63 ± 0.021	63.8 ± 1.78	0.74
		CC200	0.69 ± 0.016	66.6 ± 1.80	0.22
		AAL90	0.70 ± 0.021	67.5 ± 1.94	$\mathbf{0.08}^{*}$
	Combined	BNA246	$\mathbf{0.69 \pm 0.013}$	$\mathbf{67.8 \pm 1.60}$	$\mathbf{0.062}^{*}$
		CC200	$\mathbf{0.72 \pm 0.029}$	$\mathbf{69.8 \pm 1.76}$	$\mathbf{0.022}^{**}$
		AAL90	$\mathbf{0.74 \pm 0.023}$	$\mathbf{71.8 \pm 1.84}$	$\mathbf{0.006}^{**}$
Localization	Original	BNA246	0.74 ± 0.022	84.6 ± 0.09	
		CC200	0.75 ± 0.021	85.9 ± 0.92	
		AAL90	0.67 ± 0.023	82.32 ± 1.21	
	RefineNet Only	BNA246	0.75 ± 0.023	84.95 ± 0.91	0.261
		CC200	0.75 ± 0.018	84.6 ± 0.71	0.531
		AAL90	0.65 ± 0.021	81.8 ± 1.34	0.834
	Combined	BNA246	$\mathbf{0.77 \pm 0.021}$	$\mathbf{85.9 \pm 0.91}$	$\mathbf{0.065}^{*}$
		CC200	$\mathbf{0.78 \pm 0.017}$	$\mathbf{86.9 \pm 1.01}$	$\mathbf{0.047}^{**}$
		AAL90	0.68 ± 0.019	82.63 ± 1.09	0.312

parcellation is likely too coarse for the language localization task, as reflected in the drastically lower performance metrics.

3.3.2.3 Parcellation cohesion

Fig. 3.8 illustrates the average temporal cohesion of regions in the final parcellation, as computed on the testing data in each repeated CV fold. Once again, let $\bar{}_p$ denote the mean time series in each region p. We define the cohesion C

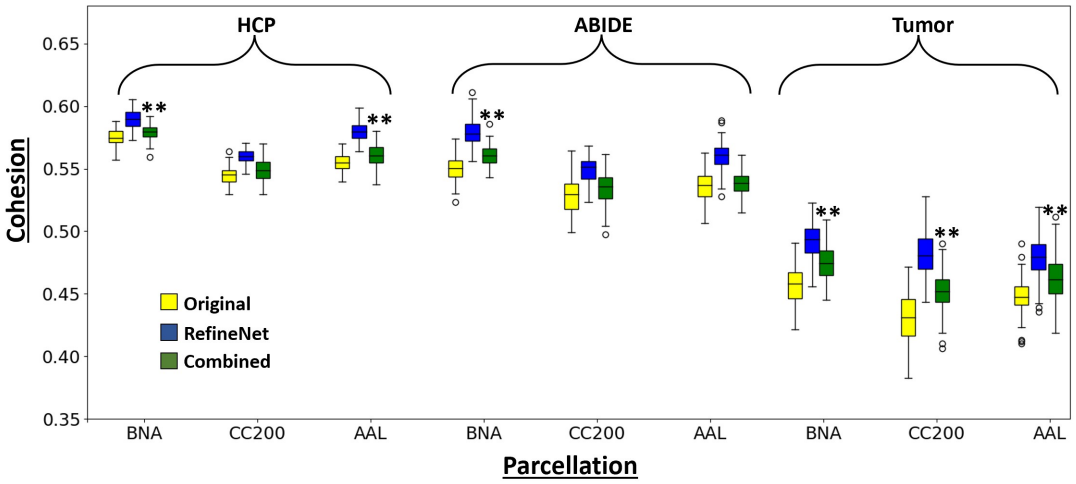

Figure 3.8: Boxplots for region cohesion across the nine experiments. Yellow refers to the original model, blue refers to RefineNet only and green refers to combined. (**) denotes a significant increase from the original to combined parcellation.

as

$$C = \frac{1}{P} \sum_{v=1}^{V} \frac{1}{|V_p|} \sum_{v \in V_p} \rho_{\mathbf{z}_v, \mu_p}. \tag{3.13}$$

Unsurprisingly, the parcellations recovered from just using RefineNet (with no downstream task awareness) achieve the highest cohesion. However, as shown in Table 3.2, these parcellations are not always suited to the prediction task. In contrast, the combined model produces more cohesive parcellations than the original atlas with statistically significant improvement denoted by (**). Taken together, attaching RefineNet to an existing model achieves a good balance between functionally-cohesive grouping and task performance.

3.4 Conclusion

In this part of the thesis, we have explored different techniques to capture the subject-specific differences in rs-fMRI parcellation construction. We have

formulated a Bayesian model that can refine a population atlas on a patient-specific basis. Our model considers both spatial contiguity as well as temporal synchrony between voxels, all while handling large and variable brain lesions. Our method outperforms established baselines for identifying a functionally consistent motor network. The use of the MRF prior along with iterative voxel reassignment shows a viable balance between properties of interest in resulting RSNs. These methodological improvements broaden the applications in which one can use rs-fMRI for analysis. We have generated a method that can be translated to other patient cohorts with anatomical brain lesions, like stroke, traumatic brain injury, or focal epilepsy. Our performance in assessing RSN cohesion shows that our method captures subect-specific functional organization well, even in a pathological population. Our MRF method outperforms both baselines in terms of motor network identification, which is an important step for preoperative planning for neurosurgical resections to avoid permanent motor network damage.

Future work for the MRF method can include using different initial atlases. Specifically, we aim to observe how our method performs with atlases of different network numbers, and different initial size (voxel membership) of networks. Methodologically, we aim to vary the number of neighbors considered in our prior model, assigning varying weights to neighbors of different geodisic distances from the center voxel. Clinically, one can use our MRF model to study reorganization of sites near the glimoa, which is known to show the most neural plasticity in these patients [152].

We then presented RefineNet, a flexible neural network module capable

of obtaining meaningful subject-specific and task-aware parcellations. Our Bayesian-inspired approach considers both spatial contiguity and temporal coherence in reassignment. We show significant performance gains across three different datasets and prediction tasks when RefineNet is appended to existing networks from the literature. Finally, we show that even the task-driven refinement procedure produces more functionally cohesive parcellations than the origial atlas. Our work is a first of its kind, as other parcellation refinement methods are not able to be jointly trained with existing deep networks for task-awareness. Now that we have explored the nuances between rs-fMRI of lesional cohorts, we will extend our work to actually performing localization of functional regions of interest in these atypical (brain tumor and focal epilepsy) cohorts in Chapters 4 and 5.

Chapter 4

Eloquent cortex localization: static connectivity analysis

4.1 Introduction

We will present our work and findings on eloquent cortex localization for brain tumor removal procedures using static connectivity as the input in this chapter of the thesis. The eloquent cortex consists of sensorimotor and language areas in the brain that are essential for human functioning. Given its importance, localizing and subsequently avoiding the eloquent cortex is a crucial step when planning a neurosurgery. Identifying and subsequently avoiding these areas during a neurosurgey is crucial for recovery and postoperative quality of life. Namely, an incision in the eloquent cortex can cause permanent physical and cognitive damage [153].

The gold standard for mapping the eloquent cortex is invasive electrocortical stimulation (ECS) performed during surgery [154]. While ECS is highly specific, it imposes a significant burden on patients, who must remain awake and functioning during the procedure. Complications due to ECS

arise for obese patients, patients with severe dysphasia, patients with severe respiratory complications, and patients with psychiatric history or emotional instability [155]. Furthermore, ECS is unavailable at the presurgical planning stage and is usually not available within the depth of the sulci, which puts more demands on the neurosurgeon and can increase surgical times [156, 157].

As a result, noninvasive task-fMRI (t-fMRI) has been increasingly popular for preoperative brain mapping [23, 24]. Namely, high activations in response to a language or motor paradigm are considered biomarkers of the respective eloquent areas [18, 33]. While task-fMRI is the most popular noninvasive mapping modality [158, 159], the activations can be unreliable for certain populations, like children, the cognitively impaired, or aphasic patients, due to an inability to follow the task protocol, or excessive head motion [25, 160].

While t-fMRI paradigms must be carefully designed to target a specific cognitive process, rs-fMRI provides a snapshot of the whole-brain, which can be used to isolate multiple functional systems [8, 49, 161]. Equally important, rs-fMRI is a passive modality and does not require the patient to perform a potentially challenging task for accurate localization. As a result, there is increasing interest in using rs-fMRI for presurgical mapping to circumvent the issues of t-fMRI [160, 162].

Prior work includes a variety of statistical and machine learning approaches to localize the eloquent cortex using fMRI data. Starting with t-fMRI, the general linear model (GLM) is used to identify voxels with significant activation [33, 163]. However, this method must be done on a per-patient basis and requires manual intervention to set the correct activation threshold. A

more unified approach is presented in ([164, 165]). Here, the authors address the problem of varying anatomical boundaries through a functional embedding of the t-fMRI data based on diffusion maps and a subsequent Gaussian mixture model fit to the signal. This method was validated on a language t-fMRI paradigm in 7 tumor patients. While promising, this method has not yet been applied to rs-fMRI data.

Deep learning (DL) methods have been increasingly popular in the neuroimaging field, and consequently, have shown promise in automatically identifying the eloquent cortex from rs-fMRI in both healthy subjects and tumor patients. For examaple, the work of [166] uses a multi-layer perceptron to classify seed-based correlation maps into one of seven resting-state networks. This method first uses PCA for dimensionality reduction followed by a two hidden layer artifical neural network for classification. Trained with t-fMRI labels, the model is extended in [160] to perform eloquent cortex localization in three separate tumor cases. While the results are promising, once again, the user must select an *a priori* seed for each network, which can affect performance. Additionally, it is trained on healthy subjects and may not accommodate changes in the brain organization due to the lesion. Finally, the large-scale study in [162] uses the same neural network architecture to identify eloquent subnetworks in 191 rs-fMRI and 83 t-fMRI scans of tumor patients. However, a success refers to whether the model identified *any clinically relevant topographies* within the scan. The study does not quantify the accuracy at the voxel or ROI level, which is the metric of interest during presurgical mapping.

4.1.1 Contributions

In this section of the thesis we will present our first models for eloquent cortex localization, specifically applied to static rs-fMRI connectivity. The work presented will be from the our CNI MICCAI 2019 paper [34] and our Medical Image Analysis (MEDIA) journal paper [37]. In contrast to prior work, we propose the first end-to-end model based on deep neural networks to identify language areas in brain tumor patients. Our problem loosely resembles image segmentation, for which deep learning approaches using convolutional neural networks (CNNs) have made great strides [167]. However, rs-fMRI captures correlated patterns of activity rather than local similarities, which cannot be represented by a traditional spatial convolution. Therefore, deep learning for rs-fMRI has focused almost exclusively on patient-wise classification [168], rather than network analysis. Our approach blends the ideas of image segmentation and functional network extraction. Namely, we construct a similarity graph from rs-fMRI data that summarizes functional connectivity between ROIs. These graphs are then input to a novel graph neural network (GNN) which leverages convolutional filters designed to act topologically upon similarity matrices [169]. The output of our GNN is a vector that classifies each node in the graph as either language, tumor, or background gray matter. Our loss function reflects the large class-imbalance in our data, as language and tumor represent a small fraction of the brain. Our model outperforms three baseline approaches in language detection, overall accuracy, and AUC.

The CNI paper (GNN) only deals with language localization, which we extend upon in our MEDIA paper to include motor localization as well. In

our MEDIA paper model, we draw from the multi-task learning (MTL) literature [170, 171] to simultaneously classify motor and language networks using a shared deep representation [172]. The goal of MTL is to improve the generalizability of a model by training it to perform multiple tasks at the same time [173]. Our architecture builds off of the CNI work and uses convolutional filters that act on rows and columns of the functional connectivity matrix [169]. The resulting graph neural network (GNN) mines the topological properties of the data in order to classify the eloquent brain regions. In addition, our training strategy can easily accommodate missing patient data in a way that optimizes the available information. This setup is highly advantageous, as the fMRI paradigms administered to each patient may vary depending on their case.

Our MEDIA paper extends upon the experiments and validation from the CNI paper substantially. We validate our method using an in-house dataset collected at the Johns Hopkins Hospital (JHH) as well as publicly available data from the Human Connectome Project (HCP), in which we simulate tumors in the healthy brain and include performance on the healthy HCP data for reference. We demonstrate that our MTL-GNN achieves higher eloquent cortex detection than popular machine learning baselines. We further show that our model can recover clinically challenging bilateral language cases when trained on unilateral language cases. Using an ablation study, we assess the value of the multi-task portion of our network. We assess robustness of our method by varying the functional parcellation used for analysis, jittering the tumor segmentations, quantifying the effects of data

augmentation, and performing a hyperparameter sweep. Finally, we include experiments that separately address potential confounders to our analysis, model optimization, and performance of our MT-GNN method on healthy HCP data. Taken together, our results highlight the promise in using rs-fMRI as part of presurgical planning procedures.

4.2 A graph neural network to localize language in brain tumor patients

4.2.1 Model

The underlying assumption of our framework is that, while the anatomical boundaries of the language network may shift, its connectivity with the rest of the brain will remain consistent [33]. We construct a weighted graph from the rs-fMRI data and classify each node in the graph as either belonging to the language network or not. We approach this problem with a neural network framework to capture complex interactions that define the language network. Our GNN node classifer extracts salient edge-node relationships and node features within the graph using a combination of specialized convolutional filters along with fully-connected (FC) layers. An important distinction in our problem is the presence of large anatomical lesions, i.e, the brain tumors. Since the tumors often encroach into the gray matter, we introduce “missing" full rows and columns in our graph. As the missing rows and columns are the most salient features of the data, we introduce two background class labels, “tumor" and “background gray matter" to avoid biasing the algorithm.

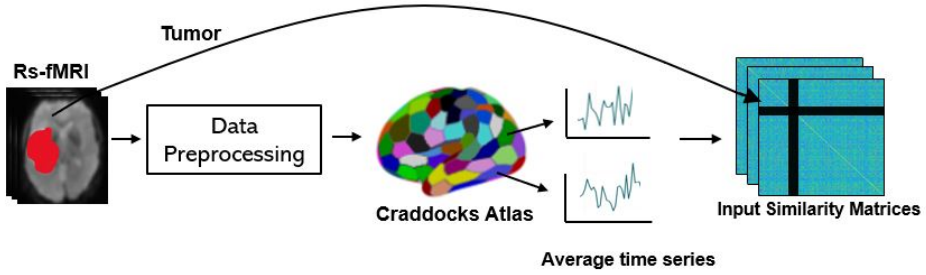

Figure 4.1: The data workflow of our model. The rs-fMRI data is preprocessed and then the Craddocks functional atlas is applied. The tumor boundaries are delineated and introduced as rows and columns of zeros in the input similarity matrix.

4.2.1.1 Graph construction

Our method treats the rs-fMRI connectivity as a weighted similarity graph, drawing inspiration from the graph theoretic literature [164, 165]. Let N be the number of brain regions in our parcellation and T be the number of time points for a rs-fMRI scan. We define $\mathbf{x}_i \in \mathbb{R}^{T\times 1}$ as the average time series extracted from region i. We normalize each time series to have zero mean unit variance. The input similarity matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$ is given by

$$\mathbf{W} = \exp\left[\mathbf{X}^T\mathbf{X} - 1\right]. \tag{4.1}$$

The tumor regions disrupt connectivity, and therefore are treated differently in our model formulation [152, 174, 34, 36]. In this work, we opt to set all edges associated with tumor nodes to zero while maintaining the value of 1 on the diagonal. We also create a separate “tumor" class at the MT-GNN output, which allows the network to learn the patterns of zero values, so that it does not bias the eloquent cortex localization. Fig. 4.1 shows the overall workflow of obtaining the input matrix.

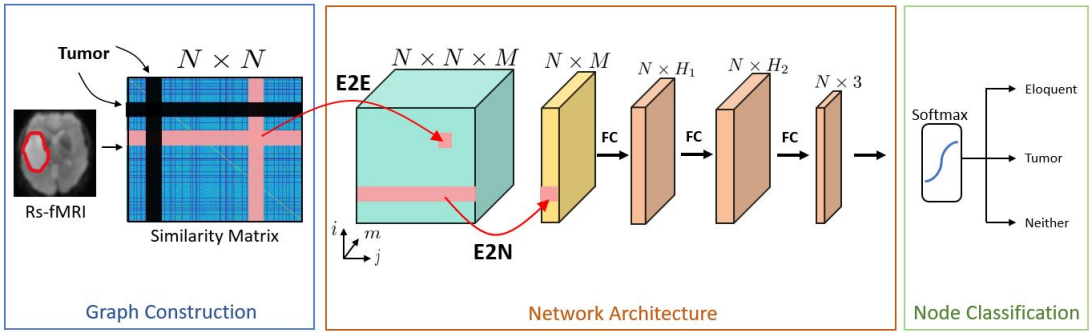

Figure 4.2: The pipeline for the GNN model. **Left**: Graph construction encodes fMRI and tumor information. **Middle**: Our GNN architecture employs E2E, E2N, and FC layers for feature extraction. **Right**: We perform a node (parcel) identification task.

Our framework assumes that tumor boundaries have been predetermined (i.e. segmented) on the voxel level. While we rely on manual segmentation in this paper, our approach is agnostic to the segmentation method and can easily be applied to automated techniques [175, 176]. Our similarity graph construction asserts that $\mathbf{W}_{i,j} > 0$ for all non-tumor regions. Therefore, even two healthy regions with a strong negative correlation will still be more functionally similar than tumor regions in our model. Our network achieves near perfect (≈ 0.99) accuracy for the tumor class due to this setup, as expected due to the zeroing out of tumor regions.

4.2.1.2 Network architecture

Fig. 4.2 shows our overall pipeline for the single-task GNN. Our GNN architecture employs both convolutional and FC layers to process node information. While traditional convolutional layers assume a grid-like organization to extract spatially local features, our GNN uses one edge-to-egde (E2E) and one

edge-to-node (E2N) layer developed in [169] that act topologically upon similarity graph data. These convolutional filters act on full rows and columns of the graph and were originally designed to perform regression for cognitive scores from diffusion MRI data. Let $m \in \{1, \cdots, M\}$ be the E2E filter index, $\mathbf{F}^m \in \mathbb{R}^{N \times N}$ be the m-th E2E filter, and $\mathbf{B}^m \in \mathbb{R}^{N \times N}$ be the bias. The feature map $\mathbf{A}^m \in \mathbb{R}^{N \times N}$ output from E2E filter m is computed as

$$\mathbf{A}^m_{i,j} = \boldsymbol{\phi}\Big(\sum_{n=1}^{N} \mathbf{W}_{i,n}\mathbf{F}^m_{i,n} + \mathbf{W}_{n,j}\mathbf{F}^m_{n,j} + \mathbf{B}^m_{i,j} \Big) \tag{4.2}$$

where $\boldsymbol{\phi}$ represents the activation function. Intuitively, an E2E filter for edge (i, j) computes a weighted sum of edge strength over all edges connected to either node i or j. Even with symmetric input $\mathbf{W}$, the learned E2E filters and corresponding feature maps are not necessarily symmetric, i.e $\mathbf{F}^m_{i,j} \neq \mathbf{F}^m_{j,i}$. This filter asymmetry is desirable, as we learn a rich representation of the data. Our motivation for using the E2E layer lies in its ability to encode multiple different views (maps) of the edge-to-edge similarities within our connectome data. This asymmetry is desirable for language localization, as these systems tend to be lateralized in the brain [177, 33]. At the E2E layer (green in Fig. 4.2), we have multiple different views along the M dimension of the edge-to-egde similarities within our connectome data.

The E2N layer condenses our representation from size $N \times N \times M$ after the E2E layer to $N \times M$, analogous to M features for each node. To obtain region-wise representations, our E2N filter performs a 1D convolution along the columns of each feature map, as the authors in [169] did not see improvement in applying the convolution to either the columns or rows of each feature map.

Furthermore, using a single orientation allows us to reduce the number of parameters in the network, which is critical given our small dataset ($S < 100$). Let $\mathbf{g}^m \in \mathbb{R}^{N\times 1}$ be the E2N filter and $\mathbf{b} \in \mathbb{R}^{M\times 1}$ be the bias. The E2N output $\mathbf{a}^m \in \mathbb{R}^N$ based on the input $\mathbf{A}^m$ from Eq. 4.2 is computed as

$$\mathbf{a}_i^m = \boldsymbol{\phi}\Big(\sum_{n=1}^{N} \mathbf{g}_n^m \mathbf{A}_{i,n}^m + \mathbf{b}_m\Big). \tag{4.3}$$

Mathematically, the E2N filter computes a single value for each node i by taking a weighted combination of edges associated with it. Our motivation for using this layer is to collapse our representation along the second dimension to obtain M features for each node. This step is similar in nature to extracting graph theoretic features, such as node centrality. In particular, we have a representation that encodes the relationship each node has to its connectivity matrix [178]. Though we use the convolutional filters developed in [169], our network and overall task are very distinct from that in [169]. There are key architectural differences to our GNN, which allow it to perform the desired eloquent cortex localization. First, the original BrainNetCNN is designed to make a single patient-wise prediction from the input connectivity matrix. In contrast, our GNN makes node-level predictions by preserving the node information through the fully-connected layers. Second, our GNN treats anatomical lesions as a separate learned class in order to remove any biases they introduce into the eloquent cortex detection.

Our node identification network uses a cascade of three FC layers of sizes $M \times H_1$, $H_1 \times H_2$ and $H_2 \times 3$ respectively. We apply activation functions between each layer. The FC layers find nonlinear combinations of the features

to best discriminate class membership for each brain parcel. Overall, our network takes $N \times N$ input and outputs an $N \times 3$ matrix for classification. Notice that the first input dimension N is maintained throughout our whole network and is not transformed. Therefore, our network maintains node structure to ultimately discriminate class membership for all nodes within one connectome at a time. As shown in Fig. 4.2, one design choice we make is to set $H_2 > H_1$ as we've observed this relationship captures the structure of our class membership well.

4.2.1.3 Loss function and implementation details

Naturally, there exists a large class imbalance in our setup, as the majority of nodes considered will be background gray matter. We cannot rely on traditional data augmentation techniques to mitigate this imbalance, as our model operates on whole-brain connectivity. To accomodate for the class imbalance, we train our model with a modified version of the Risk-sensitive cross-entropy (RSCE) loss function [179], which is designed to handle membership imbalance in multi-class classification. Let $\hat{y}_c^n$ be the output probabilitiy of our network for assigning node n to class c and y_c^n be 1 when node n belongs to class c and 0 otherwise. The loss function per patient is

$$\mathcal{L}(y_c^n, \hat{y}_c^n) = -\frac{1}{N} \sum_{n=1}^{N} \sum_{c=1}^{C} \delta_c \cdot y_c^n \log(\hat{y}_c^n) \tag{4.4}$$

where δ_c is the risk factor associated with class c. If δ_c is small, then we pay a smaller penalty for misclassifying samples that belong to class c. Our strategy is to penalize misclassifying language nodes (false negatives) larger than misclassifying background (false positives) to encourage our model to learn

the language distribution given a small number of language training samples.

We implement our network in PyTorch using the SGD optimizer with weight decay $= 5 \times 10^{-5}$ for parameter stability, and momentum $= 0.9$ to improve convergence. For our model, $\epsilon = 1$ and layer dimensions are $M = 16$, $H_1 = 9$, $H_2 = 27$ and $C = 3$. We train our model with learning rate .005 and 80 epochs, which provides for reliable performance without overfitting. The LeakyReLU$(x) = \max(0, x) + 0.33 \cdot \min(0, x)$ activation function is applied at each hidden layer. Empirically, this activation function is robust to a range of initializations. A softmax activation is applied at the final layer for classification. After cross-validation, we set $\delta = (1.1, .3, .15)$ for language, tumor, and neither classes respectively.

4.2.2 Experimental results

4.2.2.1 Baseline Comparisons

We evaluate the performance of our GNN against 3 baseline algorithms. The first is a linear SVM based on the graph theoretic measures node degree, betweenness, closeness, and eigenvector centrality [178]. The second baseline is a random forest (RF) on the stacked rs-fMRI similarity features of each node. We omit tumor class and nodes for SVM and RF as the algorithms does not exploit the spatial consistency of the similarity matrix. The last baseline is an artificial neural network (ANN) to observe how adding specialized E2E and E2N layers changes performance for this task. The ANN maintains the same input-output relationship, total parameter number, activations, and loss function as the GNN.

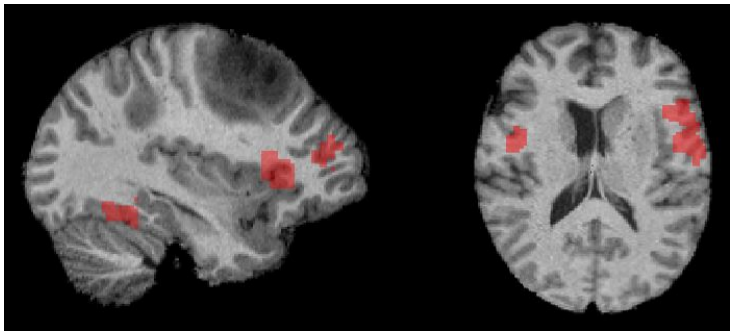

Figure 4.3: Left: One left-hemisphere language network (red) subject. **Right**: One bilateral language network subject.

4.2.3 Dataset and evaluation criteria

We evaluate the GNN language detector on rs-fMRI data from 60 patients who underwent preoperative mapping as part of routine presurgical workup. Preprocessing details can be found in Chapter 2. Our ground truth language annotations are derived from task-fMRI activations of the same patients during two language paradigms: sentence completion (SC) and silent word generation (SWG). Our dataset includes 55 patients with left-hemisphere language networks and 5 patients with bilateral networks. The tumor boundaries for each patient were manually delineated by a medical fellow using the MIPAV software. Fig. 4.3 shows language areas (red) for two separate subjects to motivate the heterogeneity of our cohort.

We parcellate our rs-fMRI data using the Craddocks functional atlas [59] with the cerebellar regions removed due to inconsistent acquisition (total $N = 384$). We assign a parcel to the language network if a majority of its voxels coincided with the ground truth task activations. We employ a ten-fold cross validation for training and testing. We stratify our folds by ensuring

at most one bilateral language subject is in each fold. We report language class accuracy as well as overall accuracy for each method that reflect a viable trade-off between true positive rate (TPR) and true negative rate (TNR). For the same set of hyperparameters that achieved the language and overall accuracies reported, we also report sensitivity (TPR) and specificity (TNR). We compute and report area under the curve (AUC) by varying hyperparameter settings to approximate ROC. We consider language vs. not language for each ROC statistic reported.

4.2.4 Results

4.2.4.1 Node classification

We present our results for the node identification task. Tumor class accuracy is not reported, as both the ANN and GNN achieved near perfect ($\approx .995$) accuracy due to the assumptions of our setup. Table 4.1 reports the node identification performance across all methods. The median for language and overall is reported. As seen, our GNN outperforms all baselines in each category. This notable performance is especially highlighted in the AUC column, as our method has the best trade-off between TPR and FPR. Our results suggest that approaching this problem with a deep learning framework is favorable, as both neural networks outperform the traditional machine learning baselines. Furthermore, we show a marked increase in performance with employing the specialized convolutional filters. This increase suggests that our network learns a more discriminative representation of language at rest than the ANN.

Table 4.1: Node identification statistics for each method.

Method	Language	Overall	Sensitivity	Specificity	AUC
Linear SVM	0.56	0.52	0.55	0.49	0.53
RF	0.38	0.77	0.34	**0.89**	0.63
ANN	0.65	0.76	0.58	0.73	0.70
GNN	**0.79**	**0.84**	**0.73**	0.81	**0.78**

4.2.4.2 Analysis of Predicted Language Area:

The specificity of our method is lower than expected due to the hemispheric symmetry of rs-fMRI data. We saw that the most frequent misclassification from our model was assigning contralateral parcels to the language class. We ran two experiments to probe whether our GNN is learning connectivity patterns associated with language rather than memorizing node locations. In Fig. 4.4 (left), we plot the histogram of true (pink) vs. predicted (blue) language parcels for frequency > 0. Each bin represents a different group of parcels, showing a large variety of language parcels in both ground truth and predicted. Second, we compared the GNN output with seed based correlation analysis (SBA), where the “seed" for each patient is selected based on the ground truth task-fMRI activations. The average rs-fMRI time course within the seed location is correlated with each of the average time courses defined by our parcellation. The correlation maps are thresholded at $\rho > 0.6$ to retain only the strong associations. Fig. 4.4 (right) illustrates a representative example. Highlighted by the white arrows, we observe right-hemisphere over prediction (blue) from our model. However, as shown by the SBA map, these right-hemisphere parcels have high resting-state connectivity with the seed

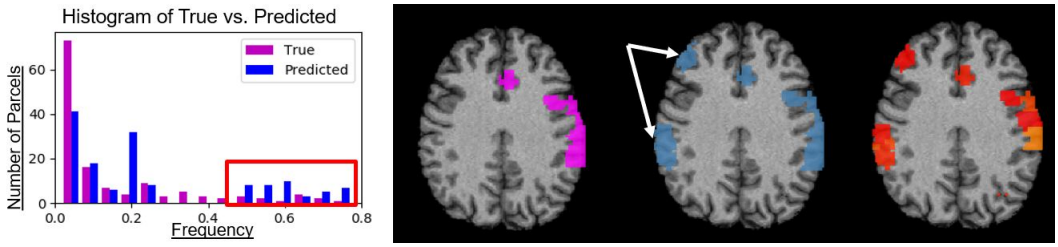

Figure 4.4: Left: Histogram of true (pink) vs predicted (blue) language parcels for frequency > 0. **Right**: Arrows show overprediction overlaps with seed based maps.

average time course. Our GNN achieves a median of **.85** dice overlap between the predicted language areas and the seed based correlation maps. As the histogram shows over prediction from our model, the SBA has high overlap with predicted areas.

4.2.4.3 Bilateral language identification

Our final experiment evaluates whether the GNN can recover a bilateral language network, even when this case is not present in the training data. Here, we trained the model on 55 left-hemisphere language network patients and tested on the remaining 5 bilateral subjects. Our model correctly predicted bilateral parcels in all five subjects. Fig. 4 shows ground truth (red) and predicted language maps (blue) for two bilateral subjects. The median language class accuracy for these five cases was **.64**. Empirically, our algorithm had trouble detecting language with the same accuracy as reported in Table 1 due to the lack of training information.

Up until now, we have demonstrated the first automated method to identify the language areas of eloquent cortex in brain tumor patients using rs-fMRI connectivity. Our model learns the resting-state functional signature of the

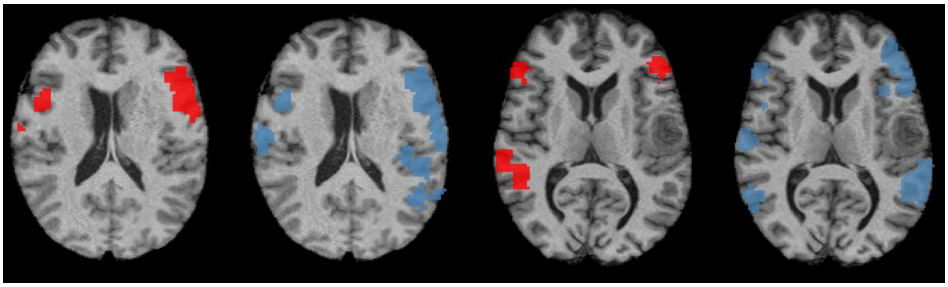

Figure 4.5: Ground truth (red) and predicted (blue) for two separate subjects. All bilateral subjects were held out of training.

language network within this tumor cohort by leveraging specialized convolutional filters that encode edge-node relationships within similarity matrices. We prove that the features extracted from our GNN are more informative for this task than standard graph theoretic features and features extracted from an ANN. We show that our model can correctly identify bilateral language networks even when trained on only unilateral network cases. Future work will focus on decoupling the lateralization and localization problems. This approach will help us overcome the intrinsic symmetry of rs-fMRI data and improve the specificity of our model. We extend upon this work by adding a multi-task learning setting to include motor localization alongside extra validation with an augmented dataset and various experiments to probe the generalizability of our deep learning approach.

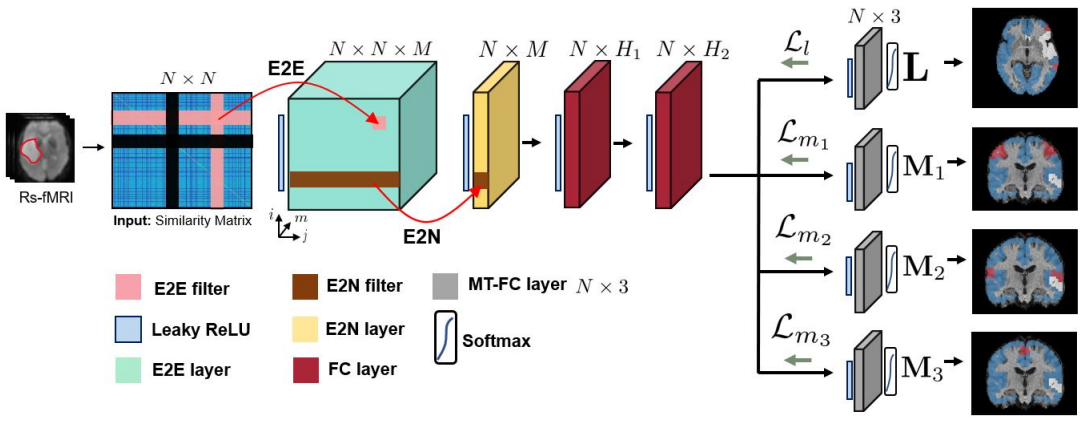

Figure 4.6: The overall workflow of our model. N is number of nodes, M is number of convolutional feature maps, H_1 is number of neurons in the first FC layer and H_2 is the number of neurons in the second FC layer. Our model uses specialized E2E and E2N filters as well as employs multi-task learning on a variety of available t-fMRI paradigms. Each grey module represents a separate 3-class segmentation task. The variables L, M_1, M_2 and M_3 represent the language, finger, tongue, and foot networks respectively, as shown by the segmentation maps where red, blue, and white refer to the eloquent, neither, and tumor classes respectively.

4.3 Multi-task graph neural networks for localization

4.3.1 Model

4.3.1.1 Extending to a multi-task learning setting

Fig. 4.6 shows the overall multi-task learning GNN (MT-GNN) network from our [37] paper. As shown, we treat the similarity graph construction the same as in [36] and our network uses the same E2E and E2N filters as previously described. The MT-GNN extends upon the base GNN by adding multi-task learning branches to localize various motor sub-networks as well as the language network.In addition, our training strategy can easily accommodate missing patient data in a way that optimizes the available information. This

setup is highly advantageous, as the fMRI paradigms administered to each patient may vary depending on their case.

We use two fully-connected layers with neurons H_1 and H_2 to extract features before the multi-task (MT) portion. The network then branches off into the MT classifier, which effectively decouples the FC weights according to which functional system it is responsible for identifying. The grey blocks in Fig. 4.6 show the MT-FC layers, where we have four separate functional systems to identify. Each grey module performs a separate 3-class classification task, shown by the segmentation maps on the RHS of Fig. 4.6. At a high level, the MT-FC layer leverages commonalities in the rs-fMRI connectivity patterns between the language and motor networks. This shared representation drastically reduces the number of parameters, relative to training the separate E2E and E2N layers in our preliminary work [36]. Clinically, our model can be extended to an arbitrary number of tasks by adding more MT branches, thus providing a valuable tool for presurgical mapping. Our MT-GNN also constructs a shared representation for language and motor areas which may shed insight into brain organization.

4.3.1.2 Classification and loss functions

Each MT-FC layer has dimension $N \times 3$ where N is the number of regions, and the three classes denote eloquent, tumor, and background gray matter, represented by the colors red, white, and blue respectively on the segmentation maps in Fig. 4.6. Recall that we treat the tumor as a separate learned class to remove any bias that zeroing out tumor edges might introduce into the

model. We emphasize that the tumor detection accuracy is not the main goal or result of this work. Instead, our goal is to maximize the eloquent detection performance. We keep the tumor regions so the input connectivity matrix is of the same dimension for each patient. Removing the tumor regions would result in different size input matrices across patients, which our model is not designed to handle. Softmax is applied and each region is classified into one of the three classes with an argmax operator. One obstacle in our datasets is the limited number of eloquent class training samples, since the language and individual motor areas are small. For the JHH cohort, the average class membership is 4.7%, 10.1% and 85.2% for the eloquent, tumor, and background gray matter class respectively. Since the convolutional filters are designed to operate upon the whole-brain connectivity matrix, our class imbalance problem cannot be mitigated by traditional data augmentation techniques. Therefore, we train our model with a modified Risk-Sensitive Cross-Entropy (RSCE) loss function [179], which is designed to handle membership imbalance in multi-class setting. Let δ_i be the risk factor associated with class i. If δ_i is large, then we pay a larger penalty for misclassifying samples that belong to class i. Due to a training set imbalance, we select different penalty values for the language class $\{\delta_i^l\}_{i=1}^3$ and motor classes $\{\delta_i^m\}_{i=1}^3$ respectively.

Let $\mathbf{L}, \mathbf{M}_1, \mathbf{M}_2$, and $\mathbf{M}_3 \in \mathbb{R}^{N\times 3}$ (Fig. 4.6) be the output of the language, finger, foot, and tongue MT-FC layers respectively. Each column of these matrices represents one of three classes: eloquent, tumor, and background. Let $\mathbf{Y}^l, \mathbf{Y}^{m_1}, \mathbf{Y}^{m_1}$, and $\mathbf{Y}^{m_3} \in \mathbb{R}^{N\times 3}$ be one-hot encoding matrices for the region-wise class labels of the language and motor subnetworks from t-fMRI. Our

loss function is the sum of four terms:

$$\mathcal{L}_{\Theta}(\mathbf{W}, \mathbf{Y}) =$$

$$\underbrace{-\sum_{i=1}^{3} \delta_i^l \log(\mathbf{L}_i)^T \mathbf{Y}_i^l}_{\text{Language Loss } \mathcal{L}_l} \underbrace{-\sum_{i=1}^{3} \delta_i^m \log(\mathbf{M}_1^i)^T \mathbf{Y}_i^{m_1}}_{\text{Finger Loss } \mathcal{L}_{m_1}} \tag{4.5}$$

$$\underbrace{-\sum_{i=1}^{3} \delta_i^m \log(\mathbf{M}_2^i)^T \mathbf{Y}_i^{m_2}}_{\text{Foot Loss } \mathcal{L}_{m_2}} \underbrace{-\sum_{i=1}^{3} \delta_i^m \log(\mathbf{M}_3^i)^T \mathbf{Y}_i^{m_3}}_{\text{Tongue Loss} \mathcal{L}_{m_3}}$$

The error from all four loss terms is backpropagated throughout the network during training, as illustrated by the green arrows in Fig. 4.6. Our framework allows for overlapping eloquent labels, as brain regions can be involved in multiple cognitive processes. To reiterate, our goal is to identify subnetworks of the eloquent cortex for presurgical planning. We take a supervised approach to this problem via multi-task classification. The model presented in this work focuses on localizing four eloquent subnetworks, as our in-house dataset contains task fMRI labels for three motor areas and one language area. We emphasize that our framework can be extended to any number of functional subsystems if the proper training labels exist. In this case, the user would simply add MT-FC layers and the corresponding cross-entropy term in the loss function. From a modeling standpoint, our edge-to-edge layer is designed to extract informative subnetworks from the rs-fMRI connectivity matrix to maximize downstream separation of the desire classes. Hence, the value of M (in this work between 8-16) is closely tied to the number of subnetworks extracted from the data.

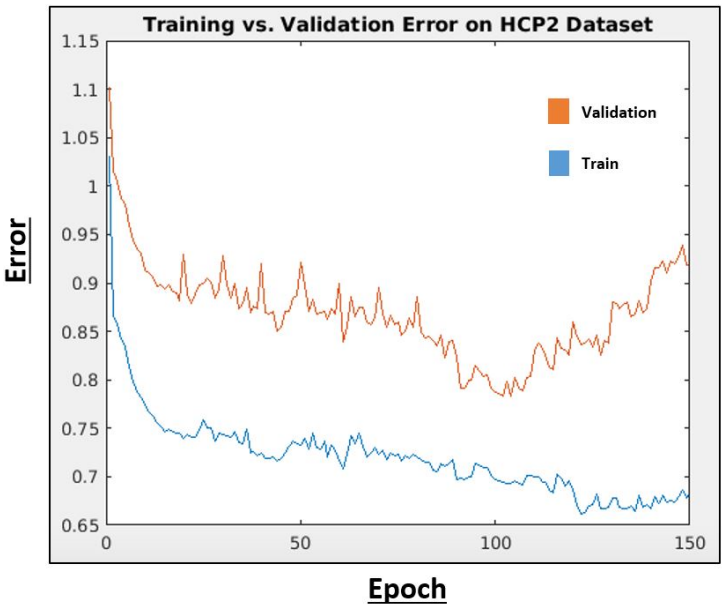

Figure 4.7: Training and validation error on HCP2 dtaset for early stopping.

4.3.1.3 Implementation details and hyperparameter selection

We used 10-fold cross validation (CV) on the HCP2 dataset (details in chapter 2) to fix the hyperparameters for all experiments. In this manner, our evaluation on the HCP1 and JHH datasets do not include biased information from the hyperparameter selection. Fig. 4.7 shows the generalization gap between training and testing, which was used to determine epoch number. Overall, we observe stable training and validation curves, which gives us confidence in the optimization of our network. For the δ hyperparameters, we performed a coarse grid search from $0 - 10$ in increments of of 10^{-1} until we found a suitable range of performance. We then performed a finer grid search in increments of 10^{-2} to obtain the final values shown in Table 4.2. We fixed the same δ values for the tumor and neither classes across branches.

Due to the clinical protocol, most JHH patients have only undergone a

Table 4.2: Hyperparameters determined via CV on the separate HCP2 dataset. lr and wd refer to learning rate and weight decay

Parameter	value	Parameter	value
N	384	wd	5×10^{-5}
M	8	$Epochs$	104
lr	0.005	δ_m	$(1.27, 0.46, 0.25)$
H_1	64	δ_l	$(2.02, 0.46, 0.25$
H_2	27		

subset of the three motor t-fMRI tasks. We handle this missing data during training by freezing the weights of the MT-FC layer in Fig. 4.6 that corresponds to the missing task when we backpropagate [180, 181]. Our strategy ensures that we mine the relevant information from the data present while preserving the fine-tuned layers of the branches that correspond to missing tasks. We train with batch size equal to one, to accommodate the missing tasks across patients. The number of subjects that performed each task is listed in Chapter 2 (datasets). We implement our network in PyTorch [182] using the SGD optimizer. The LeakyReLU$(x) = \max(0, x) + 0.33 \cdot \min(0, x)$ activation function is applied at each hidden layer. A softmax activation is applied at the final layer for classification. With GPU available, the total training time of our model is 5 minutes.

4.3.2 Localization experimental results

4.3.2.1 Baseline algorithms

We evaluate the performance of our method against three baseline algorithms.

1. A Multi-class SVM on graph theoretic features

2. Separate Random Forest Classifiers on stacked similarity matrices

3. A Fully-connected neural network with a final MT-FC layer (FC-NN)

The first baseline is a multi-class linear SVM based on node degree, betweenness centrality, closeness centrality, and eigenvector centrality [183, 178]. We include this baseline as a traditional machine learning approach for network detection in graphs. We experimented with the RBF, Gaussian, and linear kernel classes and empirically determined that the linear kernel achieves the highest AUC metrics. We set the SVM hyperparameter $c = 15.2$ using CV on the HCP2 dataset. The second baseline is a Random Forest (RF) classifier on the row vectors of the rs-fMRI similarity matrices, thus taking the connectivity as its input feature vector. Here, we train and test one separate RF classifier for each of the four functional systems. We include this baseline to assess the predictive power of the raw rs-fMRI correlations. We have implemented the RF classifier in python using 250 decision trees. The tumor nodes and class are removed for the machine learning baselines, which operate on the node level.

Our deep learning baseline is an artifical neural network that contains only fully-connected layers (FC-NN). We include this baseline to observe the performance gains in adding the specialized E2E and E2N filters. The FC-NN has five hidden layers and then a final MT-FC layer. We include more hidden layers in the FC-NN than the MT-GNN because it achieved a beter trade-off between architecture depth and width. We optimized the hyperparameters for the FC-NN using the HCP2 dataset as well, resulting in $\delta_m = (1.34, 0.43, 0.31)$ and $\delta_l = (2.13, 0.43, 0.31)$. The tumor is handled in the same way for the MT-GNN (proposed) and FC-NN (baseline).

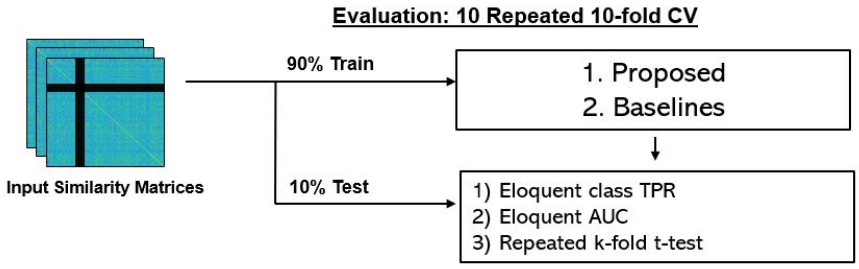

Figure 4.8: We use repeated 10-fold CV for model training and testing. We repeat each CV 10 times, ensuring that fold membership changes for each run. We report the mean and standard deviation of eloquent class true positive rate (TPR), and eloquent class area under the curve (AUC). For each baseline, we report the FDR corrected p-value from the associated t-score between our MT-GNN and the baseline, as evaluated on the AUC metric. In addition, we report the specificity, F1 and t-scores for the main classification results shown in Tables 3 and 4.

4.3.2.2 Evaluation criteria

Fig. 4.3.2.2 shows the evaluation workflow of our experiments. For each task, we report the eloquent class true positive rate (TPR) and eloquent class AUC. We note that all experiments in this work are performed on the parcel (ROI) and not voxel level. This dimensionality reduction is critical when working with a smaller clinical dataset. Eloquent class TPR is computed as the total number of correctly classified eloquent parcels divided by the total number of eloquent parcels. The AUC metric reported balances the tradeoff between the true and false positive rates of detecting the eloquent class. The reported statistics were determined using repeated 10-fold CV, where each run has a different fold membership. We report the mean and standard deviation of the metrics. To demonstrate statistically significant improvement, we perform a t-test on the repeated 10-fold CV runs, which corrects for the independence assumption between samples [151]. Formally, let r be the number of times we

repeat k-fold CV. We observe two learning algorithms A and B and measure their respective AUCs $a_{i,j}$ and $b_{i,j}$ for fold i and run j. Let $x_{i,j} = a_{i,j} - b_{i,j}$ be the performance difference, n_2 be the number of testing samples, n_1 be the number of training samples, and $\hat{\sigma}^2$ be the sample variance. The test statistic is is given by

$$t = \frac{\frac{1}{k \cdot r} \sum_{i=1}^{k} \sum_{j=1}^{r} x_{i,j}}{\sqrt{(\frac{1}{k \cdot r} + \frac{n_2}{n_1})\hat{\sigma}^2}}. \tag{4.6}$$

The variable t in Eq. 4.6 follows a t-distribution with degrees of freedom $df = kr - 1$.

4.3.2.3 HCP simulation study localization

We validate our approach on a synthetic dataset which uses healthy connectomes with fake simulated tumors. This experiment provides a proof-of-concept for our methodology on data which has similar characteristics as our main JHH cohort. The “tumors" added to this dataset are randomly positioned but created to be spatially continuous with the same size as the real tumor segmentations we obtained from the JHH cohort.

The results for this experiment are summarized in Table 4.3, where we show that the MT-GNN has superior performance in all cases when compared to the baselines. Our performance gains are underscored by the t-test, where we observe very small p-values ($p << 0.001$) for each competing baseline algorithm among each task present. Therefore, our method captures the complicated interactions between the eloquent cortex much better than the competing baseline algorithms. We also observe less performance variability across CV runs with our method compared to all of the baselines, which

Table 4.3: Mean plus or minus standard deviation for eloquent class true positive rate (TPR), specificity, F1 and AUC for the HCP cohort (100 subjects). The final column shows the FDR corrected p-values for the associated t-scores where we compare AUC between our method against each baseline.

Task	Method	Sens	Spec	F1	AUC	t-score	p-value
Lang	MTGNN	**0.67 ± 0.013**	**0.62 ± 0.012**	**0.63 ± 0.014**	**0.68 ± 0.01**		
	FCNN	0.59 ± 0.022	0.56 ± 0.021	0.58 ± 0.019	0.62 ± 0.018	14.08	3.5×10^{-44}
	RF	0.32 ± 0.036	0.61 ± 0.026	0.45 ± 0.013	0.52 ± 0.034	17.02	1.8×10^{-64}
	SVM	0.36 ± 0.026	0.49 ± 0.024	0.39 ± 0.018	0.51 ± 0.016	34.68	2.7×10^{-262}
Finger	MTGNN	**0.78 ± 0.011**	**0.75 ± 0.013**	**0.77 ± 0.014**	**0.82 ± 0.008**		
	FCNN	0.75 ± 0.014	0.69 ± 0.016	0.71 ± 0.015	0.73 ± 0.011	17.84	3.1×10^{-70}
	RF	0.41 ± 0.026	0.71 ± 0.022	0.54 ± 0.023	0.58 ± 0.028	27.61	1.2×10^{-166}
	SVM	0.41 ± 0.024	0.55 ± 0.028	0.42 ± 0.025	0.52 ± 0.015	52.66	≈ 0
Foot	MTGNN	**0.83 ± 0.009**	**0.82 ± 0.008**	**0.8 ± 0.011**	**0.79 ± 0.009**		
	FC-NN	0.73 ± 0.016	0.65 ± 0.017	0.66 ± 0.013	0.71 ± 0.015	21.45	4.9×10^{-101}
	RF	0.42 ± 0.025	0.74 ± 0.026	0.46 ± 0.021	0.58 ± 0.029	14.32	1.2×10^{-45}
	SVM	0.50 ± 0.031	0.53 ± 0.028	0.48 ± 0.027	0.51 ± 0.013	65.72	≈ 0
Tongue	MTGNN	**0.80 ± 0.01**	**0.78 ± 0.009**	**0.77 ± 0.011**	**0.78 ± 0.009**		
	FC-NN	0.76 ± 0.012	0.72 ± 0.014	0.73 ± 0.016	0.73 ± 0.015	7.63	9.1×10^{-14}
	RF	0.44 ± 0.03	0.69 ± 0.032	0.50 ± 0.026	0.57 ± 0.032	23.73	2.1×10^{-123}
	SVM	0.55 ± 0.023	0.52 ± 0.025	0.48 ± 0.024	0.53 ± 0.014	49.61	≈ 0

demonstrates robustness to the training data. We note that the RF classifier has low sensitivity and the mutli class SVM performs slightly better than chance. The performance of these machine learning baselines suggests that eloquent cortex mapping is a particularly challenging problem. Highlighted by the AUC column, the MT-GNN outperforms the FC-NN baseline in all cases. Using convolutional filters, the MT-GNN finds stereotypical connectivity patterns that identify the eloquent cortex. Compared to the motor network localization, all methods perform worse when identifying language networks, likely due to its higher anatomical variation. Fig. 4.9 shows boxplots of the AUC metric among all four methods and all four tasks. The colors red, blue, green and yellow refer to the MT-GNN, FC-NN, RF, and SVM algorithms respectively. Here we can see the performance gain and robustness of our method, which has larger median values and smaller deviations than the baselines. We repeat the performance of the algorithms on the healthy HCP dataset in the supplementary material as a way of gauging the effect that the additional tumor class has on this problem.

4.3.2.4 JHH cohort and bilateral language experiment

Our primary localization task is on the JHH tumor cohort. Table4.4 shows the eloquent class accuracy, AUC for detecting the eloquent class and t-scores for the JHH dataset. Once again, the MT-GNN has the best overall localization performance. Highlighted by the AUC and p-value column, the MT-GNN outperforms the baselines in nearly all cases, except for the tongue network. Similar to the HCP study, we observe smaller deviations with our method compared to all of the baselines, which shows robustness even when the method

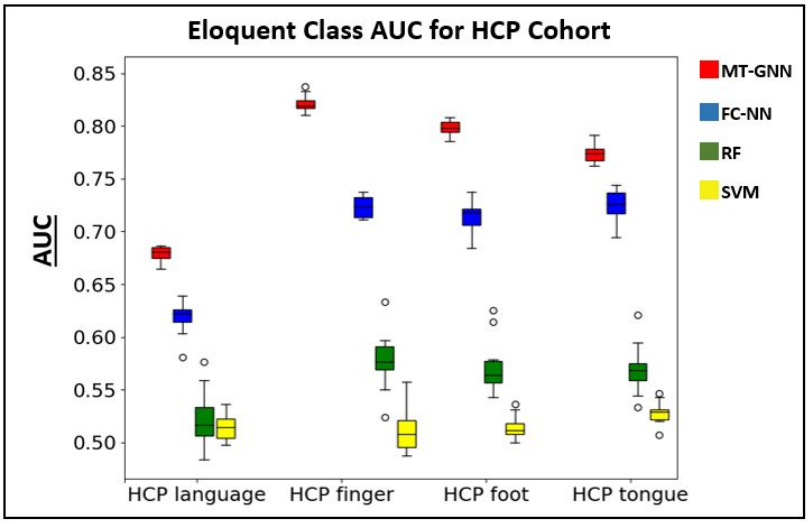

Figure 4.9: Boxplot for the AUC metric reported in Table 4.3 using 10 repeated 10-fold CVs. The colors red, blue, green and yellow refer to the MT-GNN, FC-NN, RF, and SVM methods respectively. We observe higher median performance and smaller deviations in our proposed method compared to the baseline algorithms.

is trained and tested on different subsampled versions of the data. Among both the HCP simulation study and the JHH dataset, the HCP language task was the most challenging to localize, likely due to differences between the HCP and JHH language protocols. The HCP language task was designed to target language comprehension [139] while the JHH sentence completion and silent word generation task were designed to target speech and language generation [131, 132, 33]. Fig. 4.10 shows boxplots of the AUC metric among all four methods and tasks in the JHH cohort. Once again, we can see the robustness of our MT-GNN, which has larger median values for three out of the four tasks and smaller deviations for all four tasks compared to the baselines.

Our next experiment using the JHH cohort evaluates whether the proposed model and baselines can accurately identify bilateral language networks, even

Table 4.4: Mean plus or minus standard deviation for eloquent class TPR, specificity, F1 and AUC for the JHH cohort, where the number of subjects who performed each task is shown in the first column. The final column shows the FDR corrected p-values for the associated t-scores where we compare AUC between our method against each baseline.

Task	Method	Sens	Spec	F1	AUC	t-score	p-value
Language	MTGNN	**0.75 ± 0.011**	**0.72 ± 0.01**	**0.74 ± 0.013**	**0.76 ± 0.013**		
(N = 62)	FCNN	0.68 ± 0.014	0.63 ± 0.016	0.67 ± 0.013	0.70 ± 0.015	11.56	3.8×10^{-30}
	RF	0.49 ± 0.034	0.65 ± 0.027	0.59 ± 0.029	0.61 ± 0.035	12.11	5.7×10^{-33}
	SVM	0.46 ± 0.017	0.55 ± 0.019	0.45 ± 0.02	0.52 ± 0.012	50.76	≈ 0
Finger	MTGNN	**0.85 ± 0.014**	**0.83 ± 0.016**	**0.82 ± 0.013**	**0.83 ± 0.015**		
(N = 38)	FCNN	0.77 ± 0.019	0.65 ± 0.016	0.73 ± 0.019	0.75 ± 0.017	8.36	2.7×10^{-16}
	RF	0.48 ± 0.039	0.66 ± 0.028	0.57 ± 0.034	0.60 ± 0.029	24.22	1.7×10^{-128}
	SVM	0.55 ± 0.02	0.54 ± 0.021	0.53 ± 0.015	0.54 ± 0.014	43.48	≈ 0
Foot	MTGNN	**0.81 ± 0.023**	**0.81 ± 0.021**	**0.79 ± 0.019**	**0.78 ± 0.025**		
(N = 18)	FC-NN	0.71 ± 0.023	0.62 ± 0.025	0.68 ± 0.024	0.73 ± 0.025	9.32	5.5×10^{-20}
	RF	0.45 ± 0.044	0.67 ± 0.038	0.51 ± 0.039	0.66 ± 0.047	10.58	2.0×10^{-25}
	SVM	0.53 ± 0.028	0.57 ± 0.023	0.49 ± 0.025	0.54 ± 0.021	25.63	1.2×10^{-143}
Tongue	MTGNN	0.82 ± 0.015	**0.81 ± 0.012**	0.82 ± 0.014	0.80 ± 0.014		
(N = 41)	FC-NN	**0.83 ± 0.019**	0.80 ± 0.011	**0.83 ± 0.018**	**0.80 ± 0.019**	−0.91	0.82
	RF	0.38 ± 0.028	0.65 ± 0.029	0.52 ± 0.024	0.60 ± 0.031	18.96	3.5×10^{-79}
	SVM	0.58 ± 0.021	0.51 ± 0.022	0.50 ± 0.025	0.53 ± 0.015	37.69	1.34×10^{-309}

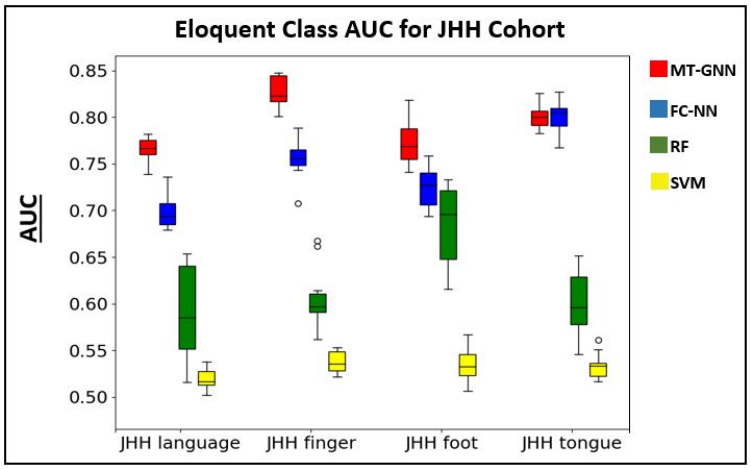

Figure 4.10: Boxplot for the AUC metric reported in Table 4.4 The colors red, blue, green and yellow refer to the MT-GNN, FC-NN, RF, and SVM methods respectively. We observe higher median performance in three out of four tasks and smaller deviations in all four tasks with the MT-GNN.

when this case is not present in the training set. This experiment assessess how well the models can identify unseen language regions based on intrinsic rs-fMRI connectivity patterns. We only perform this experiment on the JHH cohort because the JHH sentence completion and silent word generation tasks are designed to target lateralized systems, as compared to the HCP language processing and comprehension tasks. Here, we trained the model on 57 left-hemisphere language network patients and tested on the remaining 5 bilateral subjects. Table 4.5 shows the mean eloquent class and the overall accuracies for the 5 held out subjects.

Our proposed model outperforms all baselines in both per-class and overall accuracy. Fig. 4.11 shows the ground truth (blue) and predicted (yellow) labels for one bilateral language network across methods. The MT-GNN shows the best trade-off between true positives and false positives compared to the baselines. We observe that the FC-NN overpredicts too many incorrect

Table 4.5: Mean class and overall accuracy for testing on 5 bilateral language subjets. As a comparison, the mean eloquent class TPR from table 4.4 is also shown in the final column.

Method	Bilateral TPR	overall	Eloquent TPR
MT-GNN	**0.70**	**0.77**	**0.75**
FC-NN	0.51	0.72	0.68
RF	0.33	0.76	0.49
SVM	0.41	0.63	0.46

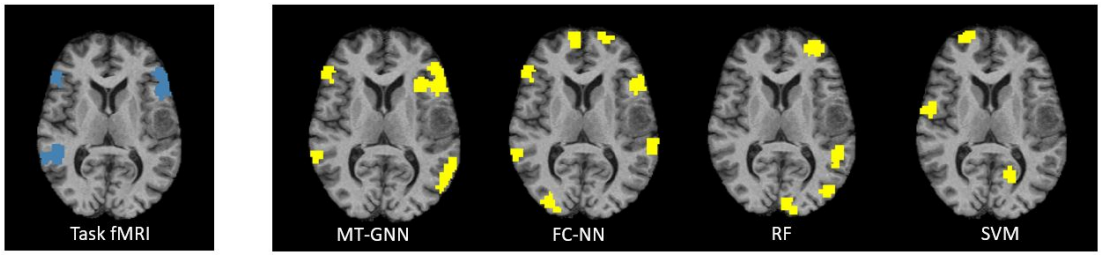

Figure 4.11: Task-fMRI "ground truth" activations (blue) and predicted (yellow) labels for one bilateral language network example across all algorithms. The MT-GNN has the highest localization accuracy.

regions, the RF is unable to detect bilateral activation, and the SVM completely misses the correct activation pattern. We point out that due to a small sample size, the bilateral language identification experiment is not as conclusive as the main results, but rather provides a proof-of-concept and clinically valuable assessment on the JHH cohort. Specifically, this experiment provides evidence that our MT-GNN does not simply memorize nodes, but rather finds intrinsic connectivity patterns associated with language. In addition, language lateralization is a key problem in clinical neuroradiology, and the bilateral experiment is exciting preliminary evidence that our MT-GNN can be applied to other clinical problems in the future.

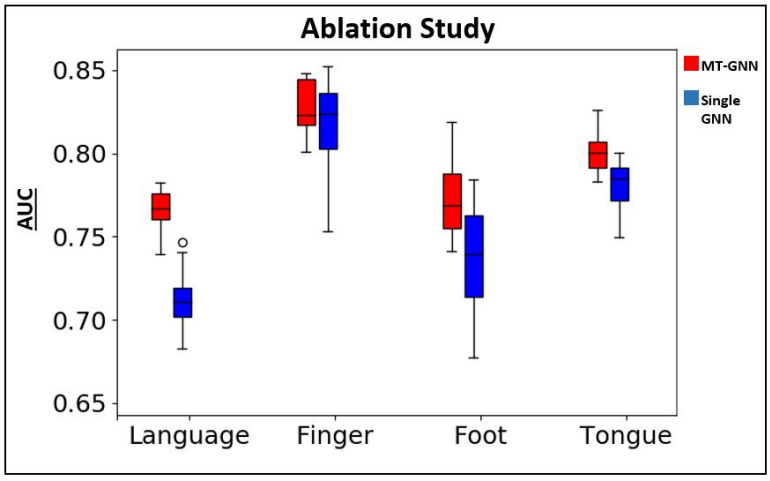

Figure 4.12: Ablation study boxplots for AUC between both cohorts. Red refers to MT-GNN and blue refers to single GNN.

4.3.3 Additional experiments

4.3.3.1 Ablation study

In this section, we assess the value of adding the multi-task learning component to our network via an ablation study. Specifically, we evaluate performance on each of the four networks by removing the other three MT-FC layers from the model during training and testing. Therefore, each single GNN (SGNN) is trained separately for each task, and evaluated on that same task, without any information from the other three tasks present. Table 4.6 shows the mean eloquent TPR, AUC for eloquent class detection, and corrected p-value for AUC between the MT-GNN and SGNN for the JHH cohort. Highlighted by $p < 0.01$, our MT-GNN outperforms the SGNN in three out of four experiments. Fig. 4.12 shows the side-by-side boxplots for AUC between the MT-GNN and SGNN, where we can see a clear divide in performance between the two methods. The MT-GNN also has smaller variability, which

Table 4.6: Mean plus or minus standard deviation for eloquent TPR and AUC for the ablation study, where the cohort is shown in the first column. The final column shows the corrected p-values from the associated t-scores where we compare AUC between our method against the single GNN (SGNN)

Task	Method	TPR	AUC	p-value
Lang.	MT-GNN	$\mathbf{0.75 \pm 0.011}$	$\mathbf{0.76 \pm 0.013}$	
	SGNN	0.73 ± 0.019	0.72 ± 0.026	1.5e-9
Finger	MT-GNN	$\mathbf{0.85 \pm 0.014}$	$\mathbf{0.83 \pm 0.015}$	
	SGNN	0.82 ± 0.021	0.81 ± 0.027	0.28
Foot	MT-GNN	$\mathbf{0.81 \pm 0.023}$	$\mathbf{0.78 \pm 0.025}$	
	SGNN	0.71 ± 0.032	0.72 ± 0.034	3.3e-3
Tongue	MT-GNN	$\mathbf{0.82 \pm 0.015}$	$\mathbf{0.80 \pm 0.014}$	
	SGNN	0.79 ± 0.019	0.77 ± 0.023	2.2e-3

shows robustness in our method.

4.3.3.2 Degrading tumor segmentation

Next, we evaluate the performance of the MT-GNN on the JHH tumor cohort without perfect manual tumor segmentations. Here, we corrupt the tumor segmentations using a combination of translation, dilation, and/or shrinking operators on the original manual segmentations. We include this experiment to assess how robust our method is to the segmentation accuracy.

Fig. 4.13 shows boxplots for the AUC metric as the tumor segmentations become more corrupt, expressed by the dice coefficient between the corrupted and true segmentations on the x-axis. As expected, overall detection performance decreases as tumor corruption increases. This result is likely due to the network learning connectivity patterns from tumor regions, which are confounding features. Also, the corrupted tumor segmentations could

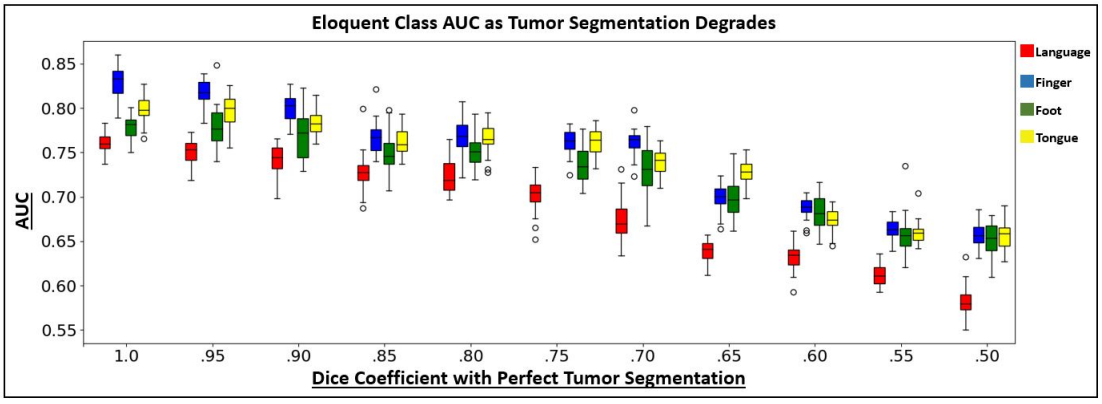

Figure 4.13: AUC boxplots using the MT-GNN on the JHH dataset as the tumor segmentations decrease in accuracy. The x-axis shows the dice coefficient of the corrupted tumor segmentation used for evaluation with the manual tumor segmentation. Corruption occurred via a combination of translating, dilating, or shrinking the manual segmentations. The colors red, blue, green and yellow refer to the JHH language, finger, foot and tongue tasks.

encroach into the eloquent cortex regions, which would also derease performance. For relatively higher dice coefficients ($> .85$), we observe only a slight decrease in performance. Therefore, the model does not require perfect tumor segmentations to work, which is valuable in a clinical setting.

We acknowledge that a restriction of our model is to have tumor segmentations manually dilineated, which can be time consuming. However, we note that there exists a large body of work describing automated techniques for tumor segmentation [175, 176] where state-of-the-art performance is up to 0.85 dice overlap with the true segmentations. We observe that our method only slightly decreases in performance at this dice coefficient, shown by Fig.4.13. Therefore, we believe our MT-GNN is a valuable tool for presurgical evaluation.

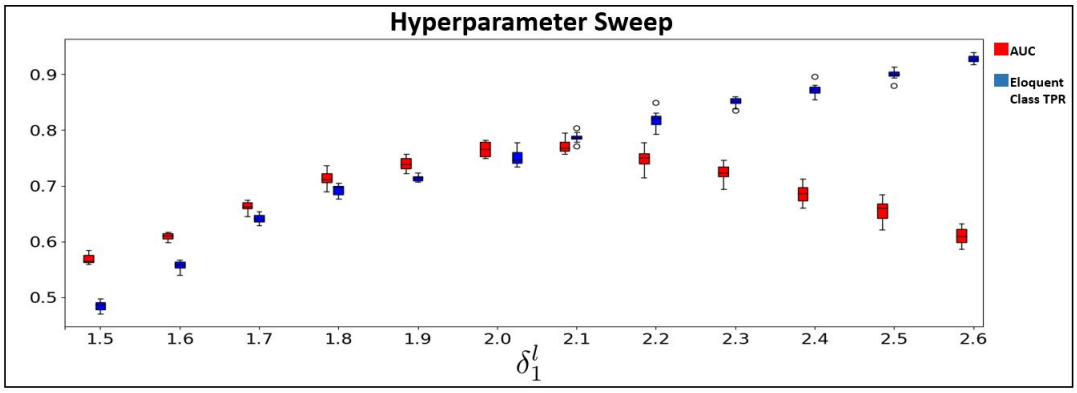

Figure 4.14: AUC (red) and class accuracy (blue) for the language class on the JHH cohort as $\delta_{l,1}$ is swept in increments of 0.1.

4.3.3.3 Hyperparameter sweep for δ_1^l

To further probe the generalizability of our model, we sweep the language class hyperparameter δ_1^l while keeping the other hyperparameters constant and plot the AUC and class accuracy on the JHH dataset. For brevity, we only show the sweep for the language class, as the tradeoff between AUC and TPR for the motor class shows the same trend. Fig. 4.14 shows the results, where AUC is in red, eloquent class TPR is in blue, and δ_1^l is swept in increments of 0.1. As δ_1^l increases, we observe an increase in false-positives, for example, when δ_1^l exceeds 2.1, AUC drops as the true positive rate continues to rise. Clinically, it is more important to minimize false negatives (missing the eloquent cortex) than to minimize false positives, as there is a greater cost for damaging the eloquent cortex during surgery. Therefore, our weighted cross-entropy strategy proves useful, even if our model tends to over predict the eloquent cortex class.

We note that the risk factor δ_c plays a role in the model performance Specifically, large values of δ_c encourage overprediction of the eloquent class, as illustrated in Fig.4.14. However, we emphasize that in this clinical application, false positive predictions are more desirable than false negative predictions, due to the severe outcomes of accidental damage to the eloquent cortex [19, 153]. Nonetheless, rectifying these overpredictions is a valuable direction for future work. In addition, we acknowledge that due to partial volume effects, our framework is conservative in handling the tumor, as the boundary parcels usually contain some number of healthy voxels. One future workaround is to use a spatially hierarchical learning scheme that increases resolution to the voxel level.

4.3.3.4 Varying parcellation choice

It is understood that the choice of parcellation can affect the rs-fMRI connectivity due to varying spatial resolution [184, 185]. Therefore, we perform eloquent cortex localization using our MT-GNN on three additional scales of the Craddocks atlas ($N = 262$, $N = 432$, and $N = 432$ regions). We choose scales that are either coarser or finer than the original $N = 384$ atlas to observe the effect that varying parcel size has on performance.

Table 4.7 shows the evaluation metrics using the MT-GNN for the JHH cohort among all three atlases considered, where the p-values for are computed with respect to the original $N = 384$ atlas. Considering a $p < 0.01$ threshold, we observe only a significant difference in performance among one of four tasks present. We observe the $N = 318$ atlas outperforming the original in

Table 4.7: Mean plus or minus standard deviation for eloquent TPR and AUC when varying the parcellation atlas. The final column shows the corrected p-values for the associated t-scores where we compare AUC between $N = 384$ against $N = 318$ and $N = 262$.

Task	Atlas	TPR	AUC	p-value
Language	384	0.75 ± 0.011	0.76 ± 0.013	
	432	0.77 ± 0.014	0.77 ± 0.012	0.11
	318	0.73 ± 0.018	0.75 ± 0.016	0.06
	262	0.71 ± 0.014	0.74 ± 0.017	1.3e-3
Finger	384	0.85 ± 0.014	0.83 ± 0.015	
	432	0.88 ± 0.013	0.84 ± 0.013	0.73
	318	0.8 ± 0.016	0.80 ± 0.017	2.7e-3
	262	0.76 ± 0.019	0.79 ± 0.014	7.0e-7
Foot	384	0.81 ± 0.023	0.78 ± 0.025	
	432	0.82 ± 0.012	0.78 ± 0.023	0.52
	318	0.81 ± 0.021	0.79 ± 0.023	0.99
	262	0.78 ± 0.027	0.77 ± 0.024	0.49
Tongue	384	0.82 ± 0.015	0.80 ± 0.014	
	432	0.83 ± 0.011	0.82 ± 0.015	0.96
	318	0.81 ± 0.016	0.79 ± 0.015	0.19
	262	0.78 ± 0.019	0.76 ± 0.017	4.9e-5

the foot functional subnetworks, denoted by a large p-value. Regarding the $N = 262$ atlas, however, three of the four tasks have a significant decrease in AUC. Our method is robust across the $N = 384$ and $N = 318$ scales but degrades in performance when the parcels become too coarse, as is the case with $N = 262$. This result implies that there is a certain spatial resolution in atlas choice that is necessary for our method to remain robust, likely due to the relatively small size of the networks we identify. However, we observe that the $N = 432$ atlas does not significantly outperform the $N = 384$ atlas, which suggests that there may be a limit of spatial resolution to which the chosen model architecture can achieve additional performance gains.

Table 4.8: Mean plus or minus standard deviation for eloquent TPR and AUC with and without data augmentation. The final column shows the corrected p-values associated with the t-scores where we compare AUC between the original and augmented.

Task	Augment	TPR	AUC	p-value
Language	No	0.75 ± 0.011	0.76 ± 0.013	
	Yes	0.76 ± 0.01	0.76 ± 0.011	0.21
Finger	No	0.85 ± 0.014	0.83 ± 0.015	
	Yes	0.86 ± 0.011	0.84 ± 0.012	0.94
Foot	No	0.81 ± 0.023	0.78 ± 0.025	
	Yes	0.80 ± 0.012	0.79 ± 0.015	0.85
Tongue	No	0.82 ± 0.015	0.80 ± 0.014	
	Yes	0.80 ± 0.017	0.80 ± 0.013	0.39

4.3.3.5 Boosting training set via data augmentation

Next, we use data augmentation to artificially increase the training set size. We include this experiment to probe the limitations of our small clinical dataset when training the highly parameterized deep network. Data augmentation has been shown to improve the performance of deep learning models due to obtaining a more comprehensive training set to help close the generalization gap [186, 187]. For the JHH cohort, we subsampled the time series data using a continuous sliding window to create 25 distinct new training similarity matrices for each subject. Our evaluation strategy remained otherwise consistent and relies on the full connectivity matrix.

Table 4.8 shows the localization performance, where the second row for each task corresponds to the augmented dataset. Overall, we observe similar performance with and without data augmentation, as highlighted by the lack of significant differences. However, we do observe smaller deviations with using augmentation, likely due to having more training samples. Ultimately,

this experiment gives us confidence that the MT-GNN method effectively mines information from the original data and is probably not overfitting on a small dataset.

4.3.4 Model analysis experimental results

In this section, we summarize the supplementary results associated with our MedIA paper. These extra experiments test the robustness and generalizability of our proposed multi-task GNN. We conduct a confounder analysis experiment, which assesses the performance of our model against various potential experimental confounders, such as the size of the tumor present. Then we conduct various model optimization experiments such as increasing model capacity or adding dropout to the model to assess how commonly used deep learning techniques for model optimization effects performance. Finally, as a proof-of-concept comparison, we include performance on localization using the healthy HCP dataset with the artificially created tumor. Taken together, our results highlight the efficiency and performance strength of our model.

4.3.4.1 Confounder analysis

Confounders, or extraneous variables that can affect the outcome of an experiment, can reduce the reliability of deep learning algorithms. There is existing work that aims to develop confounder free models applied to medical imaging [188]. Our next experiment is to assess model performance while accounting for certain potential confounders such as language laterality, tumor size, age, and gender. We include this experiment to observe any correlation

Table 4.9: Gender confounder analysis.

Task	Male TPR	Male AUC	Female TPR	Female AUC	P-value
Language	$.74 \pm .013$	$.75 \pm .031$	$.76 \pm .016$	$.76 \pm .026$	.52
Finger	$.85 \pm .017$	$.84 \pm .029$	$.84 \pm .017$	$.83 \pm .019$	.42
Foot	$.82 \pm .031$	$.79 \pm .032$	$.81 \pm .039$	$.83 \pm .019$	.37
Tongue	$.80 \pm .019$	$.81 \pm .027$	$.81 \pm .018$	$.79 \pm .021$	.33

between the confounders and model performance.

We assess model performance (both AUC and TPR for all four tasks) against four separate confounders, language laterality, tumor size, age, and gender to observe if there is a strong correlation between performance and the confounding variables. Here, laterality refers to a quantitative measure between -1 and 1 that describes handedness of the subject, and the same 10 fold-CV evaluation was used. The performance metrics are based on the same repeated 10-fold CV splits used in the paper. Likewise, we separated the testing performance based on gender and used a t-test to determine significance of model performance on men vs. women. We include the gender analysis table in and correlation plots with associated lines of best fit and p-values for the quantitative confounders.

Table 4.9 shows the gender analysis performance, where each task has a p-value greater than 0.05, indicating no significant change in performance.

Figs.4.15-4.17 shows 8 separate plots of model AUC and TPR performance across all four tasks using tumor size and age as the respective confounding variables. SFig. 3 shows the AUC and TPR for just the language task against language laterality. The p-values were calculated from the correlation coefficient between the confounder and model performance, where the line of best

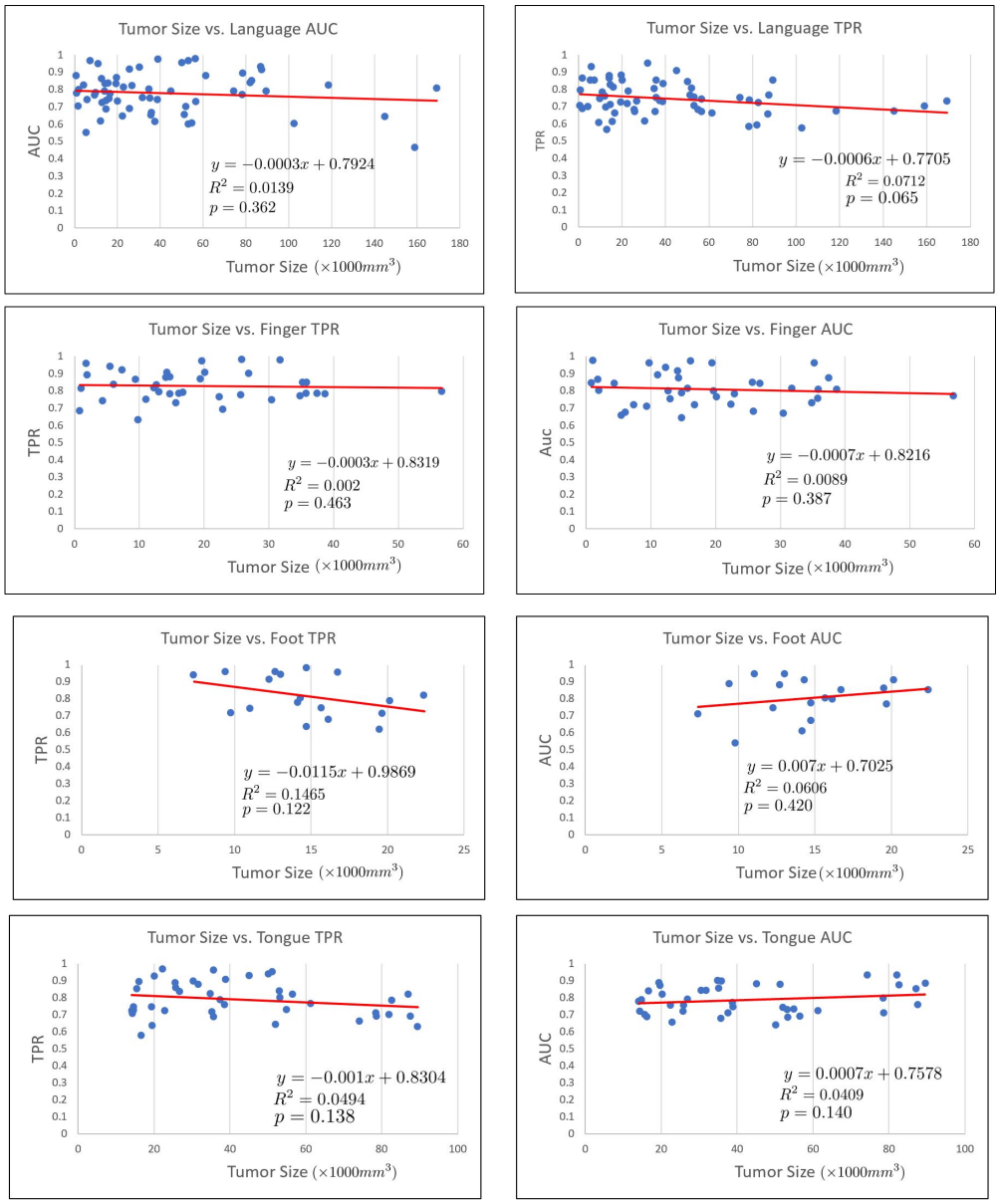

Figure 4.15: Tumor size vs. AUC and TPR for each task

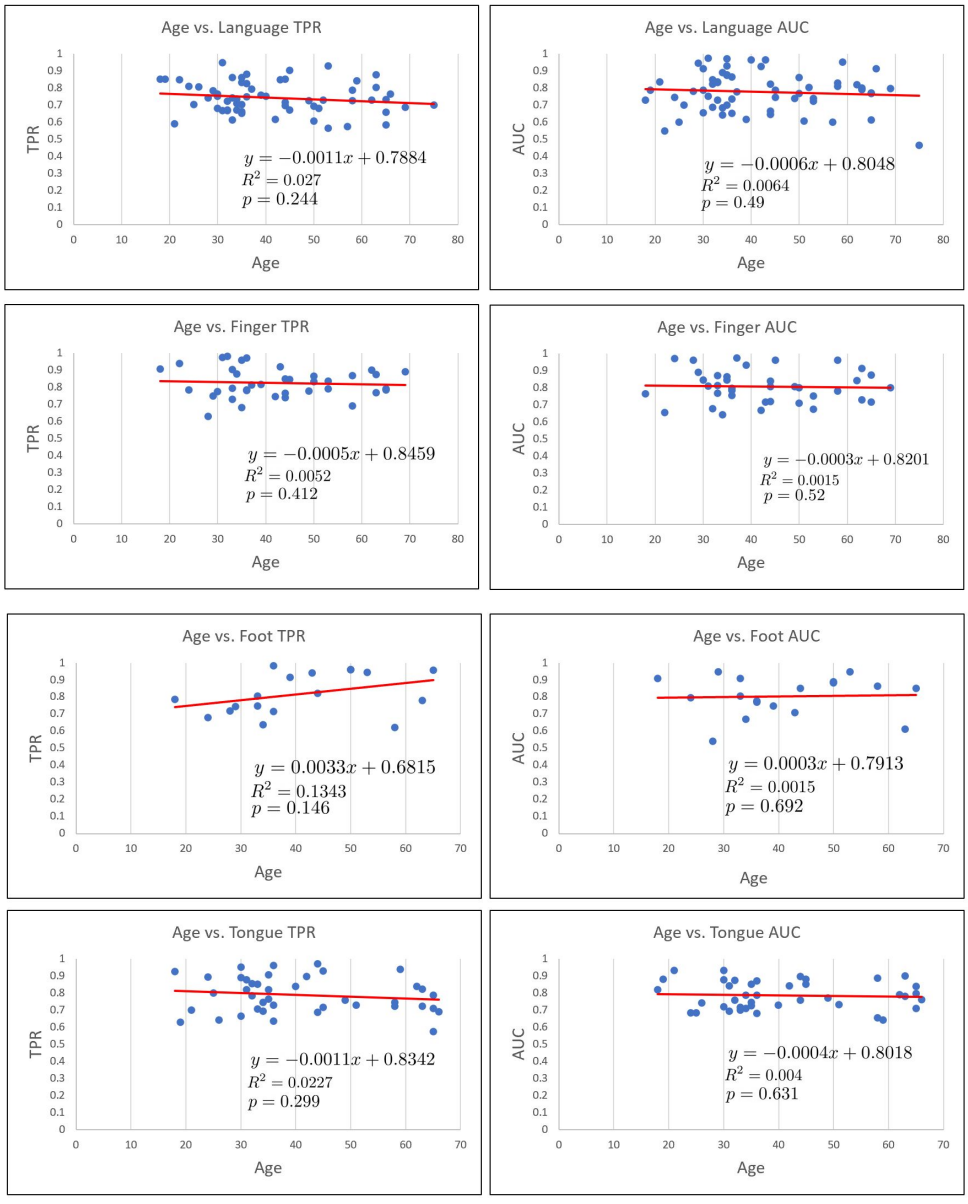

Figure 4.16: Age vs. AUC and TPR for each task

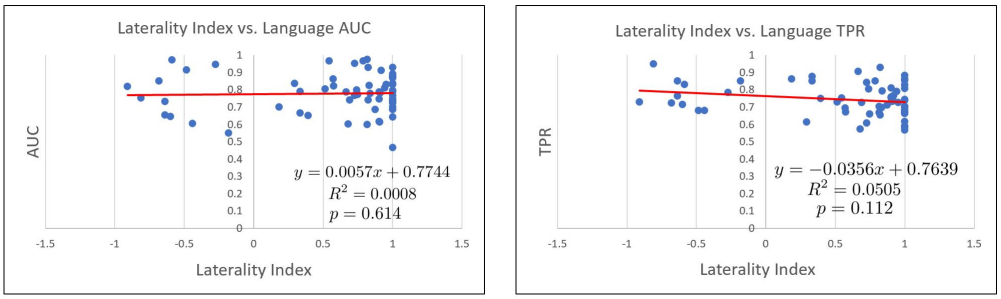

Figure 4.17: Language laterality index vs. language AUC and TPR.

fit is shown in red. As shown by $p > 0.05$, there is no significant correlation between any of the confounders and model performance for any of the four tasks.

4.3.4.2 Model optimization: model augmentation

One common strategy in deep learning is to construct an architecture that overfits to the training data and then use regularization tricks, such as dropout, to close the generalization gap on a separate validation set. Figs.4.18-4.20 show training (blue) and validation (orange) curves for the task-specific TPR with the original model across three different scales of the Craddocks atlas ($N = 318, N = 384, N = 432$). The dotted black line represents the main JHH results from the manuscript using the $N = 384$ atlas. We can observe that the original model does not fully overfit to the training data, as all four blue curves do not saturate at 1.

To arrive at a model that will overfit the training data, we increased capacity of the original model by increasing the number of feature maps in the convolutional layer and adding two fully-connected (FC) layers. Specifically, we increased M from 8 to 16, increased H_1 from 27 to 50 and added two FC

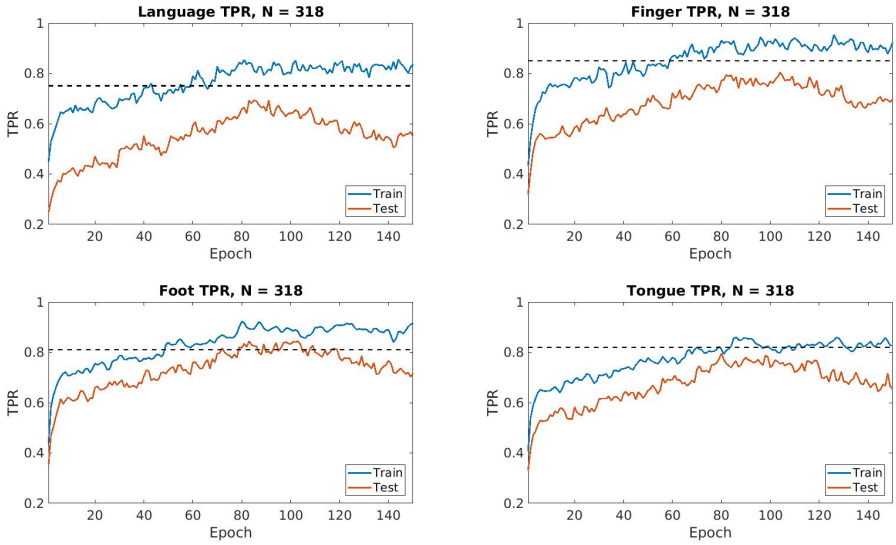

Figure 4.18: Task-specific TPR for $N = 318$ atlas with original model. We observe that the training TPR (blue) does not saturate at 1.

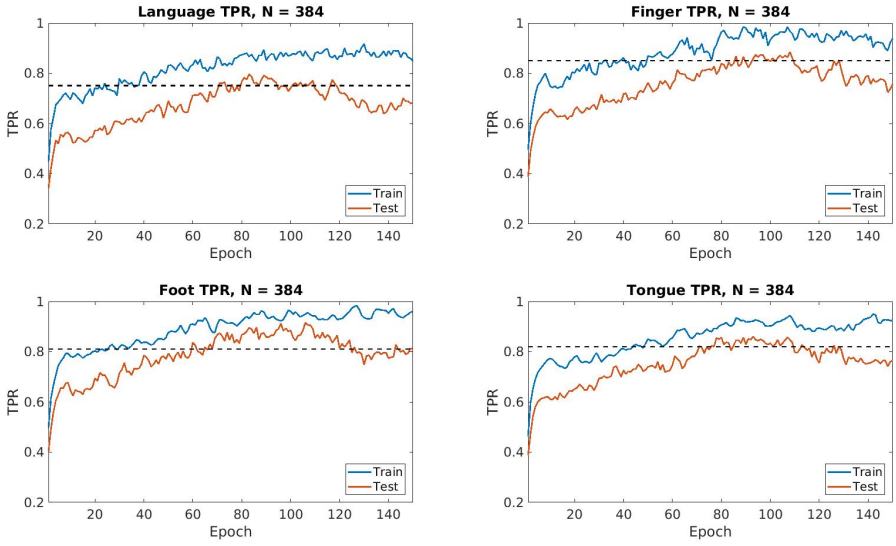

Figure 4.19: Task-specific TPR for $N = 384$ atlas with original model. We observe that the training TPR (blue) does not saturate at 1.

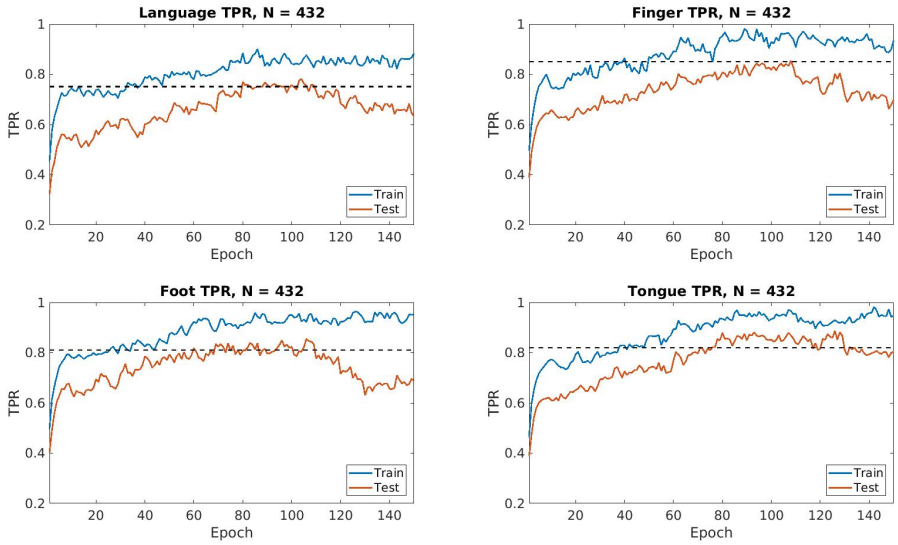

Figure 4.20: Task-specific TPR for $N = 432$ atlas with original model. We observe that the training TPR (blue) does not saturate at 1.

layers of sizes $H_3 = 25$ and $H_4 = 20$. The overfit model is shown in Fig.4.21. As shown in Figs.4.22-4.24, the model presented in Fig.4.21 overfits to the training data, as each training curve saturates at around 1.

4.3.4.3 Model optimization: adding dropout

Though the model in Fig. 4.21 fits the training data well, it performs poorly when applied to unseen test data. Now that we have identified a model with enough capacity to fit the training data, our next goal is to decrease the generalization gap. To do this, we employ dropout with $p = 0.5$ in between each hidden layer of the model except for after the E2E layer. Across three different scales of the Craddocks atlas, Figs.4.25-4.27 show the training (blue) and validation (orange) curves for the overfit model with dropout. The black dashed line indicates the original model performance on the $N = 384$ atlas.

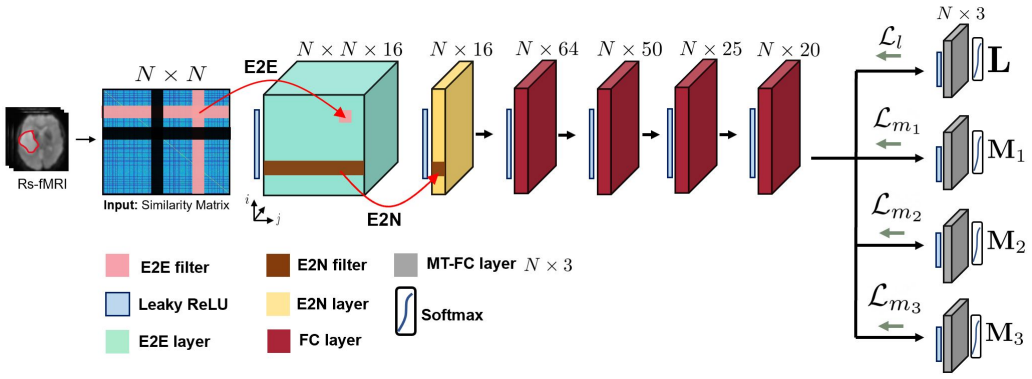

Figure 4.21: Higher capacity model used for model optimization experiment.

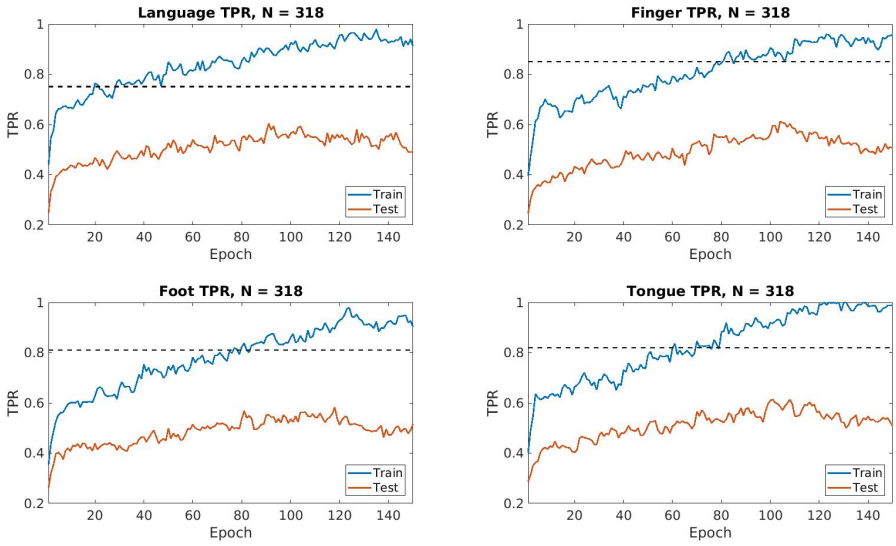

Figure 4.22: Task-specific TPR for $N = 318$ atlas with higher capacity model, where training (blue) saturates at 1 but validation (orange) decreases.

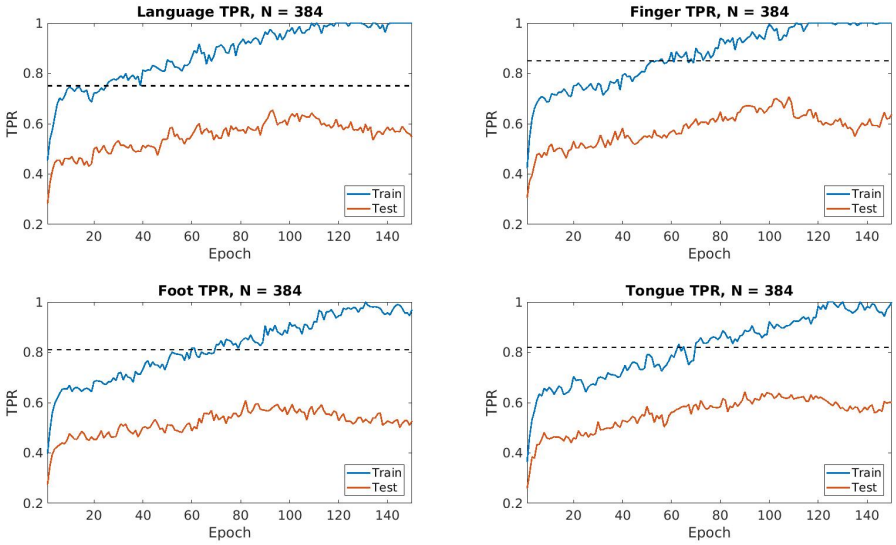

Figure 4.23: Task-specific TPR for $N = 384$ atlas with higher capacity model, where training (blue) saturates at 1 but validation (orange) decreases.

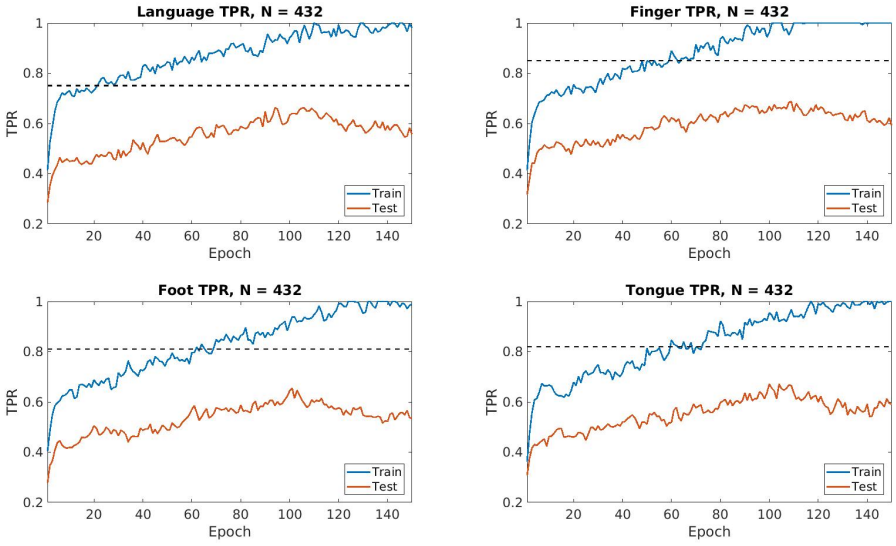

Figure 4.24: Task-specific TPR for $N = 432$ atlas with higher capacity model, where training (blue) saturates at 1 but validation (orange) decreases.

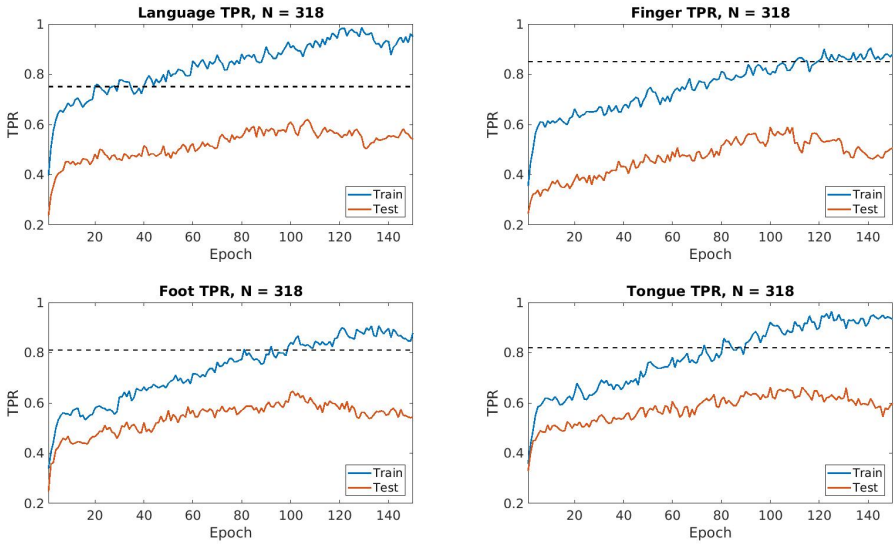

Figure 4.25: Task-specific TPR for $N = 318$ atlas with higher capacity model and dropout. The validation accuracy is lower than that of the original model.

We observe that the validation is consistently lower than the black dashed line, indicating that the original model in the manuscript has the best testing performance to unseen data.

4.3.4.4 Healthy HCP experiment

The work presented in the original manuscript treats eloquent cortex detection for tumor patients as a three-class classification problem, where tumor nodes are given their own class (i.e. not healthy and not belonging to the eloquent cortex). Removing the tumor class makes this a two-class classification problem, as each parcel is considered as either belonging to the eloquent cortex or not. Therefore, the MT-FC layers are now of size $N \times 2$ Other than the last MT-FC layers, we keep the layer dimensions consistent. To prevent biasing our hyperparameter selection, we once again use 10-fold CV on the

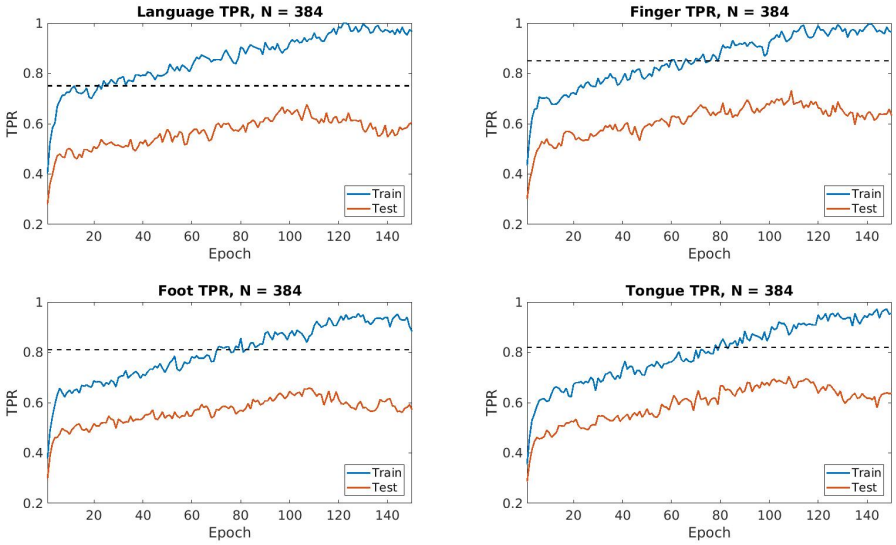

Figure 4.26: Task-specific TPR for $N = 384$ atlas with higher capacity model and dropout. The validation accuracy is lower than that of the original model.

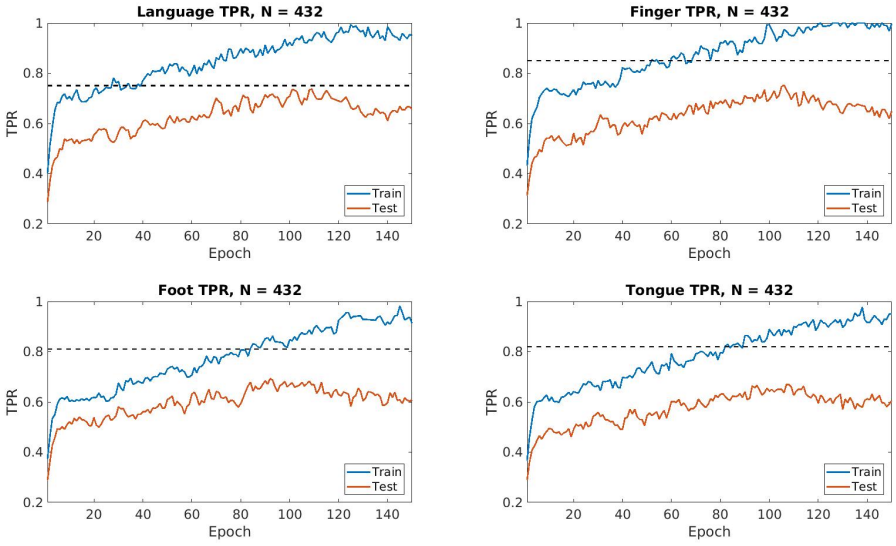

Figure 4.27: Task-specific TPR for $N = 432$ atlas with higher capacity model and dropout. The validation accuracy is lower than that of the original model.

separate healthy HCP2 dataset for hyperparameter selection, resulting in $\delta_l = (1.94, 0.54)$ and $\delta_m = (1.33, 0.54)$.

We use a 10 repeated 10-fold CV evaluation strategy, where fold membership is different for each CV. Once again, we compare the MT-GNN with a multi-class linear SVM, a RF classifier, and a fully-connected neural network (FC-NN). The FC-NN hyperparameters were selected via 10-fold CV on the healthy HCP2 dataset as well and were set to be $\delta_l = (2.04, 0.44)$ and $\delta_m = (1.52, 0.44)$. Table 4.10 shows the eloquent class true positive rate (TPR), AUC, and FDR corrected p-value for the associated t-score comparing AUC's from the MT-GNN with the baseline methods. We observe that our model outperforms each baseline at each task. Compared to the results presented in Table 4.3, we observe that each method performs better, likely due to the absence of the simulated tumor, which disrupted healthy connections in these subjects. The performance gains from the MT-GNN to the baselines are slightly higher than those in Table 4.3, shown by even smaller p-values.

4.3.5 Qualitative results

We present a novel multi-task deep learning framework to identify language processing and motor sub-regions in brain tumor patients using rs-fMRI connectivity. In comparison to baseline methods, our model achieves higher and statistically significant region-based localization performance on both a synthetic and real world clinical dataset. We show that our model can recover clinically challenging bilateral language cases when trained on unilateral cases. Our ablation study further demonstrates the value of the multi-task portion

Table 4.10: Mean plus or minus standard deviation for eloquent class TPR and AUC for the HCP cohort (100 subjects).

Task	Method	Eloquent TPR	AUC	p-value
Language	MTGNN	$\mathbf{0.70 \pm 0.011}$	$\mathbf{0.72 \pm 0.009}$	
	FCNN	0.64 ± 0.02	0.66 ± 0.017	3.7**e-**51
	RF	0.35 ± 0.034	0.55 ± 0.032	5.1**e-**100
	SVM	0.40 ± 0.026	0.53 ± 0.019	≈ 0
Finger	MTGNN	$\mathbf{0.84 \pm 0.013}$	$\mathbf{0.84 \pm 0.007}$	
	FCNN	0.78 ± 0.013	0.75 ± 0.012	7.6**e-**106
	RF	0.44 ± 0.03	0.59 ± 0.026	1.6**e-**243
	SVM	0.42 ± 0.021	0.53 ± 0.014	≈ 0
Foot	MTGNN	$\mathbf{0.86 \pm 0.01}$	$\mathbf{0.82 \pm 0.012}$	
	FC-NN	0.75 ± 0.014	0.74 ± 0.013	2.7**e-**73
	RF	0.44 ± 0.027	0.59 ± 0.028	4.2**e-**100
	SVM	0.52 ± 0.022	0.52 ± 0.013	≈ 0
Tongue	MTGNN	$\mathbf{0.82 \pm 0.011}$	$\mathbf{0.80 \pm 0.008}$	
	FC-NN	0.77 ± 0.011	0.74 ± 0.011	3.7**e-**35
	RF	0.45 ± 0.027	0.58 ± 0.031	3.1**e-**155
	SVM	0.58 ± 0.021	0.55 ± 0.012	≈ 0

of our network. Finally, we evaluate the robustness of our method, including varying the functional parcellation used, corrupting the tumor segmentations, performing data augmentation, and sweeping our weighted cross entropy loss hyperparameter for detecting the language class.

We observe that including the specialized convolutional layers aids in identifying patterns within the eloquent cortex distribution. To assess whether our network learns reproducible patterns, we visually inspected the weights with the highest E2E filter magnitudes. In this manner, we can assess which network features are considered the most important. Fig.4.28 shows one example of a language connectivity hub that our model consistently identifies on the JHH dataset. We observe that this hub is lateralized on the left hemisphere, which is in line with the bulk of the JHH training data. Fig. 4.29 shows a

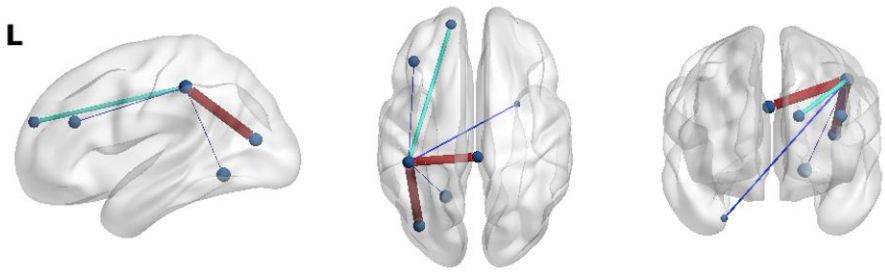

Figure 4.28: An example of a reproducible left-hemisphere only connectivity hub identified by our E2E filter when trained on the JHH dataset. We observe the nodes implicated resemble the activations in the language networks.

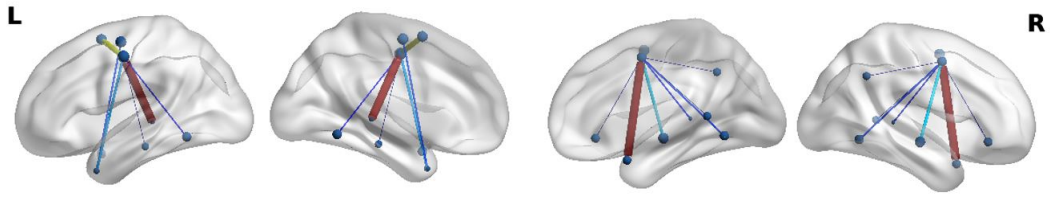

Figure 4.29: An example of a reproducible language network hub found in both hemispheres, when the MT-GNN is trained on the HCP dataset. The HCP story comprehension task is designed to target symmetric areas, which is captured in the identified language hub.

symmetric language network hub that is consistently found during the HCP experiments. This network is bilateral because the HCP task is designed to target symmetrical areas of the anterior temporal lobe (ATL) while the JHH task is not. Though the network has many layers responsible for feature extraction, we conjecture that the MT-GNN performance gains relative to the FC-NN baseline are likely due to these reproducible connectivity hubs, which aid the downstream classification task. However, as deep learning models can lack interpretability, we emphasize that our speculation is heuristic and should be taken with a grain of salt.

To highlight our localization performance, Fig.4.30 illustrates the correct

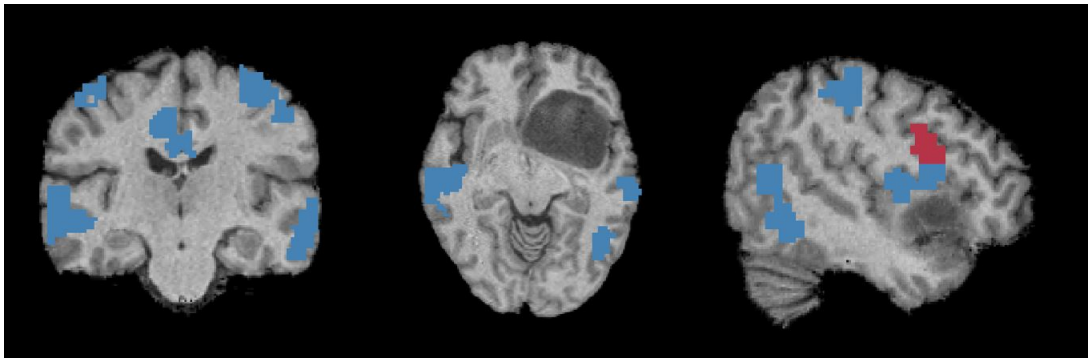

Figure 4.30: From (**L-R**) we show coronal, axial and saggital views of correct (blue) and incorrect (red) prediction by our model for the eloquent cortex in a challenging inferior frontal gyrus tumor case.

(blue) and false positive (red) detections by our MT-GNN in a patient with a large tumor in the inferior frontal gyrus. These results are aggregated across all four task branches of the model. We observe perfect sensitivity for the motor cortex localization (no false negative detections) and high accuracy for language despite the anatomical lesion.

4.4 Conclusion

In [36], we have demonstrated a GNN approach to identify the language and motor areas of eloquent cortex in brain tumor patients using rs-fMRI connectivity. Our model learns the resting-state functional signature of both the language and motor network within this tumor cohort by leveraging specialized convolutional filters that encode edge-node relationships within similarity matrices. With higher AUC for eloquent cortex detection, we prove that the features extracted from our GNN are more informative for this task than standard graph theoretic features and features extracted from a MLP. For language, we show that our model can correctly identify bilateral language

networks even when trained on only unilateral network cases.

Then, in [37], we extend the work to perform multi-task learning, which shares the network parameters among all four tasks for both cohorts. Our method shows a substantial improvement in a threefold manner: (1) we save a large number of parameters, which is essential when working with smaller clinical datasets, (2) we find a shared latent representation of the eloquent cortex functional systems, and (3) we reduce training time by a factor of three. Highlighted by the ablation study, we observe that the single GNN (SGNN) cannot localize the eloquent regions as well as the MT-GNN. Due to our multi-branch loss function, our model has access to more training data compared to the SGNN case. Also, compared to the SGNN, our network finds a shared latent representation that models the complex interactions between the eloquent cortex that eventually helps with simultaneous classification.

In comparison to baseline methods, our model achieves higher and statistically significant region-based localization performance on both a synthetic and real world clinical dataset. We show that our model can recover clinically challenging bilateral language cases when trained on unilateral cases. Our ablation study further demonstrates the value of the multi-task portion of our network. We evaluate the robustness of our method, including varying the functional parcellation used, corrupting the tumor segmentations, performing data augmentation, and sweeping our weighted cross entropy loss hyperparameter for detecting the language class. We explored different potential confounders' effect on model performance, various model optimization

strategies, and performance on the healthy HCP data. We observe no statistical significance in the correlation coefficients for each of the confounders. Regarding model optimization, we show confidence in the original model to generalize well to unseen testing data, as the overfit model or overfit model with dropout does not generalize as well. The healthy HCP result shows our method can identify localized functional subsystems of the eloquent cortex in healthy rs-fMRI scans. Finally, we showed qualitative results associated with our model, such as localization outputs on a challenging case and a feature analysis plotted on the brain.

Chapter 5

Eloquent cortex localization: dynamic connectivity and attention models

5.1 Introduction

In this chapter of the thesis, we present the work we have done on eloquent cortex localization that builds upon chapter 4, specifically with the use of dynamic connectivity analysis and various deep learning attention models. There is growing evidence in the field that functional connectivity patterns are not static, but evolve over time. In particular, studies have shown that individual functional systems are more strongly present during specific intervals of the rs-fMRI scan [189, 190]. The models presented in Chapter 4 of this thesis only make use of static connectivity, and thus are ignoring important temporal information that could aid in localization. Several studies have leveraged these dynamic connectivity patterns for classification. For example, the work in [97] uses a long-short term memory (LSTM) cell to learn time dependencies within the rs-fMRI to discriminate patients with autism from controls.

More recent work by [191] and [124] has shown that combining static and dynamic connectivity can achieve better patient versus control classification performance than either set of features alone. However, these works focus on group-level discrimination. We will leverage similar principles in this chapter to classify ROIs within a single patient.

We introduced a brief overview of attention models in chapter 2 of this thesis. Attention is a powerful mechanism that employs neural networks to hone in on the most salient or important part of the features or input to aid in the downstream task. In this chapter, we will not only explore dynamic connectivity analysis, but also explore attention models through the context of dynamic connectivity (temporal attention). Recent work in the deep learning literature has introduced the idea of spatial attention, which mimics information processing in the human visual system. For example, a 2D spatial attention model learns where in the image to focus, thus improving the quality of the learned representations [192]. The models presented in this chapter will sequentially add various attention models to improve localization.

5.1.1 Contributions

We present the work that we published in the Machine Learning for Clinical Neuroimaging (MLCN) workshop [38] as a part of MICCAI 2020 and our information processesing in medical imaging (IPMI) 2021 paper [39].

First, we propose a novel multi-task deep learning framework that uses both convolutional nerual networks (CNNs) and an LSTM attention network to extract and combine dynamic connectivity features for eloquent cortex

localization. The final stage of our model employs multi-task learning (MTL) to implicitly select the relevant time points for each network and simultaneously identify regions of the brain involved in language processing and motor functionality. Our model finds a shared representation between the cognitive networks of interest, which enables us to handle missing data. This coupling also reduces the number of model parameters, so that we can learn from limited patient data. We evaluate our framework on rs-fMRI data from 56 brain tumor patients while using task fMRI activations as surrogate ground-truth labels for training and testing. Our model achieves higher localization accuracies than a variety of baseline techniques, thus demonstrating its promise for preoperative mapping.

We then develop a spatiotemporal attention model to localize eloquent cortex from dynamic whole-brain rs-fMRI connectivity matrices. Unlike a 2D image, our "spatial" field corresponds to salient interactions in connectivity data, captured via graph-based convolutional filters. Our multi-scale spatial attention model pools three levels of granularity to amplify important interactions and suppress unnecessary ones. Then, our temporal attention mechanism selects key intervals of the dynamic input that are most relevant for either language or motor localization. Our model operates on a fine resolution parcellation and can handle missing training labels. We use t-fMRI activations as ground truth labels and validate our framework on rs-fMRI data from 100 subjects in the publicly available Human Connectome Project (HCP) [137] with artificially-inserted tumors as well as 60 subjects from an in-house dataset. Our model uniformly achieves higher localization accuracies than competing

baselines. Our attention mechanisms learn interpretable feature maps, thus demonstrating the promise of our model for preoperative mapping.

5.2 A Multi-Task Deep Learning Framework to Localize the Eloquent Cortex in Brain Tumor Patients Using Dynamic Functional Connectivity

5.2.1 Model

Our framework makes two underlying assumptions. First, while the anatomical boundaries of the eloquent cortex may shift across individuals, its functional connectivity with the rest of the brain will be preserved [33]. Second, the networks associated with the eloquent cortex phase in and out of synchrony across the rs-fMRI scan [193]. Hence, isolating these key time points will help to refine our localization. Fig. 5.1 illustrates our framework. In the top branch, we use specialized convolutional filters to capture rs-fMRI co-activation patterns from the dynamic connectivity matrices. In the bottom branch, we use an LSTM to identify key time points where the language and/or motor networks are more synchronous. We tie the activations from the LSTM branch of our model into our MTL classification problem via our specialized loss function.

5.2.1.1 Input Connectivity Matrices

We use the sliding window technique to obtain our connectivity matrices [17]. Let N be the number of brain regions in our parcellation, T be the total number of sliding windows (i.e., time points in our model), and $\{\mathbf{W}^t\}_{t=1}^{T} \in \mathbb{R}^{N \times N}$ be the dynamic similarity matrices. $\mathbf{W}^t$ is constructed from the input time

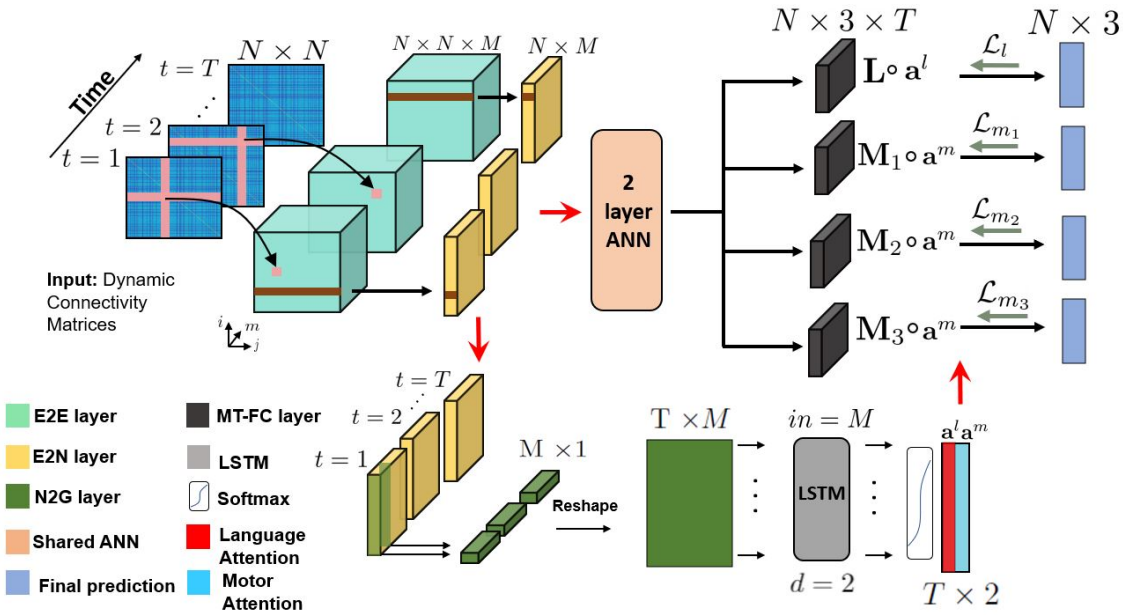

Figure 5.1: Top: Specialized convolutional layers identify dynamic patterns that are shared across the functional systems. **Bottom**: The dynamic features are input to an LSTM network to learn attention weights $\mathbf{a}^l$ (language) and $\mathbf{a}^m$ (motor). **Right**: MTL to classify the language ($\mathbf{L}$), finger ($\mathbf{M}_1$), foot ($\mathbf{M}_2$) and tongue ($\mathbf{M}_3$) networks.

courses $\{\mathbf{X}^t\}_{t=1}^T \in \mathbb{R}^{D \times N}$, where each $\mathbf{X}^t$ is a segment of the rs-fMRI obtained with window size D. The input $\mathbf{W}^t \in \mathbb{R}^{N \times N}$ is

$$\mathbf{W}^t = \exp\left[\frac{(\mathbf{X}^t)^T \mathbf{X}^t}{\epsilon} - 1\right] \tag{5.1}$$

where $\epsilon \geq 1$ is a user-specified parameter that controls decay speed [36]. Regarding the tumor, we follow the approach of [36] and treat the corresponding rows and columns of the simlarity matrix as “missing data" by fixing them to zero.

5.2.1.2 Representation Learning for Dynamic Connectivity

Our network leverages the specialized convolutional layers developed in [169] for static analysis. The edge-to-edge (E2E) layer in Fig. 5.1 acts across rows and columns of the input matrix $\mathbf{W}^t$. Mathematically, let $f \in \{1, \cdots, F\}$ be the E2E filter index, $\mathbf{r}^f \in \mathbb{R}^{1\times N}$ be the row filter f, $\mathbf{c}^f \in \mathbb{R}^{N\times 1}$ be the column filter f, $\mathbf{b} \in \mathbb{R}^{F\times 1}$ be the E2E bias, and $\boldsymbol{\phi}(.)$ be the activation function. For each time point t the feature map $\mathbf{H}^{f,t} \in \mathbb{R}^{N\times N}$ is computed as follows:

$$\mathbf{H}_{i,j}^{f,t} = \boldsymbol{\phi}\left(\sum_{n=1}^{N} \mathbf{r}_n^f \mathbf{W}_{i,n}^t + \mathbf{c}_n^f \mathbf{W}_{n,j}^t + \mathbf{b}_f\right). \tag{5.2}$$

As previously discussed, the E2E filter output $\mathbf{H}_{ij}^{f,t}$ for edge (i, j) extracts patterns associated with the neighborhood connectivity of node i and node j. The edge-to-node (E2N) filter in Fig. 5.1 is a 1D convolution along the columns of each feature map. Mathematically, let $\mathbf{g}^f \in \mathbb{R}^{N\times 1}$ be E2N filter f and $\mathbf{p} \in \mathbb{R}^{F\times 1}$ be the E2N bias. The E2N output $\mathbf{h}^{f,t} \in \mathbb{R}^{N\times 1}$ from input $\mathbf{H}^{f,t}$ is computed as

$$\mathbf{h}_i^{f,t} = \boldsymbol{\phi}\left(\sum_{n=1}^{N} \mathbf{g}_n^f \mathbf{H}_{i,n}^{f,t} + \mathbf{p}_f\right). \tag{5.3}$$

Following the convolutional layers in the top branch, we cascade two fully-connected (FC) layers to combine these learned topological features for our downstream multi-task classification. In the bottom branch, we use a node-to-graph (N2G) layer to extract features that will be input to our LSTM network. The N2G filter acts as a 1D convolution along the first dimension of the E2N output, effectively collapsing the node information to a low dimensional representation for each time point. Let $\mathbf{k}^f \in \mathbb{R}^{N\times 1}$ be N2G filter f and

$\mathbf{d} \in \mathbb{R}^{F\times 1}$ be the bias. The N2G filter gives a scalar output $q^{f,t}$ for each input $\mathbf{h}^{f,t}$ by

$$q^{f,t} = \boldsymbol{\phi}\left(\sum_{n=1}^{N} \mathbf{k}_n^f \cdot \mathbf{h}_n^{f,t} + \mathbf{d}_f\right). \tag{5.4}$$

5.2.1.3 Dynamic Attention Model

Per time point, we define $\mathbf{q}^t = [q^{1,t} \cdots q^{F,t}]$ and feed the vectors $\{\mathbf{q}^t\}_{t=1}^T$ into an LSTM module to learn attention weights for our classification problem. The LSTM adds a cell state to the basic recurrent neural network to help alleviate the vanishing gradient problem, essentially by accumulating state information over time [194]. LSTMs have demonstrated both predictive power for rs-fMRI analysis [97, 124] and the ability to identify different brain states [195]. We choose $d = 2$ as the output dimension, and perform a softmax over each column of the LSTM output to get the attention vectors $\mathbf{a}^l \in \mathbb{R}^{T\times 1}$ (language) and $\mathbf{a}^m \in \mathbb{R}^{T\times 1}$ (motor). These attention vectors provide information on which input connectivity matrices are more informative for identifying the language or motor networks. The attention model outputs are combined with the classifer during backpropogation in our novel loss function.

5.2.1.4 Multi-task Learning with Incomplete Data

The black blocks in Fig. 4.6 show the multi-task FC (MT-FC) layers, where we have four separate branches to identify the language, finger, foot, and tongue areas. Up until this point, there has been an entirely shared representation of the feature weights at each layer. Let $\mathbf{L}^t, \mathbf{M}_1^t, \mathbf{M}_2^t$, and $\mathbf{M}_3^t \in \mathbb{R}^{N\times 3}$ be the output of the language, finger, foot, and tongue MT-FC layers, respectively,

at time t. The $N \times 3$ matrix represents the region-wise assignment into one of three classes; eloquent, tumor, and background. As in [36], we introduce the tumor as its own learned class to remove any bias these regions may have introduced to the algorithm.

Following our previous models, we use a modified version of the risk-sensitive cross-entropy loss function [179, 36], which is designed to handle membership imbalance in multi-class problems. Let δ_c be the risk factor associated with class c. If δ_c is small, then we pay a smaller penalty for misclassifying samples that belong to class c. Due to a training set imbalance, we set different values for the language class (δ_c^l) and motor classes (δ_c^m) respectively. Let $\mathbf{Y}^l, \mathbf{Y}^{m_1}, \mathbf{Y}^{m_1}$, and $\mathbf{Y}^{m_3} \in \mathbb{R}^{N \times 3}$ be one-hot encoding matrices for the ground-truth class labels of the language and motor subnetworks. Notice that our framework allows for overlapping eloquent labels, as brain regions can be involved in multiple cognitive processes. Our loss function is the sum of four terms:

$$\mathcal{L}_\Theta(\{\mathbf{W}^t\}_{t=1}^T, \mathbf{Y}) = \sum_{n=1}^{N} \sum_{c=1}^{3} \Big[\underbrace{-\delta_c^l \log\Big(\sigma\Big(\sum_{t=1}^{T} \mathbf{L}_{n,c}^t \cdot \mathbf{a}^{l,t}\Big)\Big)\mathbf{Y}_{n,c}^l}_{\text{Language Loss } \mathcal{L}_l}$$

$$\underbrace{-\delta_c^m \log\Big(\sigma\Big(\sum_{t=1}^{T} \mathbf{M}_{1n,c}^t \cdot \mathbf{a}^{m,t}\Big)\Big)\mathbf{Y}_{n,c}^{m_1}}_{\text{Finger Loss } \mathcal{L}_{m_1}} \underbrace{-\delta_c^m \log\Big(\sigma\Big(\sum_{t=1}^{T} \mathbf{M}_{2n,c}^t \cdot \mathbf{a}^{m,t}\Big)\Big)\mathbf{Y}_{n,c}^{m_2}}_{\text{Foot Loss } \mathcal{L}_{m_2}} \tag{5.5}$$

$$\underbrace{-\delta_c^m \log\Big(\sigma\Big(\sum_{t=1}^{T} \mathbf{M}_{3n,c}^t \cdot \mathbf{a}^{m,t}\Big)\Big)\mathbf{Y}_{n,c}^{m_3}}_{\text{Tongue Loss } \mathcal{L}_{m_3}} \Big]$$

where $\sigma(\cdot)$ is the sigmoid function. Our loss in Eq.5.5 allows us to handle

missing information during training. For example, if we only have ground-truth labels for some of the functional systems, then we can freeze the other branches and just backpropagate the known loss terms. This partial backpropagation will continue to refine the shared representation, thus maximizing the amount of information mined from our training data. Note that our formulation is agnostic to the length of the rs-fMRI scan (i.e. T), which is useful in clinical practice.

5.2.1.5 Implementation details and baselines

We implement our network in PyTorch using the SGD optimizer with weight decay $= 5 \times 10^{-5}$ for parameter stability, and momentum $= 0.9$ to improve convergence. We train our model with learning rate $= 0.002$ and 300 epochs, which provides for reliable performance without overfitting. We used $D = 45$ and a stride length of 5 for the sliding window. We specified $F = 25$ feature maps in the convolutional branch, and 2 layers in our LSTM. The LeakyReLU with slope $= -0.1$ was used for $\boldsymbol{\phi}(.)$. Using cross validation, we set the cross-entropy weights to $\delta^m = (1.5, 0.5, 0.2)$, and $\delta^l = (2.25, 0.5, 0.2)$.

We compare the performance of our model against three baselines:

1. PCA + Multi-class linear SVM on dynamic connectivity matrices (SVM)
2. A multi-task GNN on static connectivity (MT-GNN)
3. A multi-task ANN with LSTM attention model (MT-ANN)

The first baseline is a traditional machine learning SVM approach to our problem. The MT-GNN operates on static connectivity and does not have an LSTM

module. We include the MT-GNN to observe the difference in performance with and without using dynamic information. The MT-ANN maintains the same number of parameters as our model but has fully-connected layers instead of convolutional layers. Therefore, the MT-ANN does not consider the network organization of the input dynamic connectivity matrices.

5.2.2 Experimental Results

5.2.2.1 Dataset and localization results

We evaluate the methods on rs-fMRI data from 56 brain tumor patients who underwent preoperative mapping at our institution. These patients also underwent t-fMRI scanning, which we use to derive pseudo ground-truth labels for training and validation. Our dataset includes the three motor and two language paradigms described in Chapter 2, which also includes preprocessing details.

We used the Craddocks atlas to obtain N=384 brain regions [59]. Tumor boundaries for each patient were manually delineated by a medical fellow using the MIPAV software package [128]. An ROI was determined as belonging to the eloquent class if a majority of its voxel membership coincided with that of the t-fMRI activation map. Tumor labels were determined in a similar fashion according to the MIPAV segmentations. A general linear model implemented in SPM8 was used to obtain t-fMRI activation maps.

We use 8-fold cross validation (CV) to quantify our eloquent cortex localization performance. Table 5.1 reports the eloquent per-class accuracy and the area under the receiver operating characteristic curve (AUC) for detecting the

Table 5.1: Class accuracy, overall accuracy, and ROC statistics. The number in the first column indicates number of patients who performed the task.

Task	Method	Eloquent	Overall	AUC
Language (56)	SVM	0.49	0.59	0.55
	MT-ANN	0.70	0.71	0.70
	MT-GNN	0.73	0.74	0.74
	Proposed	**0.85**	**0.81**	**0.80**
Finger (36)	SVM	0.54	0.61	0.57
	MT-ANN	0.73	0.75	0.74
	MT-GNN	0.87	**0.86**	**0.84**
	Proposed	**0.88**	0.85	**0.84**
Foot (17)	SVM	0.58	0.63	0.60
	MT-ANN	0.72	0.77	0.74
	MT-GNN	0.82	0.79	0.79
	Proposed	**0.86**	**0.85**	**0.82**
Tongue (39)	SVM	0.54	0.60	0.58
	MT-ANN	0.74	0.76	0.73
	MT-GNN	0.85	0.81	0.82
	Proposed	**0.87**	**0.83**	**0.84**

eloquent class on the testing data. Each MT-FC branch has separate metrics. Our proposed method has the best overall performance, as highlighted in bold. Even with attention from the LSTM layer, we observe that a fully-connected ANN still is sub-par for our task compared to using the specialized E2E, E2N, and N2G layers. Furthermore, our performance gains are most notable when classifying the language and foot networks. The former is particularly relevant for preoperative mapping, due to the difficulties in identifying the language network even with ECS [20, 196]. We observe that the inclusion of dynamic connectivity alongside the LSTM for temporal attention performs better than the MT-GNN model from chapter 4, suggesting that approaching this problem with dynamic connectivity analysis is favorable.

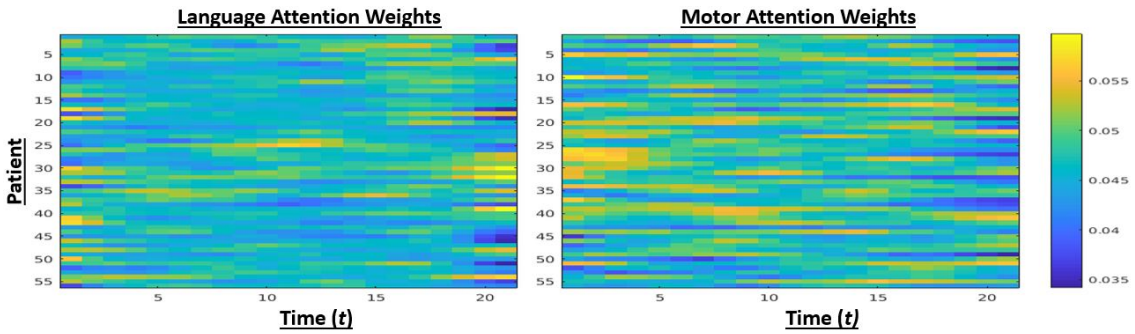

Figure 5.2: Language (**L**) and motor (**R**) attention weights for all patients.

5.2.2.2 Attention weight analysis and bilateral identification

As a side experiment, we plotted the attention weights recovered for each of the 56 subjects during the test phase. We did this to observe if there was a trend in the magnitude of the weights over time, and that the attention mechanism is the key novelty associated with this work. Fig.5.2 shows the language (left) and motor (right) attention vectors for all patients across time. We observe that both systems phase in and out, such that when one system is more active, the other is less active. This pattern lends credence to our hypothesis that identifying the critical intervals is key for localization. Hence, our model outperforms the static MT-GNN.

Finally, we test whether our model can recover a bilateral language network, even when this case is not present in the training data. Here, we trained the model on 51 left-hemisphere language network patients and tested on the remaining 5 bilateral patients. Our model correctly predicted bilateral parcels in all five subjects. Fig.5.3 shows ground truth (blue) and predicted language maps (yellow) for two example cases. The mean language class accuracy for these five cases was **0.72**. This is slighly lower than reported in Table 5.1 likely

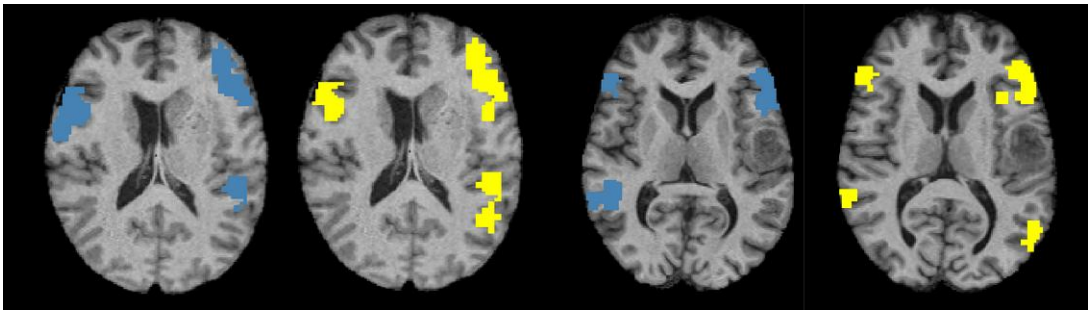

Figure 5.3: Ground truth (Blue) and predicted (Yellow) language labels for two subjects.

due to the mismatch in training information.

5.3 A Multi-Scale Spatial and Temporal Attention Network on Dynamic Connectivity to Localize The Eloquent Cortex in Brain Tumor Patients

In this section, we review the main findings from our IPMI 2021 paper, which incorporates a spatial attention mechanism alongside temporal attention mechanism to improve localization. We use multiple scales of spatial attention which operates on graph-based features extracted from the connectivity matrices, thus honing in on the inter-regional interactions that collectively define the eloquent cortex. This is the last model we developed for eloquent cortex localization, which builds upon our previous model with the addition of spatial attention and extra data for validation.

5.3.1 Model

Our framework begins with the same assumptions as our previous models, that is, the resting-state connectivity of the eloquent cortex with the rest of

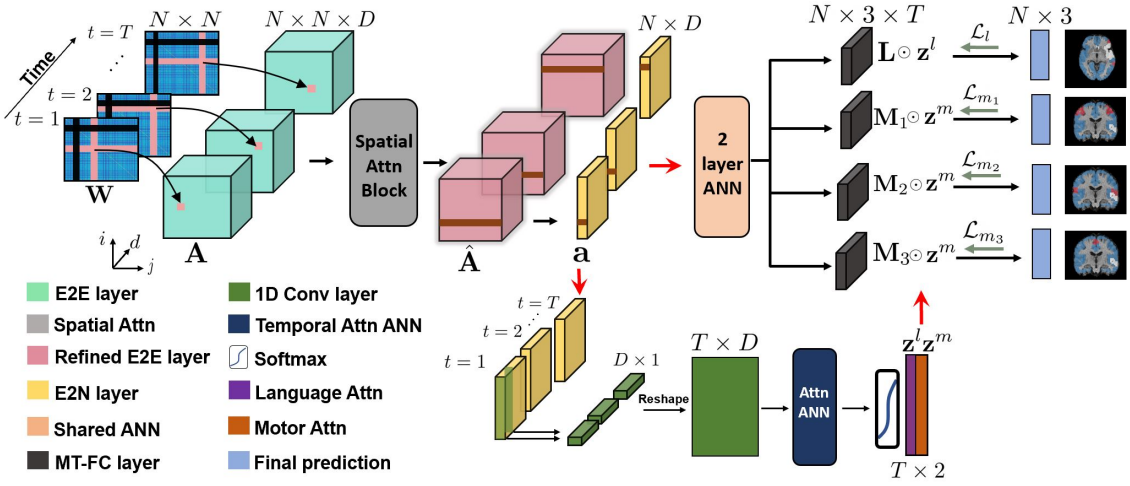

Figure 5.4: Top: Convolutional features extracted from dynamic connectivity are refined using a multi-scale spatial attention block. **Bottom**: The dynamic features are input to an ANN temporal attention network to learn weights $\mathbf{z}^l$ (language) and $\mathbf{z}^m$ (motor). **Right**: Multi-task learning to classify language ($\mathbf{L}$), finger ($\mathbf{M_1}$), tongue ($\mathbf{M_2}$), and foot ($\mathbf{M_3}$) subnetworks, where each subnetwork is a 3-class classification which is shown in red, white, and blue respectively on segmentation maps.

the brain is identifiable and consistent among subjects. Adding a layer of complexity, the eloquent cortex represents a relatively small portion of the brain. This is the motivation for our spatial attention mechanism, i.e., to zone in on the key connectivity patterns. Furthermore, the networks associated with the eloquent cortex will likely phase in and out of synchrony across the rs-fMRI scan [38]. Our temporal attention mechanism will track these changes. Fig. 5.4 shows our overall framework. As seen, we explicitly model the tumor in our dynamic similarity graph construction and feed this input into a deep neural network which uses specialized convolutional layers designed to handle connectome data [169].

Once again, we use the sliding window technique to construct our dynamic inputs [17]. Let N be the number of brain regions in our parcellation, T be

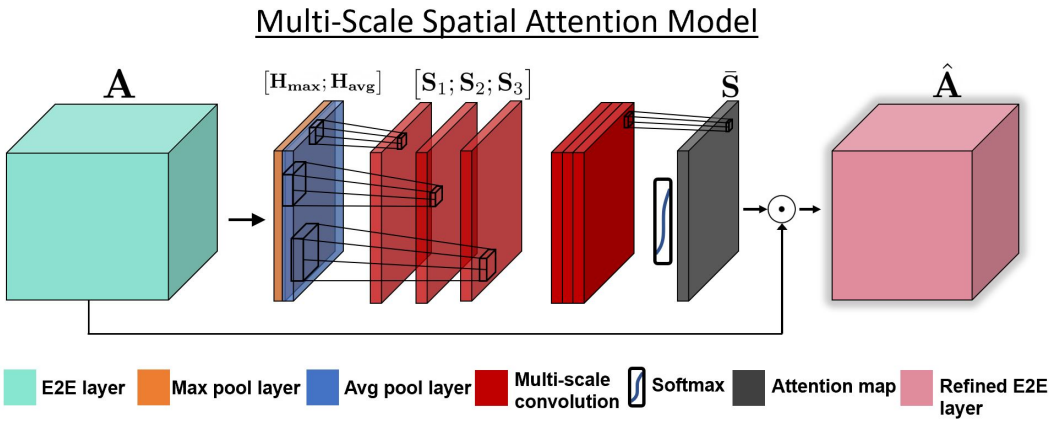

Figure 5.5: Our multi-scale spatial attention model extracts features from max pool and average pool features along the channel dimension. We use separate convolutional filters with increasing receptive field size to extract multi-scale features, and use a 1×1 convolution and softmax to obtain our spatial attention map $\bar{\mathbf{S}}$. This map is element-wise multiplied along the channel dimension of the original E2E features.

the total number of sliding windows (i.e., time points in our model), and $\{\mathbf{W}^t\}_{t=1}^T \in \mathbb{R}^{N \times N}$ be the dynamic similarity matrices. $\mathbf{W}^t$ is constructed from the normalized input time courses $\{\mathbf{X}^t\}_{t=1}^T \in \mathbb{R}^{G \times N}$, where each $\mathbf{X}^t$ is a segment of the rs-fMRI obtained with window size G. Formally, the input $\mathbf{W}^t \in \mathbb{R}^{N \times N}$ is

$$\mathbf{W}^t = \exp\left[(\mathbf{X}^t)^T \mathbf{X}^t - 1\right]. \tag{5.6}$$

where the tumor is handled similarly among all models.

5.3.1.1 Multi-scale Spatial Attention on Convolutional Features

Our network leverages the specialized convolutional layers developed in [169] for feature extraction on each of the dynamic inputs. The edge-to-edge (E2E) filter (pink in Fig. 5.4) acts across rows and columns of the input matrix $\mathbf{W}^t$. Mathematically, let $d \in \{1, \cdots, D\}$ be the E2E filter index, $\mathbf{r}^d \in \mathbb{R}^{1 \times N}$ be

the row filter d, $\mathbf{c}^d \in \mathbb{R}^{N\times 1}$ be the column filter d, $\mathbf{b} \in \mathbb{R}^{D\times 1}$ be the E2E bias, and $\boldsymbol{\phi}(.)$ be the activation function. For each time point t the feature map $\mathbf{A}^{d,t} \in \mathbb{R}^{N\times N}$ is computed as follows:

$$\mathbf{A}^{d,t}_{i,j} = \boldsymbol{\phi}\Big(\mathbf{W}^t_{i,:}(\mathbf{r}^d)^T + (\mathbf{c}^d)^T\mathbf{W}^t_{:,j} + \mathbf{b}_d\Big). \tag{5.7}$$

The E2E filter output $\mathbf{A}^{d,t}_{ij}$ for edge (i,j) extracts information associated with the connectivity of node i and node j with the rest of the graph. We use the same D E2E filters $\{\mathbf{r}^d, \mathbf{c}^d\}$ for each time point to standardize the feature computation.

Fig.5.5 illustrates our multi-scale spatial attention model. The attention model acts on the E2E features and implicitly learns “where" informative connectivity hubs are located for maximum downstream class separation. The multi-scale setup uses filters of different receptive field sizes to capture various levels of connectivity profiles within the E2E features [197]. Following [192], we apply an average pooling and max pooling operation along the feature map axis and concatenate them to generate an efficient feature descriptor. Mathematically,

$$\mathbf{H_{avg}} = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}\mathbf{A}^{d,t} \tag{5.8}$$

is the $N \times N$ average pool features and

$$\mathbf{H^{i,j}_{max}} = \max_{d,t}\mathbf{A}^{d,t}_{i,j} \tag{5.9}$$

is the $N \times N$ max pool features. Note that we extract the maximum and average activations across all feature maps and time points simultaneously. We then apply a multi-scale convolution to this feature descriptor, which implicitly

identifies the deviation of the maximum activation from the neighborhood average, thus highlighting informative regions to aid in downstream tasks [198].

We apply three separate convolutions with increasing filter sizes to the concatenated feature descriptor to obtain different scales of resolution of our analysis. The convolution outputs $\mathbf{S}_1, \mathbf{S}_2$ and $\mathbf{S}_3 \in \mathbb{R}^{N \times N}$ are computed using a 3×3, 7×7, and 11×11 kernel, respectively, on the concatenated maps $\left[\mathbf{H_{avg}}; \mathbf{H_{max}}\right]$. The convolutions include zero padding to maintain dimensionality. Each successive convolutional filter has an increasing receptive field size to help identify various connectivity hubs within the E2E layer. We obtain our spatial attention map $\bar{\mathbf{S}} \in \mathbb{R}^{N \times N}$ with an element-wise softmax operation on the weighted summation, derived using a 1×1 convolution with bias b, across the three scales;

$$\bar{\mathbf{S}} = \text{Softmax}\left(\sum_{i=1}^{3} w_i \mathbf{S}_i + b\right). \tag{5.10}$$

This weighted combination is designed to highlight salient hubs in the network which appear across different spatial scales. The softmax transforms our attention into a gating operation, which we use to refine our convolutional features $\mathbf{A}^{d,t}$ by element-wise multiplication with $\bar{\mathbf{S}}$. Let $\odot$ denote the Hadamard product. The refined features $\hat{\mathbf{A}}^{d,t} \in \mathbb{R}^{N \times N}$ are computed as

$$\hat{\mathbf{A}}^{d,t} = \mathbf{A}^{d,t} \odot \bar{\mathbf{S}}. \tag{5.11}$$

Finally, we condense our representation along the column dimension by using the edge-to-node (E2N) filter [169]. Our E2N filter (brown in Fig. 5.4)

performs a 1D convolution along the columns of each refined feature map to obtain region-wise representations. Mathematically, let the E2N output be $\mathbf{a}^{d,t} \in \mathbb{R}^{N\times 1}$ from input $\hat{\mathbf{A}}^{d,t}$. Again, we apply the same E2N filters to each time point. At a high level, the E2N computation is similar to that of graph-theoretic features, such as node degree. The E2N outputs are fed into both the temporal attention model (bottom branch of Fig. 5.4) and the multi-task node classifier (right branch of Fig. 5.4).

5.3.1.2 Temporal Attention Model

We use a 1D convolution to collapse the region-wise information into a low dimensional vector for our temporal attention network. Let $\mathbf{k}^d \in \mathbb{R}^{N\times 1}$ be the weight vector for filter d and $\mathbf{j} \in \mathbb{R}^{D\times 1}$ be the bias across all filters. A scalar output $q^{d,t}$ for each input $\mathbf{a}^{d,t}$ is obtained

$$q^{d,t} = \boldsymbol{\phi}\Big((\mathbf{k}^d)^T \mathbf{a}^{d,t} + \mathbf{j}_d\Big). \tag{5.12}$$

The resulting $T \times D$ matrix $[q^{d,t}]^T$ is fed into a fully-connected layer of two perceptrons with size D to extract our temporal attention weights. We obtain one language network attention vector $\mathbf{z}^l \in \mathbb{R}^{T\times 1}$ and one motor network attention vector $\mathbf{z}^m \in \mathbb{R}^{T\times 1}$, which learn the time intervals during which the corresponding eloquent subnetwork is more identifiable. The FC attention model is more flexible than a recurrent architecture and can be easily trained on small clinical datasets (<100 subjects). We observed that the FC attention shows a good trade-off between representation and robustness to training with a limited sample size.

5.3.1.3 Implementation details and baselines

We backpropagate the same loss function as Eq. 5.5, which uses a weighted cross-entropy term, four different branches, and a dot product between the attention vectors and node predictions. We implement our network in PyTorch using the SGD optimizer with weight decay $= 5 \times 10^{-5}$ for parameter stability, and momentum $= 0.9$ to improve convergence. We train our model with learning rate $= 0.005$ and 140 epochs, which provides for reliable performance without overfitting. We specified $D = 50$ feature maps in the convolutional branch. The LeakyReLU with slope $= -0.1$ was used for $\boldsymbol{\phi}(.)$.

We compare the performance of our model against three baselines:

1. Random forest on dynamic connectivity matrices (RF)
2. A fully-connected network with temporal attention (FC-tANN)
3. Same as proposed without spatial attention (w/o sp. attn.)

The first baseline is a traditional machine learning RF approach to our problem. The FC-tANN maintains the same number of parameters as our model but has fully-connected layers instead of convolutional layers. Finally, we compare against our same architecture without spatial attention to observe the performance gain of focusing on different neighborhoods. To avoid biasing performance, we selected the hyperparameters using a development set of 100 subjects downloaded from the Human Connectome Project (HCP). The final settings are: $\boldsymbol{\delta}^m = (1.48, 0.44, 0.18), \boldsymbol{\delta}^l = (2.16, 0.44, 0.18)$ for proposed, $\boldsymbol{\delta}^m = (1.57, 0.42, 0.22), \boldsymbol{\delta}^l = (2.31, 0.42, 0.22)$ for FC-tANN and $\boldsymbol{\delta}^m = (1.51, 0.46, 0.19), \boldsymbol{\delta}^l = (2.22, 0.46, 0.19)$ for w/o sp. attn.

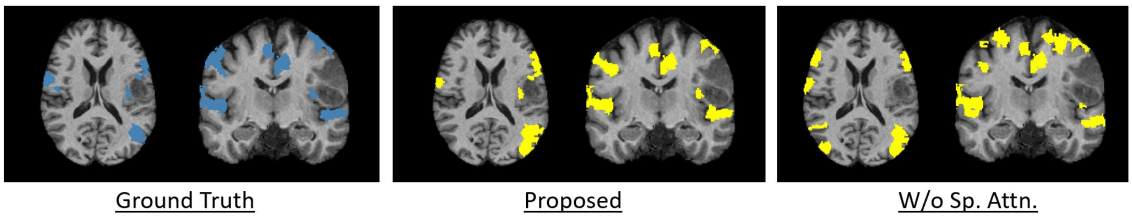

Figure 5.6: Ground truth (blue) and predicted (yellow) for a bilateral language subject.

5.3.2 Experimental Results

5.3.2.1 Dataset and localization results

We evaluate the methods on rs-fMRI data from an additional HCP cohort [137] in which we artificially insert "fake tumors" by zeroing out entries of the connectivity matrix, and an in-house brain tumor dataset. Similarly to previous experiments, we use the t-fMRI acquired to act as ground truth labels. All pre processing details are in chapter 2 of this thesis. We used the Schaefer atlas to obtain $N = 1000$ brain regions [199], which is on par with the resolution of eloquent areas we are trying to detect. An ROI was determined as belonging to the eloquent class if a majority of its voxel membership coincided with that of the t-fMRI activation map. Tumor labels were determined in a similar fashion according to the MIPAV segmentations.

We use 10-fold cross-validation to evaluate each method. Table 1 shows the performance metrics for detecting the eloquent class. In the second column, the number next to the task refers to the number of subjects whom we have training labels. As highlighted in bold, our proposed method outperforms the baseline algorithms in nearly all cases. We observe that the spatial attention model improves the specificity by improving the ratio of true negatives to

Table 5.2: Overall accuracy, and ROC statistics. The number in the second column indicates number of patients who performed the task.

Dataset	Task	Method	Accuracy	Sens.	Spec.	F1	AUC
HCP	Language (100)	RF	0.58	0.32	0.55	0.42	0.5
		FC-tANN	0.65	0.61	0.58	0.59	0.64
		w/o Sp. Attn.	0.77	0.73	0.68	0.69	0.72
		Proposed	**0.83**	**0.79**	**0.81**	**0.82**	**0.80**
	Finger (100)	RF	0.70	0.53	0.67	0.64	0.56
		FC-tANN	0.76	0.70	0.72	0.73	0.72
		w/o Sp. Attn.	0.87	0.83	0.78	0.80	0.86
		Proposed	**0.91**	**0.86**	**0.85**	**0.85**	**0.88**
	Foot (100)	RF	0.67	0.48	0.65	0.62	0.53
		FC-tANN	0.79	0.77	0.69	0.73	0.76
		w/o Sp. Attn.	0.86	0.86	0.83	0.84	0.85
		Proposed	**0.90**	**0.87**	**0.86**	**0.86**	**0.88**
	Tongue (100)	RF	0.70	0.46	0.68	0.63	0.53
		FC-tANN	0.75	0.72	0.68	0.72	0.73
		w/o Sp. Attn.	0.81	0.83	0.80	0.81	0.81
		Proposed	**0.89**	**0.87**	**0.85**	**0.85**	**0.86**
In-house	Language (60)	RF	0.65	0.40	0.66	0.59	0.53
		FC-tANN	0.78	0.76	0.70	0.71	0.73
		w/o Sp. Attn.	0.84	0.85	0.74	0.79	0.82
		Proposed	**0.93**	**0.91**	**0.85**	**0.87**	**0.91**
	Finger (36)	RF	0.67	0.43	0.67	0.61	0.55
		FC-tANN	0.76	0.75	0.69	0.71	0.77
		w/o Sp. Attn.	0.88	**0.88**	0.79	0.82	0.85
		Proposed	**0.91**	**0.88**	**0.85**	**0.84**	**0.89**
	Foot (17)	RF	0.68	0.49	0.65	0.60	0.56
		FC-tANN	0.79	0.73	0.68	0.72	0.75
		w/o Sp. Attn.	0.86	0.86	0.78	0.80	0.82
		Proposed	**0.89**	**0.87**	**0.83**	**0.84**	**0.86**
	Tongue (39)	RF	0.69	0.38	0.70	0.64	0.52
		FC-tANN	0.79	0.78	0.71	0.74	0.76
		w/o Sp. Attn.	0.86	0.85	0.77	0.81	0.84
		Proposed	**0.90**	**0.87**	**0.82**	**0.84**	**0.87**

false positives. Our performance gains are most notable regarding the language network, which is arguably the most challenging rea to localize during preoperative mapping. Fig. 5.6 shows the ground truth (blue) and predicted (yellow) for all four systems in a challenging bilateral language subject, with both the proposed and w/o spatial attention methods. The model without spatial attention overpredicts the right-hemipshere language nodes, and misses various parts of the motor strip. Our model can localize functional regions

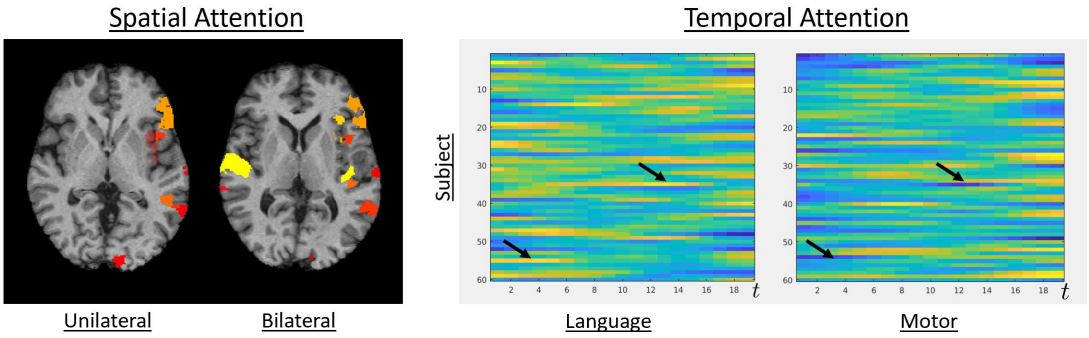

Figure 5.7: Left: Heat map for the nodes with highest total spatial attention for a unilateral and a bilateral language subject. **Right:** Temporal attention weights for language and motor networks. The black arrows indicate networks phasing in and out with each other.

right on the tumor boundary that the baseline method misses as well, which is relevant for clinical practice.

5.3.2.2 Feature Analysis

To better understand how the attention models improve the localization performance, Fig. 5.7 illustrates the spatial attention (left) and temporal attention weights (right) for our in-house dataset. These plots are generated by summing across the rows of the attention map $\bar{\mathbf{S}}$ and plotting the top ten nodes in one unilateral language and one bilateral language case. The spatial attention model is accurately able to capture right hemisphere activation in the bilateral case while correctly omitting this region in the unilateral case. This lateralization ability may be why localization performance increases for the language network. On the right-hand side of Fig.5.7, we show the temporal attention weights for both language and motor networks across all patients and time. The language and motor networks phase in and out at different times, which improves localization by identifying important time intervals within the scan

for each network.

5.4 Conclusion

In this portion of the thesis, we extend upon Chapter 4 by improving the localization performance of our network by considering two main modelling additions: (1) dynamic connectivity with temporal attention models, and (2) spatial attention models to refine the convolutional features and hone in on specific spatially continuous hubs. We present a novel deep learning framework that leverages specialized convolutional layers, multi-scale spatial attention, temporal attention, and multi-task learning to identify critical regions of the eloquent cortex in tumor patients using dynamic resting-state connectivity. We validate our method on a real in-house dataset and a synthetic dataset to show generalizability of our method. We outperform machine and deep learning baselines by a large margin. Finally, we show the spatial and temporal attention features, which can be important biomarkers for simultaneous language and motor network identification.

This concludes the portion of the thesis dedicated to eloquent cortex localization. Through ideation from current literature and deep learning trends and experimentation, we improved localization performance dramatically from our first static GNN presented in [36] to our last spatiotemporal attention model applied to dynamic connectivity [39]. We assessed robustness with a vast number of experiments on our static MT-GNN presented in our MEDIA paper [37]. Future work includes exploring the multi-modal inclusion of DTI imaging to incorporate structural connectivity pathways to the overall

network analysis.

Chapter 6

Epileptogenic Zone localization

6.1 Introduction

In this chapter, we present our work on automated epileptogenic zone (EZ) localization from rs-fMRI connectivity. Epilepsy is one of the most common neurological disorders, affecting around 50 million people worldwide, and is linked to a fivefold increase in mortality [200]. Epilepsy onset often occurs in childhood, and approximately one third of all patients have a medication refractory course that is associated with a disabling cumulative effect on neurocognitive development, lost productivity for the family, and increased societal and healthcare costs [201]. Surgical treatment is a safe and effective therapeutic approach for medication refractory epilepsy, that can provide seizure freedom and improved quality of life [202]. However, surgical candidacy and treatment outcomes are dependent on accurate localization of the EZ as defined by clinical, radiographic (magnetic resonance imaging, MRI) and physiological (electroencephalography, EEG) features [203]. Long-term treatment failures following surgery most commonly occur due to inaccurate

identification and resection of the EZ. Invasive monitoring using implanted intracranial electrodes can provide more accurate EZ localization that can help plan treatment, but is associated with surgical risks [29]. Hence an accurate EZ localization hypothesis is the foundation for effective and safe treatment in epilepsy [204, 205], and is the most important prognostic determinant for long term treatment outcomes.

6.1.1 Automated Methods for EZ Localization

Over the past two decades, there has been an increasing focus on automated methods for EZ localization. These methods are most often based on electrographic (EEG) or neuroimaging (structural MRI) modalities and can help reduce interpretative differences and delays in clinical reviews.

Automated methods for EEG localization have largely focused on improving the spatial resolution of the EEG sensors by deconvolving the signals into current dipoles or distributed sources at the millimeter scale [206, 207]. Going one step further, EEG data can be combined with noninvasive magnetoencephalography (MEG) for improved source estimation [208, 209]. Recent studies have demonstrated the translational promise of such methods. However, from a modeling standpoint, these *inverse solvers* require careful annotations of the seizure interval and are sensitive to physiological noise and the underlying head model [210, 211]. More importantly, they rely on high-density recordings of >50 EEG/MEG channels. From a logistical standpoint, the current standard-of-care for long-term EEG monitoring is the 10-20

electrode placement system [212], which contains fewer than 20 EEG channels distributed across the scalp. This resolution is insufficient for accurate and fine-grained inverse source estimation. Moreover, only 27% of epilepsy centers in the United States have access to and regularly utilize MEG [213]. Thus, while inverse source mapping remains a valuable direction of research with tremendous potential for presurgical evaluation, these methods are not amenable to most clinical workflows. Recently, Temple University Hospital (TUH) released a large public EEG dataset, which has spurred interest in seizure type classification [214, 215, 216], where the goal is to predict the epilepsy subtype from scalp EEG. While this task provides more information than seizure detection and is less reliant on human annotations than inverse source localization, the categories (focal, generalized, complex partial, absence, etc.) are too broad to accurately pinpoint the EZ.

In contrast to EEG, automated methods for MRI localization aim to identify epileptogenic lesions including Focal Cortical Dysplasias (FCDs), that are often difficult to radiographically identify on clinical imaging. Traditionally, these methods were implemented as a two-stage procedure. First, image-based features are extracted from the MRI data, such as cortical thickness, intensity, texture, asymmetry, and voxel-based morphometry [217, 218, 219, 220, 221]. Second, each voxel is classified as normal or FCD using statistical or machine learning algorithms. While these methods work well on large FCD cohorts, they tend to be unreliable for nonlesional patients [222]. In addition, epileptogenic lesions are diverse and can involve cortical, subcortical white matter [223, 224] and vascular abnormalities [225, 226, 227], which are better

suited to other data modalities [228, 229, 230].

6.1.2 Connectivity as a Biomarker for Epilepsy

Recent neuroimaging studies of epilepsy have implicated global brain network changes in seizure generation and disease progression. Accordingly, epilepsy is increasingly viewed as a network disorder that affects regional and global connectivity [223, 231, 232, 233, 234]. Diffusion MRI (d-MRI) assesses white matter properties based on free water diffusion [235] and informs us on the structure of brain networks [236]. In contrast, resting-state functional MRI (rs-fMRI) quantifies the temporal synchrony between brain regions by measuring changes in low frequency BOLD fluctuations. Alterations in structural and functional network properties have been linked to disease onset [237], duration [238], and treatment outcomes [231, 232] in epilepsy. For example, our group has investigated functional topology in subjects with epilepsy [239], demonstrating functional reorganization with a shift of network hubs to the contralateral hemisphere in temporal onset epilepsy [240].

Support Vector Machine-based analysis can discriminate network features in temporal epilepsy and healthy control subjects. In addition, neural connectivity patterns can help predict neuropsychological measures that assess language and memory function. Notably, global connectome changes in epilepsy are associated with a decrement in neurocognitive phenotype [241]. While promising, these connectivity studies are restricted to predefined structural and functional systems and careful patient subtyping (e.g., temporal lobe epilepsy). In addition, the results are correlative and do not quantify how

well the biomarkers would generalize to new patients. Recently, a few seminal studies have explored prospective EZ localization from rs-fMRI connectivity, which is presented in Section 2.3.2.2 in this thesis (background). As a recap, ICA based methods are not ideal due to lack of being automated and requiring manual intervention.

6.1.3 Contributions

We present the findings published in our IEEE transactions in biomedical engineering (TBME) paper [40] and our international symposium of biomedical imaging (ISBI) 2023 paper [41]. In [40], we introduce the first deep learning model for EZ localization using interictal rs-fMRI connectivity. The underlying assumption of our work is that the chronic and recurrent seizure activity causes subtle and distributed changes in functional connectivity across the brain. In contrast to rule-based approaches, we used supervised learning to automatically mine and leverage complex relationships in the rs-fMRI data for robust and generalizable EZ identification. Our model, which we call DeepEZ, takes as input a whole-brain connectivity graph, where nodes correspond to regions in our brain parcellation and edges denote the functional connectivity between regions. From here, DeepEZ uses graph convolutional networks (GCNs) [242, 243] to implicitly track information flow along expected anatomical pathways and fully-connected layers to classify each node (i.e., region) as belonging to the EZ or not. We encode anatomical information using d-MRI tractography, which is often viewed as the anatomical substrate for functional signaling in the brain [244, 245].

DeepEZ also incorporates the findings of previous works via an asymmetry term in the loss function to encourage lateralized predictions and a learned subject-specific bias to mitigate environmental confounds. We validate DeepEZ on a heterogeneous dataset of 14 pediatric epilepsy patients collected at the University of Wisconsin (UW) Madison. We demonstrate the DeepEZ outperforms the ICA methods of [31, 80] and ablated versions of the network. We also rigorously evaluate the sensitivity of DeepEZ to parcellation size, the number of network layers, hyperparameter tuning, and data augmentation. Taken together, our results highlight the promise of using rs-fMRI connectivity as a complementary source of information to localize the EZ in presurgical epilepsy patients.

Recent work in rs-fMRI literature has increasingly leveraged the dynamic evolution of connectivity information to improve predictive performance [17]. For example, we used an LSTM network as a temporal attention module to improve localization of eloquent cortex in brain tumor patients. The method of [246] uses a transformer network applied to dynamic functional connectivity (dFC) to predict the brain age of patients with Alzheimer's disease. The next model presented after DeepEZ builds upon DeepEZ via incorporating dFC and attention mechanisms.

In [41], we present the first deep learning model to localize the EZ in focal epilepsy patients based on dFC. Our deep network architecture uses an anatomically-regularized graph convolutional network (GCN) for feature extraction. From here, a transformer network learns a temporal attention vector, which identifies relevant time windows of the rs-fMRI scan that aid

in localization. Following [142], we train our network entirely on simulated data derived from the Human Connectome Project (HCP), and we test it on a clinical epilepsy dataset from the University of Wisconsin Madison. We demonstrate the significantly improved performance of our framework, as compared to both ablated versions of our model and the DeepEZ method. Our results highlight the promise of using rs-fMRI connectivity for preoperative EZ localization.

6.2 DeepEZ: A Graph Convolutional Network for Automated Epileptogenic Zone Localization from Resting-State fMRI Connectivity

DeepEZ is designed under the assumption that there are subtle but widespread connectivity patterns associated with the EZ. Inspired by the rs-fMRI literature, we use a weighted similarity matrix to capture whole-brain rs-fMRI connectivity [36, 164, 165]. Our DeepEZ architecture exploits the topological properties of rs-fMRI connectivity data via a set of graph convolutions. To integrate biological knowledge, we use structural connectivity, derived from d-MRI tractography, to define the underlying graph, thus emphasizing signal propagation along anatomical pathways. DeepEZ includes a subject-specific detection bias to account for patient differences and improve generalizability. Finally, we incorporate a clinically relevant asymmetry term into our loss function to provide crucial lateralization information.

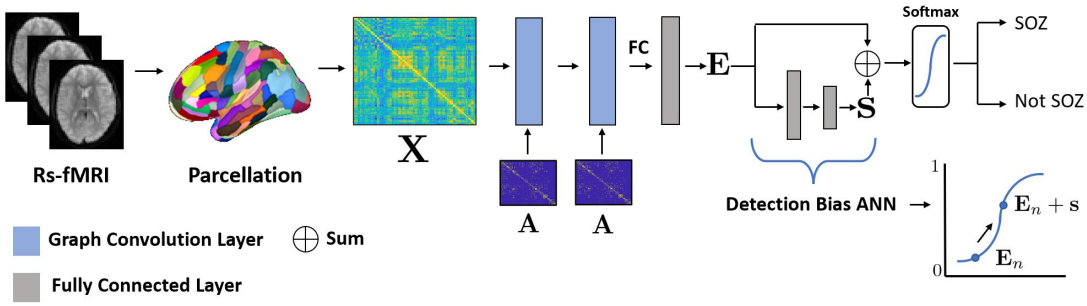

Figure 6.1: The overview of our model schematic. First, we apply a parcellation to the rs-fMRI and construct the subject specific functional connectivity information $\mathbf{X}$. Our network contains two graph convolution layers which include the adjacency matrix $\mathbf{A}$. The network uses an artificial neural network (ANN) for node classification. To improve detection of the EZ class, we added a separate ANN to learn a subject-specific bias term $\mathbf{s}$, which is added to the node-wise predictions $\mathbf{E}$. Our model classifies each ROI from the parcellation as either belonging to the EZ or not.

6.2.1 Model

6.2.1.1 Graph Convolution Network

Fig. 6.1 provides a graphical overview of our DeepEZ framework. Formally, let N be the number of brain regions in our parcellation and T be the number of time points for a rs-fMRI scan. We define $\mathbf{m}_i \in \mathbb{R}^{T\times 1}$ as the average time series extracted from region i, normalized to have zero mean and unit variance. Following the work of [36], we construct the input functional connectivity matrix as follows:

$$\mathbf{X} = \exp\left[\mathbf{M}^T\mathbf{M} - 1\right] \tag{6.1}$$

where $\mathbf{M} \in \mathbb{R}^{T\times N}$ aggregates the region-wise time series $\mathbf{m}_i$ as columns. Eq. 6.1 ensures non-negative input values, such that anti-correlated regions have connectivity close to zero, and highly correlated regions have connectivity close to one.

As shown in Fig. 6.1, DeepEZ processes the input data via two spatial

graph convolutions. Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ be a binary adjacency matrix used for spatial graph filtering [94]. As described in the previous section, we use d-MRI tractography to construct $\mathbf{A}$. In this case, an entry $\mathbf{A}_{ij} = 1$ denotes an anatomical pathway connecting regions i and j. Each graph convolution produces an activation map $\mathbf{H}_l \in \mathbb{R}^{N \times F_l}$, where $l \in \{1, 2\}$ denotes the layer number. The learnable parameters in each graph convolution are a weight matrix $\mathbf{W}_l \in \mathbb{R}^{F_l \times F_{l+1}}$ and a constant bias $\mathbf{b}_l \in \mathbb{R}^{1 \times F_{l+1}}$. The activation maps are generated via the layer propagation rule:

$$\mathbf{H}_1 = \boldsymbol{\phi}\Big(\mathbf{A}\mathbf{X}\mathbf{W}_1 + \mathbf{b}_1\Big) \tag{6.2}$$

$$\mathbf{H}_2 = \boldsymbol{\phi}\Big(\mathbf{A}\mathbf{H}_1\mathbf{W}_2 + \mathbf{b}_2\Big) \tag{6.3}$$

The multiplication by $\mathbf{A}$ in Eqs. (6.2)-(6.3) aggregates the region-wise representation based on their direct neighborhood [94].

6.2.1.2 Subject-Specific Detection Bias

We treat EZ identification as a two-class classification problem, where each region i classified as either belonging to the EZ or not. Here, the output $\mathbf{H}_2$ of our GCN cascade is fed through a fully-connected layer to obtain $\mathbf{E} \in \mathbb{R}^{N \times 2}$

$$\mathbf{E} = \boldsymbol{\phi}\Big(\mathbf{H}_2\mathbf{W}_{fc}\Big). \tag{6.4}$$

Similar to the graph convolutions, the weight matrix $\mathbf{W}_{fc} \in \mathbb{R}^{F_2 \times 2}$ is learned during training.

One challenge with our clinical rs-fMRI dataset is heterogeneity, both in

the EZ locations and in the data acquisition procedures (e.g., scanner type). To improve detection of the EZ class, we introduce a novel concept known as subject-specific detection bias (SSDB), which helps to mitigate variation in the input data distributions. The SSDB $\mathbf{s} \in \mathbb{R}^{1\times 2}$ is learned via a simple 2-layer artificial neural network (ANN) and is added to each row of $\mathbf{E}$ to obtain the final predictions. Mathematically, let $\{\mathbf{G}_l, \mathbf{c}_l\}$ denote the weight matrix and constant offset for each layer $l \in \{1, 2\}$ of the ANN. Our subject-specific bias term $\mathbf{s}$ is computed as follows:

$$\mathbf{s} = \boldsymbol{\phi}\Big(\mathbf{G}_2\boldsymbol{\phi}\Big(\mathbf{G}_1\mathbf{E} + \mathbf{c}_1\Big) + \mathbf{c}_2\Big). \tag{6.5}$$

An illustration of the effect of the SSDB on predicting whether a region n belongs to the EZ class is shown in the bottom right of Fig. 6.1. Empirically, we observe that the bias improves the sensitivity of detecting the EZ class. Following the SSDB addition, a softmax function is applied and each region is classified as belonging to the EZ or not using a max operator.

6.2.1.3 EZ Classification via Weighted Class Prediction and Contralateral Loss Function

There exists a large class imbalance in our dataset, as on average 7.3 % of the regions lie within the resection boundary that denotes the EZ. Since the GCN layers are designed to operate upon a whole-brain connectivity matrix, traditional data augmentation techniques would not solve our class imbalance problem. Following the work of [39, 37], we train our model with a modified Risk-Sensitive Cross-Entropy loss function [179], which is designed to handle a class membership imbalance. Formally, let δ_i be the risk associated with

class i. If δ_i is large, then we pay a larger penalty for misclassifying samples belonging to class i.

Beyond the class imbalance, it has been shown that contralateral areas of the brain have high rs-fMRI correlation [247, 248], often causing them to be treated similarly in downstream analyses. In contrast, we expect the EZ bto be lateralized [249, 31]. We leverage this asymmetry in the second term of our DeepEZ loss function by specifying that regions contralateral to the predicted EZ should be classified as normal.

Let N_e denote the nodes that belong to the EZ (labeled without loss of generality as class #2), and let $c(n)$ denote the contralateral counterpart to region n. Our training loss function consists of the following two terms:

$$\mathcal{L} = \underbrace{-\sum_{n=1}^{N}\sum_{i=1}^{2}\delta_i y_{n,i}\log\hat{y}_{n,i}}_{\text{Weighted Cross Entropy}} \underbrace{-\lambda\frac{1}{N_e}\sum_{n\in N_e}\left(\hat{y}_{n,2}-\hat{y}_{c(n),2}\right)}_{\text{EZ Contralateral Term}}. \tag{6.6}$$

The quantities $\hat{y}_{n,i}$ in Eq. 6.6 denote the DeepEZ prediction for the baseline ($i = 1$) and EZ ($i = 2$) classes at each region n. As seen, the first term of Eq. 6.6 accounts for the class imbalance, and the second term enforces hemispheric asymmetry in the final EZ predictions. Finally, λ balances the contributions of the two loss terms.

6.2.1.4 Implementation Details

We implement DeepEZ in PyTorch [250] using the Adam optimizer with weight decay (wd) and ϵ for regularization. The LeakyReLU $(x) = \max(0, x) + 0.1\cdot\min(0, x)$ activation function is applied at each hidden layer of the network

Table 6.1: Hyperparameters determined via cross validation on a separate cohort drawn from the HCP dataset.

Parameter	Value	Parameter	Value
(δ_1, δ_2)	$(0.29, 1.52)$	λ	0.017
lr	0.005	Epochs	200
ϵ	1×10^{-8}	wd	5×10^{-5}

in Fig. 6.1. A softmax activation is applied at the final layer for region-wise classification.

To prevent undue bias, we tune the hyperparameters $\delta_1, \delta_2, \lambda$ in Eq. 6.6 and the Adam optimization routine based on 50 subjects drawn from the HCP dataset. Specifically, we randomly selected a portion of the brain regions in each subject to denote an "artificial EZ". We then use 10-fold cross validation to fix the hyperparameters used in all experiments. For δ_1 and δ_2, we performed a coarse grid search from $0 - 10$ in increments of of 10^{-1} until we found a suitable range of performance. We then performed a finer grid search in increments of 10^{-2}. For the parameter λ, we performed a fine grid search over $0 - 1$ in increments of 10^{-3}. Table 6.1 reports the resulting parameter values for our final DeepEZ implementation.

6.2.2 Experimental results

6.2.2.1 Dataset and baseline algorithms

Our dataset consists of preoperative functional and postoperative structural MRI scans from 14 pediatric subjects with focal epilepsy that underwent a EZ resection procedure at UW Madison. Preoperative Rs-MRI (rs-fMRI) data and postoperative T1-weighted structural images were acquired. We manually

delineate the resection zone and use this boundary to define pseudo ground truth EZ labels for training and evaluation after applying the Brainnetomme atlas. All preprocessing details and a summary of seizure outcome of patients is found in Section 2.5.3 of this thesis. The GCNs in DeepEZ rely on a structural connectivity profile. To ensure consistency across subjects, we derive the graph from d-MRI tractography of 50 subjects from the Human Connectome Project (HCP) dataset [137]. The d-MRI data for each subject was preprocessed using the pipeline of [143] to obtain individual structural connectivity matrices based on the Brainnetome atlas. We compare DeepEZ against three competing methods from the literature and seven ablated versions of our model:

- ICA approach of [31] (ICA1)
- ICA approach of [80] (ICA2)
- The BrainNetCNN [169] adopted for region-wise classification (BN-CNN)
- Ablation #1: No SCT and No SSDB (GCN)
- Ablation #2: No SSDB (GCN-SCT)
- Ablation #3: No SCT (GCN-SSDB)
- Ablation #4: DeepEZ with an identity matrix replacing $\mathbf{A}$ (GCN-I)
- Ablation #5: DeepEZ with patient-specific d-MRI matrices replacing $\mathbf{A}$ (GCN-$\mathbf{A}_{subj}$)
- Ablation #6: DeepEZ with a randomly sampled matrix replacing $\mathbf{A}$ (GCN-$\mathbf{A}_{rand}$)

- Ablation #7: DeepEZ with topology preserved matrix replacing $\mathbf{A}$ (GCN-$\mathbf{A}_{top}$)

The first baseline is a traditional machine learning approach for EZ localization described in [31] Specifically, the ICA method extracts features from each independent component (IC) and then classifies each IC as belonging to the EZ or not via a linear support vector machine. The IC farthest from the boundary is selected as the final EZ for that patient. We chose this ICA baseline because it performed the best in the meta-analysis of [251], which compares seven different ICA methods for EZ localization using rs-fMRI. The second baseline described in [80] employs a screening process where ICs are sequentially discarded based on rules such as contralateral correlation and power spectrum density and the remaining ICs are considered belonging to the EZ class. We note that the original method is not fully automated. For example, visual inspection was used on a subject that had multiple independent components (ICs) that survived the rule-based screening process. In an effort to provide fair comparison across methods, we automate the work of [80] by combining the predictions of any ICs that pass the rule-based screening.

The third baseline is a modified version of the BrainNetCNN model developed in [169], which was originally designed to regress cognitive scores from structural connectivity matrices. The BrainNetCNN architecture uses cross-shaped convolutional filters to leverage topological relationships in connectivity data. We have modified the final layers of the original architecture to perform region-wise classification input rather than patient-level phenotypic prediction.

Our next three baselines focus on the EZ contralateral term (SCT) and the subject specific detection bias (SSDB) in our DeepEZ framework. As seen, we systematically ablate the components to determine the performance gain derived from each one. The last four baselines use the same architecture and loss function but vary the anatomical connectivity matrix $\mathbf{A}$ in the spatial graph convolutions. The GCN-I baseline replaces $\mathbf{A}$ with the identity matrix, effectively removing the anatomical regularization from our model. The GCN-$\mathbf{A}_{subj}$ baseline replaces $\mathbf{A}$ with personalized graphs computed from the patient-specific d-MRI. Due to variations in the data and tractography outputs, the edges in $\mathbf{A}_{subj}$ vary across patients, i.e., the graphs are slightly mismatched. The GCN-$\mathbf{A}_{rand}$ baseline replaces $\mathbf{A}$ with a randomly-sampled symmetric matrix $\mathbf{A}_{rand}$. Finally, the GCN-$\mathbf{A}_{top}$ baseline replaces $\mathbf{A}$ with a matrix that reflects the same geometric topology as $\mathbf{A}$. To obtain $\mathbf{A}_{top}$, we first bin the edge weights from the unthresholded version of $\mathbf{A}$ and then randomly shuffle edges within each bin [252]. We then threshold to obtain $\mathbf{A}_{top}$. Similar to the proposed model, we kept $\mathbf{A}_{rand}$ and $\mathbf{A}_{top}$ fixed across patients within each CV fold. For robustness, we run a repeated 7-fold CV procedure for all methods, and we sample the matrix $\mathbf{A}_{rand}$ 50 times and report the average statistics.

6.2.2.2 EZ Detection Performance

We evaluate the performance of each method using a repeated 7-fold CV setup, where each fold contains 2 subjects and each repeated sampling (i.e., run) ensures different fold membership. Fig. 6.2 shows the evaluation workflow of our experiments. We report the mean and standard deviation performance across 90 unique runs for the following metrics: sensitivity (TPR), specificity

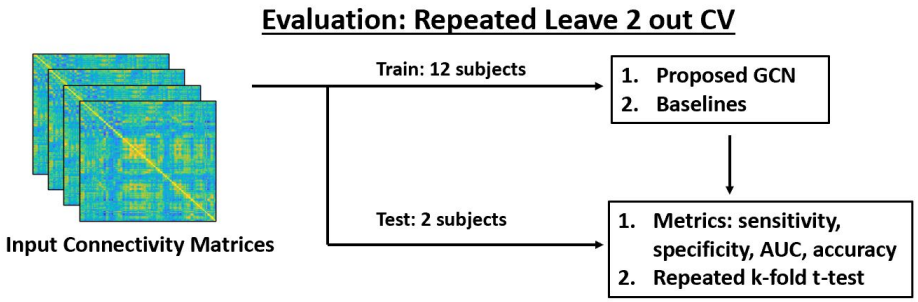

Figure 6.2: We use repeated 7-fold CV for model training and testing. We report the mean and standard deviation of the sensitivity, specificity, AUC, and accuracy across runs. For each baseline, we report the FDR corrected p-value that DeepEZ achieves significantly higher AUC.

(TNR), accuracy, and area under the receiver operating characteristic curve (AUC). To demonstrate a statistically significant performance gain, we perform a t-test on the AUC metric comparing each baseline with DeepEZ. The test statistic corrects for dependencies between the resampled folds, as outlined in [151]. Using this statistic, we compute a p-value and apply FDR correction. Since the ICA2 method is based on deterministic rules and not learned by a classifier, the results are the same across CV folds. Thus, we report only a single average across patients for each metric, as opposed to a mean $\pm$ standard deviation. We use a one-sample t-test to determine statistical significance for ICA2.

Table 6.2 summarizes the EZ detection performance for each method. We observe that DeepEZ achieves the highest sensitivity, precision, F1 and AUC. While the specificity is slightly lower than the GCN-$\mathbf{A}_X$ baselines that swap out the HCP matrix $\mathbf{A}$, the performances are within a standard deviation. Minor variations in accuracy can also be attributed to the class imbalance between EZ and non-EZ regions. The performance gains of DeepEZ are

Table 6.2: Mean plus or minus standard deviation for sensitivity, specificity, precision, F1-score, accuracy, and AUC. The t-score compares the AUC of DeepEZ with each baseline; we also note the corresponding FDR corrected p-value. We use a two-sample t-test based on the repeated 7-fold CV for all models, except ICA2; here, we use a one-sample t-test.

Method	Sensitivity	Specificity	Precision	F1	Accuracy	AUC	t-score	p-value
ICA1	0.071 ± 0.034	0.78 ± 0.045	0.09 ± 0.035	0.08 ± 0.029	0.57 ± 0.033	0.52 ± 0.023	22.26	$< 10^{-10}$
ICA2	0.25	0.7	0.31	0.28	0.69	0.6	7.13	$< 10^{-10}$
BN-CNN	0.099 ± 0.048	0.72 ± 0.036	0.11 ± 0.046	0.11 ± 0.027	0.78 ± 0.015	0.56 ± 0.035	11.84	$< 10^{-10}$
GCN	0.17 ± 0.051	0.78 ± 0.032	0.26 ± 0.039	0.20 ± 0.038	0.81 ± 0.032	0.62 ± 0.035	7.66	$< 10^{-10}$
GCN-SCT	0.22 ± 0.059	0.81 ± 0.036	0.29 ± 0.042	0.25 ± 0.041	0.83 ± 0.029	0.65 ± 0.038	5.13	8.6×10^{-7}
GCN-SSDB	0.31 ± 0.056	0.83 ± 0.031	0.43 ± 0.048	0.36 ± 0.039	0.85 ± 0.03	0.70 ± 0.034	2.15	0.023
GCN-I	0.27 ± 0.061	0.86 ± 0.034	0.39 ± 0.037	0.31 ± 0.028	0.87 ± 0.041	0.68 ± 0.033	3.69	3.3×10^{-4}
GCN-$\mathbf{A}_{subj}$	0.28 ± 0.046	0.87 ± 0.029	0.37 ± 0.039	0.32 ± 0.034	0.89 ± 0.031	0.70 ± 0.026	2.03	0.031
GCN-$\mathbf{A}_{rand}$	0.33 ± 0.041	0.86 ± 0.031	0.41 ± 0.042	0.37 ± 0.037	0.88 ± 0.034	0.71 ± 0.028	1.91	0.046
GCN-$\mathbf{A}_{top}$	0.35 ± 0.039	0.86 ± 0.035	0.47 ± 0.038	0.4 ± 0.035	0.88 ± 0.036	0.72 ± 0.019	1.63	0.078
DeepEZ	$\mathbf{0.4 \pm 0.044}$	0.85 ± 0.033	$\mathbf{0.52 \pm 0.039}$	$\mathbf{0.45 \pm 0.041}$	0.88 ± 0.034	$\mathbf{0.73 \pm 0.031}$		

underscored by the AUC t-test, where we observe a statistically significant ($p < 0.05$) improvement for DeepEZ over all baseline methods except for GCN-$\mathbf{A}_{top}$ method. This result highlights that the performance gain of DeepEZ can be largely attributed to the graph topology (e.g., small-world connectivity) when fusing the rs-fMRI connectivity information across layers.

We note that the competing ICA1 and BN-CNN methods are not well-suited to the task, possibly due to the heterogeneity of our clinical cohort. While ICA2 performs much better than ICA1, it cannot match the performance of DeepEZ. One issue with the rule-based ICA methods is that the selection criteria on one dataset may not generalize well to another. While an end-to-end model such as DeepEZ can easily be retrained on new data, modifying a rule-based approach is nontrivial. The ablated models perform slightly better than ICA1, ICA2 and BN-CNN, but still not on par with DeepEZ. In fact, we observe a notable performance gain using the SSDB, which suggests that a subject-specific approach may be useful to overcome heterogeneity in clinical prediction tasks. There is a similar performance gain when incorporating the

contralateral loss term (SCT), which emphasizes the asymmetry associated with our problem. We observe a marked decline in sensitivity when replacing the d-MRI connectivity matrix $\mathbf{A}$ with identity. This suggests that using information about anatomical pathways is crucial for EZ localization. Interestingly, we also note a performance decline when using the patient-specific d-MRI information encoded in $\mathbf{A}_{subj}$. We hypothesize that this is due to the inconsistency of the edges across patients, particularly in our small dataset. This hypothesis is supported by the results for GCN-$\mathbf{A}_{rand}$ and GCN-$\mathbf{A}_{top}$, which performs slightly better than GCN-$\mathbf{A}_{subj}$. Recall that, while random, the graphs in GCN-$\mathbf{A}_{rand}$ and GCN-$\mathbf{A}_{top}$ are fixed. We find that model training is more stable when the same $\mathbf{A}$ matrix is used for each patient. Finally, GCN-$\mathbf{A}_{top}$ performs the best out of the baselines, implying a benefit to using topological information.

Fig. 6.3 illustrates axial views of the ground truth (red) and predicted (blue) labels for all 14 patients across each method considered. Each row represents a patient (numbered 1-14 in Table 6.2) and each column represents one method. As shown, DeepEZ localizes correct regions in most patients while omitting non-EZ regions. An example of this can be see for Patient 1 (first row), where DeepEZ aligns well with the ground truth labels, while avoiding the spurious predictions by the GCN-X methods. Furthermore, DeepEZ is the only method to localize the resections in Patient 3 and Patient 10, while not incurring incorrect contralateral predictions. We observe that the ICA and BNCNN methods are poorly suited for the task and rarely produce correct predictions. Overall, we observe similar predictions made by the DeepEZ and

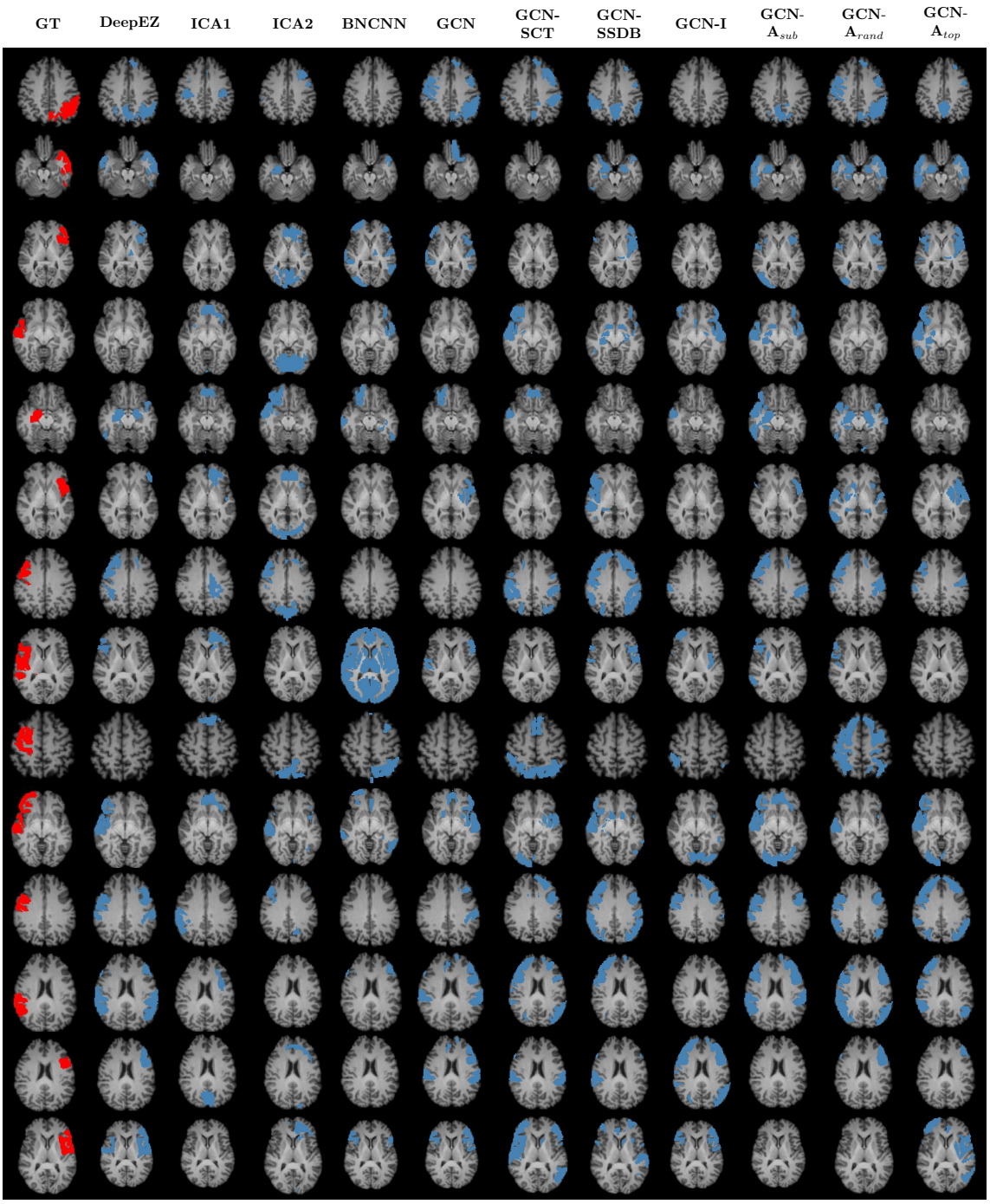

Figure 6.3: Axial view of ground truth (red) and model predictions (blue) for all patients in the UW Madison dataset. Each row corresponds to a single patient, organized from 1–14 according to Table 6.2. Model names are displayed at the top of each column.

GCN-$\mathbf{A}_{top}$ methods, likely due to the preserved network topology. Overall, we observe that DeepEZ achieves a good balance between correct localization while avoiding false positives.

An interesting result of this work is that using subject-specific anatomical connectivity information does not improve localization performance. In fact, the sensitivity, precision, F1 and AUC are worse than when the graph $\mathbf{A}$ is fixed according to the normative HCP dataset. There are three different facets to this result. From an imaging standpoint, there is more variability in our UW Madison dataset due to scanner differences (e.g., 1.5T versus 3T), patient age (9–18 years), and heterogeneous pathologies. This variability may lead to "incorrect" tractography outputs, as compared to the underlying neurophysiology. In contrast, the HCP acquisition sequences have been carefully validated on a standardized cohort, and the preprocessing pipelines have been optimized for the data. Consequently, the matrices $\mathbf{A}$ may reflect long-range and distributed anatomical pathways more accurately than $\mathbf{A}_{subj}$. From an optimization standpoint, the edges are inconsistent across the patient-specific matrices $\mathbf{A}_{subj}$. This inconsistency can lead to instability during training, particularly given the small sample size ($S = 14$). Future work will include comparing the influence of $\mathbf{A}$ and $\mathbf{A}_{subj}$ as the dataset grows in size. Finally, from a neurobiological standpoint, our robustness study in Section III-B suggests that two GCN layers is optimal for EZ localization. Thus, DeepEZ only fuses information across two-stage pathways. Given that we operate at the region level, the HCP template may be sufficient for this operation without needing patient-specific connectivity.

In line with the above observation, our experiments demonstrate that replacing the anatomical connectivity matrix with $\mathbf{A}_{top}$ achieves similar performance. This result underscores the importance of network topology over individual anatomical connections. It also suggests that DeepEZ is robust to variations in the anatomical connectivity matrix used for $\mathbf{A}$. Thus, we conclude that acquiring rs-fMRI data alone is sufficient for EZ localization, which reduces the logistical burden of integrating DeepEZ into the clinical workflow.

We note that there is considerable prior work that uses ICA for EZ localization in rs-fMRI [31, 253, 254]. The meta-analysis of [251] determined that among these, the machine learning approach of [31] (baseline 1 in this work) achieves the best odds ratio. However, as reported in Table 6.2, this method fails to localize the EZ for our cohort. One possible reason is that the handcrafted features chosen in [31] may not generalize well to different cohorts. Another drawback of this method is that it selects just one independent component as the EZ, when there might exist multiple epileptic sources across overlapping components [255]. In fact, we observed in our experiments that the independent components did not overlap well with the surgical resection boundaries used as the pseudo ground truth EZ. We also note that since this method performs ICA at the subject level, there is substantial variability in the component locations across patients. In contrast, DeepEZ uses a well-defined functional parcellation, which allows for both a fine resolution analysis and group-level concordance in the region definitions. We also demonstrate that DeepEZ is robust to the choice of parcellation, which gives the user more

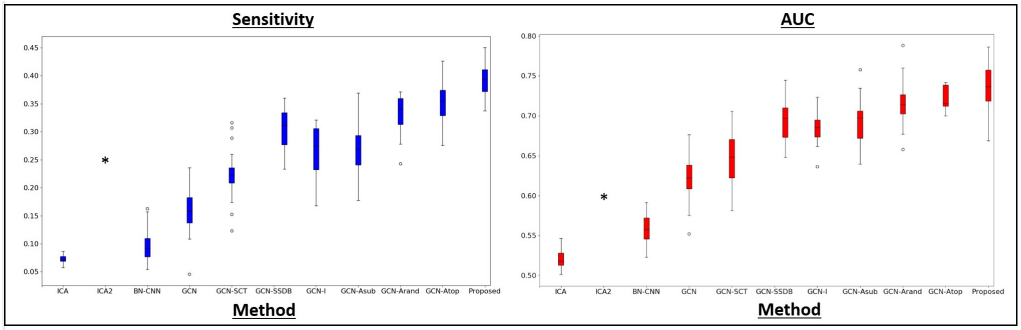

Figure 6.4: Boxplots for sensitivity (left) and AUC (right) across each method. The proposed DeepEZ has the best performance across methods considered.

flexibility when applying the framework to clinical data.

Fig. 6.4 shows boxplots of the sensitivity (left) and AUC (right) among all methods. We include this figure to show the spread of performance based on fold membership. As described in the main text, we observe statistically higher sensitivity and AUC for the proposed DeepEZ method. Fig. 6.5 shows three views of ground truth (red) and DeepEZ predictions (blue) of four representative patients. Based on clinical evaluation, the EZ for these patients are categorized as left temporal, left extra-temporal, right temporal and right extra-temporal, as organized from top to bottom in the figure. We note that DeepEZ can accurately localize the EZ in the second and fourth patients with minimal false positives. For the firth and third patients, DeepEZ exhibits high sensitivity with spurious contralateral predictions.

6.2.2.3 Feature analysis

Finally, we visualize the most commonly learned graph convolutional filters of DeepEZ. Accordingly, we extracted the first-stage graph convolution weight matrices $\mathbf{W}_1 \in \mathbb{R}^{246\times 120}$ learned during each of our repeated 7-fold CV runs.

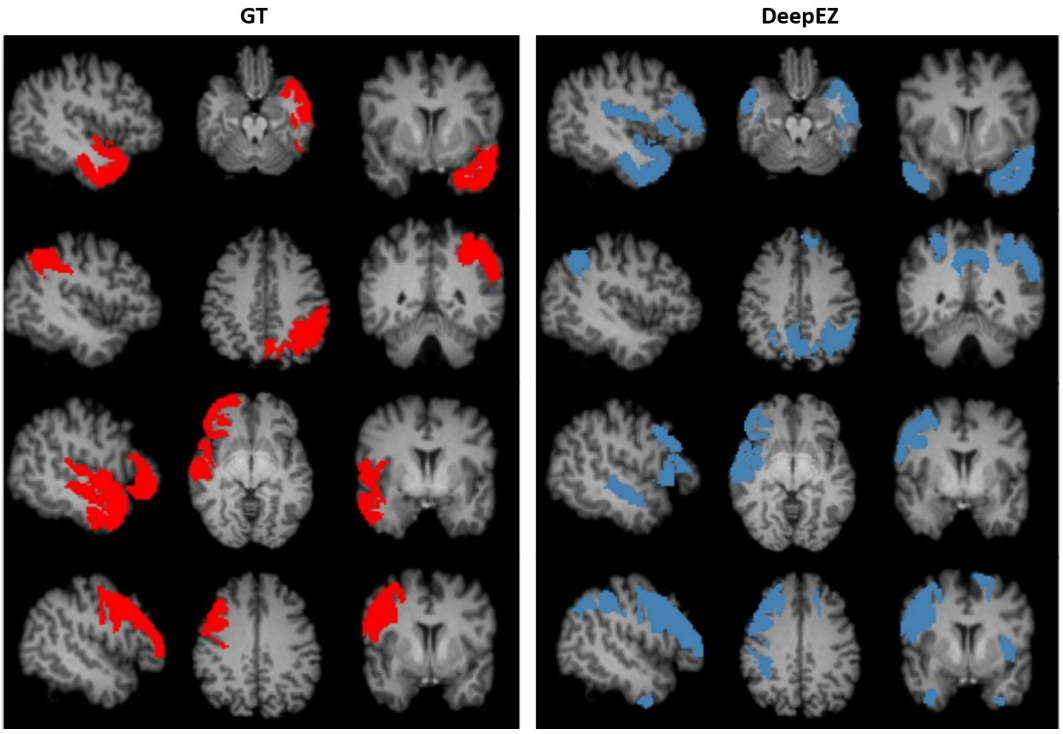

Figure 6.5: Ground truth (red) and DeepEZ predictions (blue) for four representative patients. The patients have the EZ located in left temporal, left extra-temporal, right temporal and right extra-temporal from top to bottom respectively.

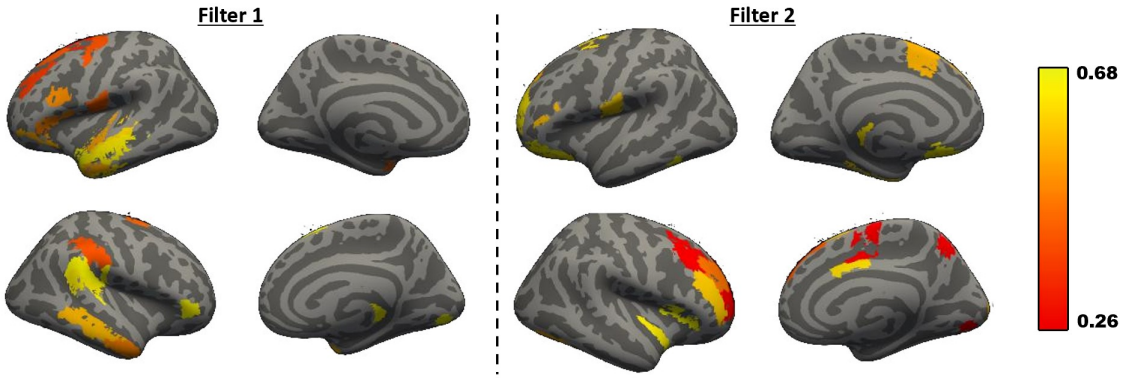

Figure 6.6: Regions implicated by the two most commonly learned convolutional filters in DeepEZ. One filter (**L**) identifies the temporal lobe while the other (**R**) identifies the frontal regions. The distribution of these regions mimic the EZ labels in our dataset.

Each column of $\mathbf{W}_1$ represents a different convolutional filter that can be visualized by plotting the magnitude of the weights back on to the brain. We first aligned the columns of $\mathbf{W}_1$ across the repeated CV folds using a Procrustes algorithm and then masked each column to identify the top ten regions implicated by that filter. We identified two activation patterns that were consistently learned by DeepEZ. Fig. 6.6 shows these filters projected onto the cortical surface, where the color denotes the average activation across repeated CV folds.

We note that one filter (left) implicates regions within the temporal lobe while the other filter (right) implicates the regions associated with the frontal lobe. These patterns mimic the distribution of EZ labels in our UW Madison dataset. Thus, in a data-driven manner, DeepEZ focuses its analysis on stereotypical patterns associated with the EZ.

6.2.2.4 Assessing Model Robustness

In this experiment section, we assess the robustness of DeepEZ to four aspects of our experimental setup: (1) the choice of brain parcellation, (2) the number of GCN layers, (3) the relative weighting δ_2 for the EZ class in Eq. 6.6, and (4) the small dataset size used for model training.

It has been shown that the choice of parcellation can have a tremendous impact on rs-fMRI analyses [184, 185]. For example, coarse parcellations mitigate the effects of noise but can blur subtle effects, whereas fine parcellations preserve detailed phenomena but can be overwhelmed by environmental confounds. In addition, the Brainnetome atlas (BNA) used in Table **??** is symmetric where each region has a direct contralateral region, which is not the case with all parcellations. To explore this, we apply DeepEZ using three different scales of the Craddocks functional parcellation [59]. The Craddocks atlas was derived using a spectral clustering algorithm on the rs-fMRI data from healthy subjects. The different scales come from varying the number of clusters. In this work, we use the $N = 178$, $N = 318$, and $N = 384$ scales, which include both coarser and finer parcellations than the BNA $N = 246$ atlas to assess the effect that resolution has on performance. Once again, we use repeated 7-fold CV to quantify performance.

Table 6.3 reports the accuracy and AUC when applying DeepEZ to each of the parcellations defined above. The p-values are computed with respect to the original BNA atlas. To account for the fact that regions in the Craddocks atlases are not symmetrically defined across the hemispheres, our SCT loss function considers the region with centroid closest to the contralateral location

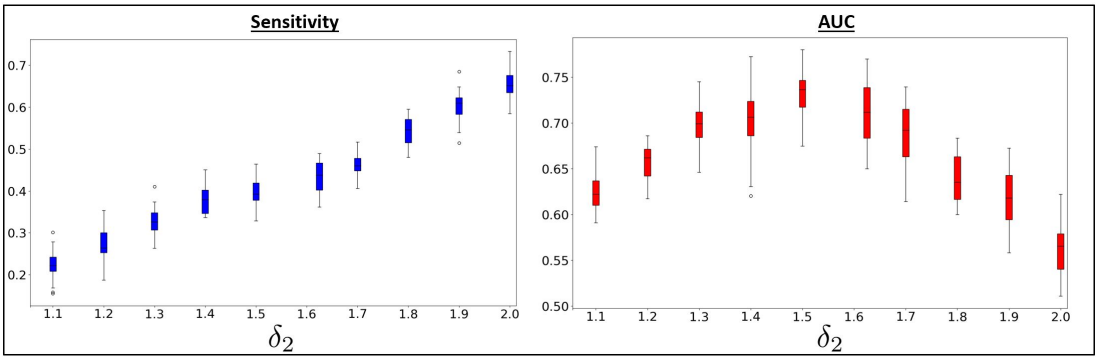

Figure 6.7: Sensitivity (left) and AUC (right) for proposed method while sweeping δ_2 in increments of 0.1. We observe as δ_2 increases, sensitivity increases, but AUC eventually decreases.

Table 6.3: Mean plus or minus standard deviation for accuracy and AUC for different parcellations. The final column shows the FDR corrected p-values when comparing the AUCs of the BNA and Craddocks (CC) parcellations.

Atlas	Accuracy	AUC	p-val
BNA-246	0.88 ± 0.034	0.73 ± 0.031	
CC-178	0.85 ± 0.037	0.70 ± 0.032	0.018
CC-318	0.87 ± 0.038	0.72 ± 0.039	0.367
CC-384	0.87 ± 0.035	0.72 ± 0.031	0.312

as the counterpart $c(n)$ in Eq. 6.6. Based on a $p < 0.05$ threshold, we only observe a significant performance difference in AUC with the CC-178 atlas. This observation suggests that a finer parcellation is better suited for EZ localization. In contrast, DeepEZ is robust using either the CC-318, CC-384, or BNA atlas, which suggests model stability across different parcellations.

At a high level, the GCN layers of DeepEZ perform a random walk on the brain graph defined by the anatomical connections in d-MRI. Our choice of two GCN layers in DeepEZ can analyze the rs-fMRI connectivity patterns associated with path lengths ≤ 2 but cannot capture higher-order information.

Table 6.4: Localization performance as the number of GCN layers is varied. The t-score compares the AUC of the proposed DeepEZ with each baseline; we also note the corresponding FDR corrected p-value.

Layers	Sensitivity	Specificity	Precision	F1	Accuracy	AUC	t-score	p-value
1	0.09 ± 0.035	0.93 ± 0.021	0.12 ± 0.032	0.1 ± 0.029	0.91 ± 0.039	0.55 ± 0.032	12.04	$< 10^{-10}$
2 (Proposed)	$\mathbf{0.4 \pm 0.044}$	0.85 ± 0.033	$\mathbf{0.52 \pm 0.039}$	$\mathbf{0.45 \pm 0.041}$	0.88 ± 0.034	$\mathbf{0.73 \pm 0.031}$		
3	0.27 ± 0.042	0.87 ± 0.035	0.38 ± 0.039	0.32 ± 0.037	0.88 ± 0.031	0.70 ± 0.027	2.01	0.039
4	0.33 ± 0.039	0.84 ± 0.029	0.45 ± 0.041	0.38 ± 0.029	0.83 ± 0.028	0.68 ± 0.029	3.6	4.9×10^{-4}

To probe this effect, we conduct a robustness experiment in which we vary the number of GCN layers in DeepEZ and use the repeated 7-fold CV strategy in Fig. 6.2 to quantify the performance of each method.

Table 6.4 reports the performance across 1–4 GCN layers. We observe that the proposed architecture (2 GCN layers) achieves the best trade-off between true positive and false positive detections, as quantified via a t-test on the AUC. There are two interpretations for this result. First, it appears that the rs-fMRI connectivity patterns associated with the EZ are the most prominent at a walk of length of two, with diminishing returns beyond this point. Second, increasing the number of GCN layers also increases the number of model parameters, which may lead to overfitting. Taken together, we believe that two GCN layers balances the trade-off between capturing discriminative patterns without overfitting on small datasets.

One of the key aspects of DeepEZ is the weighted cross-entropy loss to handle class imbalance. To probe this effect, we sweep the EZ detection hyperparameter δ_2 in in increments of 0.1 while keeping the other hyperparameters fixed. Fig. 6.7 shows the sensitivity (left) and AUC (right) metrics as δ_2 varies over the range $[1.1, 2.0]$. As expected, sensitivity increases with δ_2 due to the higher penalty for incorrectly classifying EZ regions as baseline. However, the

Table 6.5: Mean plus or minus standard deviation for sensitivity, specificity, AUC, and accuracy with and without data augmentation. The final column shows the FDR corrected p-values for the associated t-score for comparing AUC.

Augmentation	Sensitivity	Specificity	Precision	F1	Accuracy	AUC	t-score	p-value
Without	0.4 ± 0.044	0.85 ± 0.033	0.52 ± 0.039	0.45 ± 0.041	0.88 ± 0.034	0.73 ± 0.031		
With	0.43 ± 0.033	0.87 ± 0.026			0.89 ± 0.015	0.74 ± 0.019	-1.28	0.76

AUC metric peaks at 1.5 and steadily decreases, which suggests that DeepEZ incurs too many false positives at larger values of δ_2. We observe relatively stable performance in both metrics over the range $\delta_2 = [1.4 - 1.6]$. Finally, we note that the weighted cross-entropy loss is useful from a clinical perspective, as it is more important not to miss the EZ regions at this stage of therapeutic planning for epilepsy.

Data augmentation has been shown to improve the performance of deep learning models due to providing more information about the underlying data distribution [186, 187]. Given the small sample size ($N = 14$), we explore whether DeepEZ would benefit from data augmentation. Here, we sub-sampled the time series data using a continuous sliding window to create 10 distinct new training similarity matrices for each subject. Our augmented dataset contains an order of magnitude more data points ($N = 140$). Our evaluation strategy remained the same as depicted in Fig. 6.2. Table 6.5 reports the EZ detection performance for DeepEZ with and without data augmentation. We observe a small performance boost in each metric and smaller standard deviations when using data augmentation during training. However, we note that the performance gain in AUC is not statistically significant, as indicated in the last column of Table 6.5. This result demonstrates that DeepEZ is able to effectively mine the information present in our original dataset for

generalizable EZ localization.

We have presented DeepEZ, a graph convolutional network on static connectivity data to localize the EZ in focal epilepsy patients. We have shown the robustness and overall performance of DeepEZ. Similar to our eloquent cortex localization work, we extend the DeepEZ model to operate on dynamic connectivity to improve performance. The next section of this thesis will summarize the findings presented in [41], where we use a transformer-based attention mechanism applied to dynamic functional connectivity alongside data augmentation techniques to improve EZ localization metrics.

6.3 A Deep Learning Framework To Localize the Epileptogenic Zone From Dynamic Functional Connectivity Using A Combined Graph Convolutional and Transformer Network

6.3.1 Model

In this work, we present the first automated framework that uses dynamic functional connectivity from rs-fMRI to localize the EZ across a heterogeneous epilepsy cohort. Our deep network architecture builds off of DeepEZ and uses an anatomically-regularized graph convolutional network (GCN) for feature extraction. From here, a transformer network learns a temporal attention vector, which identifies relevant time windows of the rs-fMRI scan that aid in localization. Following [142], we train our network entirely on simulated data derived from the Human Connectome Project (HCP), and we test it on a clinical epilepsy dataset from the University of Wisconsin Madison. We

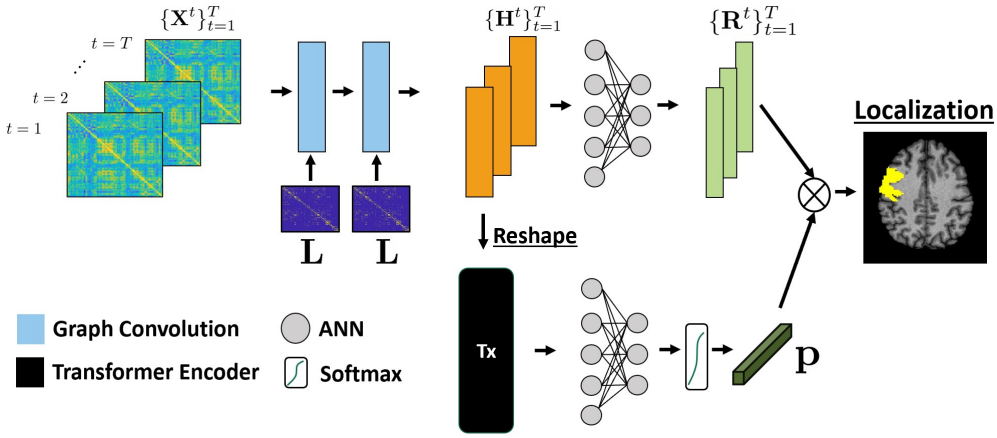

Figure 6.8: Network overview. **Top**: We use a multi-modal GCN and fully-connected layers to obtain node-wise predictions of the EZ over time **R**. **Bottom**: Our transformer and fully-connected layers network extract a temporal attention vector **p** that selects specific windows of the dFC input. The attention **p** is combined with **R** to obtain the final EZ predictions.

demonstrate the significantly improved performance of our framework, as compared to both ablated versions of our model and the method of [40].

Fig. 6.8 shows an overview of our framework. Our method uses a two layer GCN to obtain intermediate node-wise features over time. These intermediate features are fed into both a temporal attention network as well as a node-classifier network for EZ localization. We use the sliding window technique to generate the dFC inputs to our framework. Formally, let N be the number of brain regions in our parcellation, T be the number of sliding windows, and $\{\mathbf{X}^t\}_{t=1}^T$ be the dFC matrices. $\mathbf{X}^t \in \mathbb{R}^{N \times N}$ is computed from a segment $\mathbf{Z}^t \in \mathbb{R}^{N \times d}$ of the rs-fMRI time series, where d is the sliding window size.

6.3.1.1 GCN for feature extraction

The first stage of our model is an anatomically-regularized GCN for feature extraction. We use diffusion MRI (d-MRI) tractography to construct the binary adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ used for graph filtering [94]. In this case, an entry $\mathbf{A}_{ij} = 1$ denotes an anatomical connection between regions i and j. Let $\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ be the normalized graph Laplacian of $\mathbf{A}$, where $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$.

Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ denote the binary adjacency matrix used for graph filtering [94]. We use d-MRI tractography to construct $\mathbf{A}$. In this case, an entry $\mathbf{A}_{ij} = 1$ denotes an anatomical connection between regions i and j. Let $\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ be the normalized graph Laplacian of $\mathbf{A}$, where $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$. Each layer produces an activation map $\mathbf{H}_l \in \mathbb{R}^{N \times G_l}$, where $l \in \{1, 2\}$ denotes the layer number. The learnable parameters in each graph convolution are a weight matrix $\mathbf{W}_l \in \mathbb{R}^{G_l \times G_{l+1}}$ and a constant bias $\mathbf{b}_l \in \mathbb{R}^{1 \times G_{l+1}}$. The intermediate activation $\mathbf{H}_1$ is generated via the propagation rule:

$$\mathbf{H}_1^t = \phi\Big(\mathbf{L}\mathbf{X}^t\mathbf{W}_1 + \mathbf{b}_1\Big), \tag{6.7}$$

with the activation $\mathbf{H}_2$ generated likewise from $\mathbf{H}_1$.

6.3.1.2 Transformer-Based Temporal Attention

The outputs $\{\mathbf{H}_2^t\}_{t=1}^T$ of the GCN correspond to intermediate node-level features per time point. From here, the temporal attention module leverages the encoder stage of a transformer network [101], followed by a fully-connected artificial neural network (ANN). Our transformer employs multi-headed self-attention (MHA) and feed-forward networks with residual connections to

process sequential data. Formally, let $\mathbf{H}' \in \mathbb{R}^{T \times NG_2}$ be a flattened version of $\{\mathbf{H}_2^t\}_{t=1}^T$. A single encoder layer in the transformer is computed as follows:

$$\mathbf{C}_1 = \text{MHA}(\mathbf{H}') + \mathbf{H}' \qquad \mathbf{C}_2 = \text{FF}(\mathbf{C}_1) + \mathbf{C}_1, \tag{6.8}$$

where the $\text{FF}(\cdot)$ operation denotes a feed-forward network.

The $\text{MHA}(\cdot)$ function in Eq. 6.8 consists of multiple self-attention (SA) operations, where each SA_i for $i \in \{1 \cdots I\}$ is computed as $\text{SA}_i(\mathbf{V}_i) = \mathbf{M}_i \mathbf{V}_i$. As introduced in [101], the attention mask $\mathbf{M}_i \in \mathbb{R}^{T \times T}$ captures the similarity between a query matrix $\mathbf{Q}_i = \mathbf{W}_i^q \mathbf{H}'$ and a key matrix $\mathbf{K}_i = \mathbf{W}_i^k \mathbf{H}'$, both of which are linear functions of the input data:

$$\mathbf{M}_i = \text{Softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_i^T}{\sqrt{NG_2}}\right) \tag{6.9}$$

Likewise, the value matrix $\mathbf{V}_i = \mathbf{W}_i^v \mathbf{H}'$ is also obtained via a linear layer. The matrices $\mathbf{W}_i^q$, $\mathbf{W}_i^k$, and $\mathbf{W}_i^v$ in the above expressions denote the learned weights for head i.

The self-attention outputs are concatenated across heads and fed through a linear layer to obtain one MHA operation. The attention mask $\mathbf{M}_i$ identifies which time points have similar representations and scales the output accordingly. The transformer combines the $\text{MHA}(\cdot)$ operation with a residual connection. The subsequent $\text{FF}(\cdot)$ operation consists of two fully-connected layers plus another residual connection. The encoding procedure in Eq. 6.8 optimizes the mixing across the sequential input features for the downstream task.

The output of the transformer is fed through two fully-connected layers

and a softmax function to obtain our temporal attention vector $\mathbf{p} \in \mathbb{R}^{T\times 1}$. The attention $\mathbf{p}$ is designed to identify which time points are more relevant for downstream node classification. Both the intermediate features $\{\mathbf{H}\}_{t=1}^{T}$ and attention vector $\mathbf{p}$ appear in the EZ classification stage.

6.3.1.3 Classification and Loss Function

We treat the problem of EZ localization as a region-wise classification problem, where each region is identified as either belonging to the EZ class or to the "normal" class. Formally, the intermediate features $\{\mathbf{H}\}_{t=1}^{T}$ are fed into a two-layer ANN to obtain node-wise predictions over time $\{\mathbf{R}\}_{t=1}^{T}$, where $\mathbf{R}^t \in \mathbb{R}^{N\times 2}$. Our temporal attention vector $\mathbf{p}$ is combined with $\{\mathbf{R}\}_{t=1}^{T}$ via an inner product to obtain a single prediction for each region.

We adopt a modified version of the loss function presented in [40], which uses a weighted cross-entropy loss and a regularization term to suppress activations in regions contralateral to the EZ. Let $\mathbf{Y} \in \mathbb{R}^{N\times 2}$ be the one-hot encoded labels, N_e denote the nodes that belong to the EZ class and let $c(n)$ denote the contralateral counterpart to region n. Our training loss function consists of the following two terms:

$$\begin{aligned}\mathcal{L}(\{\mathbf{X}^t\}_{t=1}^{T}, \mathbf{Y}) = &\underbrace{-\sum_{n=1}^{N}\sum_{i=1}^{2}\delta_i \log\Big(\sigma\Big(\sum_{t=1}^{T}\mathbf{R}_{n,c}^{t}\cdot\mathbf{p}^t\Big)\Big)\mathbf{Y}_{n,c}}_{\text{Weighted Cross Entropy}} \\ &\underbrace{-\lambda\frac{1}{N_e}\sum_{n\in N_e}\Big(\sigma\Big(\sum_{t=1}^{T}\mathbf{R}_{n,2}^{t}\cdot\mathbf{p}^t\Big)-\sigma\Big(\sum_{t=1}^{T}\mathbf{R}_{c(n),2}^{t}\cdot\mathbf{p}^t\Big)\Big)}_{\text{EZ Contralateral Term}}.\end{aligned} \tag{6.10}$$

6.3.1.4 Data Augmentation for Training

Rs-fMRI studies of focal epilepsy patients are often limited in size. Therefore, following [142], we train our deep network entirely on augmented data derived from a neurotypical control dataset. For each training sample, we augment the healthy rs-fMRI data by first randomly selecting a spatially continuous neighborhood of voxels to form the EZ and then modifying the time series at those voxels via one of six noise models: (1) adding normally distributed noise, (2) adding uniformly distributed noise, (3) adding power-law noise, (4) adding Brownian noise, (5) adding noise generated by a Levy walk process, and (6) randomly permuting the time series. Since there is no established ground truth for how the EZ affects rs-fMRI, the combination of these six noise models exposes our network to a broad range of data abnormalities during training [142]. To our knowledge, our work is the first to use data augmentation for EZ localization based on rs-fMRI connectivity.

We implement our network in Pytorch [182] using the ADAM optimizer and a leaky-ReLU activation function with slope = -0.1 between each layer. To prevent data leakage, all hyperparameters of our network are set using cross-validation on 100 EZ-augmented subjects from the Human Connectome Project (HCP) dataset.

6.3.2 Experimental results

6.3.2.1 Datasets and baseline algorithms

Our training data consists of 300 HCP subjects [137]. We generate training three samples per subject ($S = 900$ total) by varying the EZ location and/or

noise model used for data augmentation. We use the Brainnetomme atlas [60] to define $N = 246$ cortical and subcortical regions for our analysis. We construct the adjacency matrix $\mathbf{A}$ used in our GCN from d-MRI tractography of 50 additional HCP subjects. Individual structural connectivity matrices are generated according to [143]. We average and threshold these matrices to compute $\mathbf{A}$, used for both training and testing.

Our clinical dataset consists of 14 pediatric patients with focal epilepsy from the University of Wisconsin (UW) Madison. All preprocessing details can be found in Chapter 2 of this thesis. As shown in Fig. 6.9, we manually segment the resection cavity and consider this area as the pseudo ground truth EZ for each patient.

We compare our proposed framework against competing methods from the literature (first two below) and ablated versions of our framework (last four below):

- **BN-CNN:** A modified version of the BrainNetCNN architecture developed in [169] that performs region-wise, rather than subject-level, prediction.
- **DeepEZ:** The model developed by [40] to localize the EZ based on static rs-fMRI connectivity.
- **NoAttn:** Ablation #1 that removes the temporal attention mechanism. Final predictions are averaged over time.
- **ANNattn:** Ablation #2 that uses a fully-connected ANN rather than a transformer as the temporal attention model.

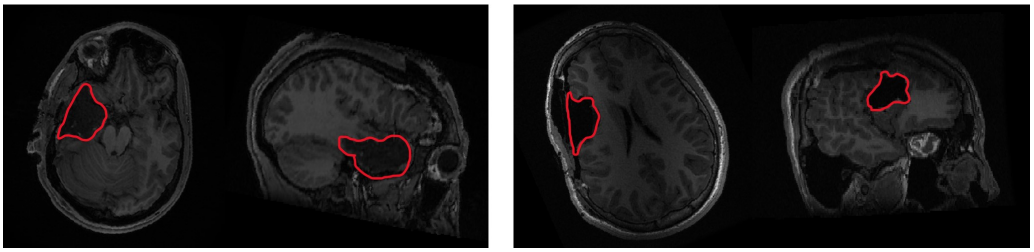

Figure 6.9: Resection boundaries (red) for two epilepsy patients.

- **LSTM:** Ablation #3: that uses an LSTM rather than a transformer as the temporal attention model.
- **NoAugment:** Ablation #4 that trains the deep network directly on the clinical data with no augmentation.

6.3.2.2 Localization performance

Table 6.6 reports the performance of each model. We use a De Long's test on the AUC metric [256] to determine statistically significant improvement between our proposed framework and each baseline. We note an improvement in sensitivity and AUC when using the transformer to extract the temporal attention weights. Likewise, we note an improvement when using data augmentation for training. This is likely because our clinical dataset is too small to extract information from using our dynamic model, due to the larger parameterization. Finally, Fig. 6.10 shows model the ground truth (red) and predicted (yellow) labels for three Epilepsy patients among all models. The observed trends reflect the metrics in Table 6.6.

Method	Sens	Spec	Acc	AUC	p-value
BN-CNN	0.17	0.81	0.69	0.58	$< 10^{-8}$
DeepEZ	0.34	0.88	0.87	0.7	0.011
NoAttn	0.35	0.86	0.86	0.69	< 0.01
ANNattn	0.41	0.86	0.88	0.71	0.019
LSTM	0.41	0.88	0.88	0.73	0.052
NoAugment	0.28	0.90	0.90	0.68	< 0.01
Proposed	**0.51**	0.89	**0.92**	**0.77**	

Table 6.6: Performance metrics for EZ classification.

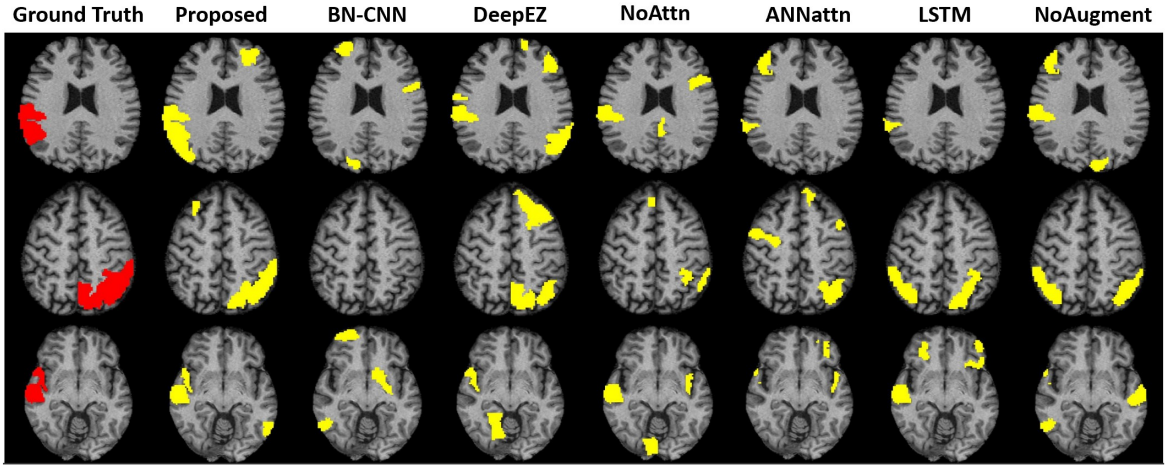

Figure 6.10: Ground truth (red) and model predictions (yellow) for three test subjects.

6.3.2.3 Temporal Attention

Fig. 6.11 shows the temporal attention weights recovered from each method that uses attention (proposed, ANNattn, and LSTM) for each of the 14 epilepsy patients during the testing phase. We observe a larger dynamic range in the weights recovered from the proposed framework, as compared to the ablated models. We conjecture that the transformer learns more nuances in the dFC data that improve the region-wise EZ classification. We hypothesize that the

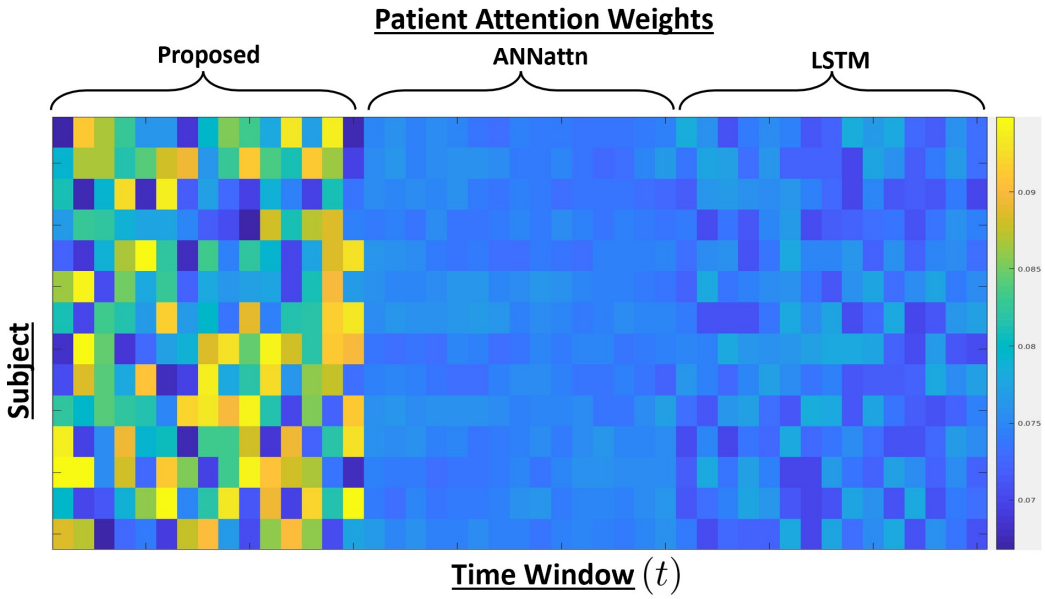

Figure 6.11: Temporal attention weights recovered for **Left:** the proposed framework, **Middle:** the ANNattn ablation model, and **Right:** the LSTM ablation model for all epilepsy patients.

MHA operation, which inherently captures similarities and differences between time-points, is responsible for better honing in on the relevant intervals for prediction.

6.4 Conclusion

To conclude this chapter, we have introduced DeepEZ, a novel deep learning approach for EZ localization based on rs-fMRI connectivity. DeepEZ relies on spatial graph convolutions that leverage biologically-inspired anatomical pathways to aggregate neighborhood information during forward propagation. These graph convolutions are complemented with a subject-specific detection bias (SSDB) to mitigate inter-patient differences in connectivity and an asymmetry loss term to encourage lateralized predictions. In comparison

to baseline methods, DeepEZ achieves statistically improved AUC for detecting EZ regions. Via ablation studies, we show that these performance gains are linked to the EZ contralateral term (SCT) and the SSDB. In subsequent anaylses, we demonstrate that DeepEZ is robust to varying the parcellation used for analysis and performs comparably with and without data augmentation. We also show that DeepEZ achieves robust performance within a range of δ_2 in the weighted cross-entropy loss term.

Clinical rs-fMRI studies often lack statistical power due to small sample sizes [257] with logistical constraints making it difficult to acquire additional data for analysis. In our case, the UW Madison dataset contains a specialized cohort of pediatric focal epilepsy patients who underwent surgical resection of the EZ. Currently, rs-fMRI is not a commonly acquired modality for epilepsy patients, which limits our ability to grow the dataset further. Thus, to maximize the sample size for model training and evaluation, we include patients that have their scans taken from two different scanners (1.5T and 3T) which is common in the literature [258, 259, 260]. Accordingly, we have designed DeepEZ to mine information from smaller heterogeneous datasets. Relevant attributes include a relatively small number of learnable parameters, a biologically informed spatial graph, the SSDB module to improve sensitivity, and a SCT to encourage clinically relevant EZ localization patterns. We demonstrate state-of-the-art performance, along with a robustness to different modeling choices and data augmentation. The fact that DeepEZ harmonizes information across scanners is particularly encouraging, as inter-scanner differences are known to confound deep learning models [261].

We then propose a novel end-to-end model based on dynamic functional connectivity to localize the EZ in focal epilepsy patients. Our model leverages a combined GCN + transformer architecture for feature extraction and temporal tracking. In parallel, we leverage a simple yet effective data augmentation strategy for robust training. We show statistically significant improvements over the baseline methods, and hypothesize that the performance gain is directly related to using the transformer-based attention module, which hones in on relevant intervals of the dFC time series for EZ prediction. Our work shows increased promise in using rs-fMRI as a preoperative protocol for noninvasive EZ localization.

Chapter 7

A framework to characterize noisy labels in EZ localization

7.1 Introduction

With increasing availability of heavy compute power and large-scale datasets, deep learning has made unprecedented breakthroughs in many common machine learning tasks such as computer vision [88], language processing [262], and medical imaging [32]. As dataset sizes increase, it has been increasingly popular to employ non-expert humans or automated systems with little supervision to automatically label datasets [263]. However, datasets collected using these methods usually suffer from very high label noise. The ratio of corrupted to clean labels in real world datasets is reported to be anywhere from 8.0% to 38.5% [264].

The challenge of curating datasets with accurate labels is especially significant in medical imaging. Datasets tend to be small to begin with, and institutional policies or patient privacy can prevent data from being shared. Labeling of medical images is especially resource-intensive and potentially

unreliable, as it requires specific domain expertise and there often exists a large degree of inter-observer variability [32]. The study in [265] has utilized a large number of experts to annotate medical image datasets, but these efforts require financial and logistical resources. The study in [266] used crowd sourcing to obtain non-expert annotations for medical imaging, but this can be reliable due to lack of proper training of the participants. Overall, lack of large datasets with trustworthy labels is considered as one of the biggest challenges associated with adoption of deep learning methods in the hospital and medical applications [267].

The models presented in this thesis thus far have assumed perfect labels, which is not usually the case in the medical imaging domains. Regarding eloquent cortex localization, we assumed that the thresholded GLM activation maps taken from the language and motor t-fMRI acted as a perfect ground truth to capture the regions of interest. However, this assumption could be flawed, as there is a lot of evidence showing reliability issues with using t-fMRI as ground-truth biomarkers [268]. Factors like inability to follow the protocol, or excessive head motion can further disrupt the reliability of t-fMRI activations [25, 160]. Furthermore for our application in tumor removal procedures, the reliability of t-fMRI activations can be even lower than with compared against a healthy cohort [269].

Regarding EZ localization, our models assumed that the resection from the post-operative T1 image acts as the ground truth EZ. That is, our models assumed that there is abnormal fMRI activity in the entire resection and

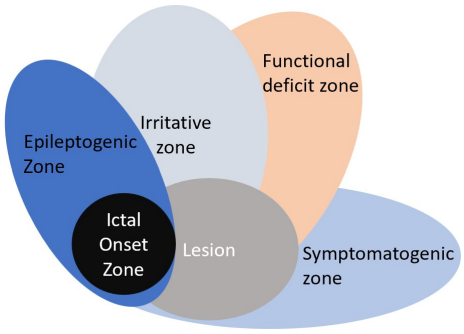

Figure 7.1: A cartoon example of the various areas of the cortex that are responsible for seizures. Figure is recreated and taken from [271].

the rest of the brain is considered healthy, or typically functioning. However, it is well known that these resections usually are larger than the EZ, as a means of removing secondary tissue that can also be problematic [270]. Specifically, there are five cortical zones defined in presurgical evaluation. The irritave zone is the area of the cortex that generates the interictal spikes, the seizure-onset zone is the area of the cortex that initiates clinical seizures. The symptomatogenic zone is the the area that, when activated, produces the initial ictal symptoms or signs, the epileptogenic lesion is causitive of epileptic seizures because the lesion itself is epileptogenic, and the functional deficit zone, which is the area of the cortex that is not functioning normally in the interictal period [271]. Fig. 7.1 shows a cartoon image recreated and taken from [271] that illustrates the various overlaps that these five regions can have with eachother. As described in [271], removing just the ictal onset zone is not enough to sustain lasting seizure freedom, and that there is no direct preoperative measurement of the EZ; its delineation is a conceptual exercise derived from many tests and presurgical evaluations.

7.1.1 Contributions

In this section of the thesis, we develop a framework to identify noisy labels within our existing experimental setups for EZ localization from connectivity graphs. Specifically, we model the probability of an incorrect label using the concrete distribution [272], which is essentially a continuous relaxation of the Bernoulli distribution. Leveraging the data augmentation techniques presented in [142] for EZ simulation, we create a simulated dataset that contains noisy labels that reflect the expected pattern of noisy labels in the context of EZ localization. We develop a neural-network architecture to learn the concrete distribution parameters of interest in a strongly pre-trained fashion and then train our localization network with just the noisy observed labels. We assess how well the proposed method with label dropout does compared to the standard localization only method presented in [41], and finally we show that our proposed method outperforms our previous method ([41]) on the real subjects from the UW dataset when trained with the noisy dataset and show the qualitative predictions and recovered parameters on all 14 examples. It is important to note that this section of the thesis is ongoing work and has not resulted in a publication yet. We plan on submitting the finalized model and results to the International Symposium of Biomedical Imaging (ISBI) 2024 conference.

7.2 Learnable label dropout for noisy label detection

Our goal in this chapter is to develop a mathematical framework coupled with a set of experiments to identify noisy labels in the context of EZ localization from rs-fMRI dynamic functional connectivity graphs. We introduce the concrete distribution to characterize label noise and use a deep learning network to learn the concrete distribution parameters in a pre-trained fashion. We take a semi-supervised approach where we pre-train a portion of our deep network with knowledge of existing latent true labels and then only use the observed potentially corrupted labels during training for localization. We begin with our graphical model and underlying assumptions.

7.2.1 Model

7.2.1.1 Graphical model representation

We will explore the noisy label problem in the context of the previously presented models. Fig. 7.2 shows the graphical model for one subject that describes our model and assumptions. Mathematically, let $n \in \{1, \cdots, N\}$ index node, $\mathbf{X} \in \mathbb{R}^{N \times N \times T}$ be the dynamic functional connectivity input, so $\mathbf{X}_n \in \mathbb{R}^{N \times T}$ is the connectivity profile, or data, associated with node n. Let $\mathbf{Y}_n$ be the observed label for node n and $\tilde{\mathbf{Y}}_n$ be the real, unobserved label for node n and let $\mathbf{Z}_n$ be a random variable that captures the latent corruption of label n, which is parameterized by α_n.

A shown by the plate notation, we assume i.i.d. between nodes and can derive the joint distribution of $\mathbf{Y}_n, \tilde{\mathbf{Y}}_n, \mathbf{Z}_n$ conditioning on $\mathbf{X}_n$ from the

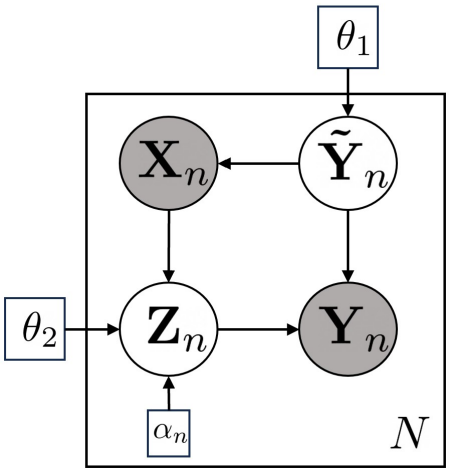

Figure 7.2: A graphical model showing the dependencies in our model. Circles represent random variables, rectangles represent parameters, and arrows represent dependencies. Shaded nodes are observed while white are latent.

graphical model as

$$P(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n, \mathbf{Z}_n | \mathbf{X}_n) = P(\mathbf{Y}_n | \tilde{\mathbf{Y}}_n, \mathbf{Z}_n) P(\tilde{\mathbf{Y}}_n | \mathbf{X}_n) P(\mathbf{Z}_n | \mathbf{X}_n). \tag{7.1}$$

First, we define the helper function $\Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n) = \mathbf{Y}_n \tilde{\mathbf{Y}}_n + (1 - \mathbf{Y}_n)(1 - \tilde{\mathbf{Y}}_n)$, which acts as an indicator to identify when the observed label is corrupted (i.e. mislabeled) or not. Using our i.i.d assumption, we can assume $P(\mathbf{Y} | \tilde{\mathbf{Y}}, \mathbf{Z}) = \prod_{n=1}^{N} P(\mathbf{Y}_n | \tilde{\mathbf{Y}}_n, \mathbf{Z}_n)$. We define our likelihood term as

$$P(\mathbf{Y}_n | \tilde{\mathbf{Y}}_n, \mathbf{Z}_n) = [\rho^{\Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n)} (1 - \rho)^{(1 - \Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n))}]^{(1 - \mathbf{Z}_n)} [\alpha_n^{(1 - \Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n))} (1 - \alpha_n)^{\Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n)}]^{\mathbf{Z}_n} \tag{7.2}$$

where $\rho \approx 0.99$, or essentially is 1. Our likelihood term asserts that when $\mathbf{Z}_n = 1$, there is an α_n probability of the label being corrupted, and when $\mathbf{Z}_n = 0$, that there is a very high chance that the label is not corrupted (i.e.

$\mathbf{Y}_n = \tilde{\mathbf{Y}}_n$). In contrast to existing methods that take a similar approach, like the one in [273], our goal is to develop a framework that can learn the α_n parameters as opposed to have them be set *a priori*.

The term $P(\tilde{\mathbf{Y}}_n|\mathbf{X}_n)$ simply describes the class label distribution while the term $P(\mathbf{Z}_n|\mathbf{X}_n)$ describes the noise content within the data. We will use two separate neural networks to characterize these distributions, parameterized by θ_1 and $\theta_2' = \theta_2 \bigcup \alpha$ respectively.

7.2.1.2 Concrete label dropout

Let α_n denote the probability that label n is incorrect, that is $P(\mathbf{Y}_n \neq \tilde{\mathbf{Y}}_n) = \alpha_n$. Instead of sampling the random variable from the discrete Bernoulli distribution, we model the latent random variable $\mathbf{Z}_n$ that describes label corruption via the concrete distribution [272]

$$\mathbf{Z}_n = \sigma\left(\frac{1}{t}\left(\log\left(\frac{\alpha_n}{1-\alpha_n}\right) + \log\left(\frac{u_n}{1-u_n}\right)\right)\right) \tag{7.3}$$

which gives a continuous relaxation of the Bernoulli distribution, where u_n is a uniform random variable on the interval of $[0, 1]$. Here, the temperature t is a hyperparameter of the distribution. Biologically, we chose the concrete relaxation of the Bernoulli to reflect that the labels in our case could be partially incorrect. For example, partial volume effects are introduced when we apply a parcellation to obtain labels. Empirically, we noticed smoother optimization and training from using the concrete Bernoulli compared to the discrete case.

7.2.1.3 Learning the parameters

We use two deep networks to to model $P(\tilde{\mathbf{Y}}|\mathbf{X}) = \prod_{n=1}^{N} P(\tilde{\mathbf{Y}}_n|\mathbf{X}_n)$ and $P(\mathbf{Z}|\mathbf{X}) = \prod_{n=1}^{N} P(\mathbf{Z}_n|\mathbf{X}_n)$ where θ_1 parameterizes $P(\tilde{\mathbf{Y}}|\mathbf{X})$ and $\theta_2' = \theta_2 \bigcup \alpha$ parameterizes $P(\mathbf{Z}|\mathbf{X})$. Our goal is to find the optimal $\theta = \theta_1 \bigcup \theta_2'$ that maximizes the incomplete log-likelihood $P(\mathbf{Y}|\mathbf{X};\theta) = \prod_{n=1}^{N} P(\mathbf{Y}_n|\mathbf{X}_n;\theta)$. We use the EM algorithm to iteratively solve this problem [274].

For an arbitrary distribution $q(\tilde{\mathbf{Y}}, \mathbf{Z}|\mathbf{Y}, \mathbf{X})$, we can derive a lower bound of the incomplete log-likelihood

$$\log P(\mathbf{Y}|\mathbf{X};\theta) = \log \sum_{\tilde{\mathbf{Y}},\mathbf{Z}} P(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{Z}|\mathbf{X};\theta) \geq q(\tilde{\mathbf{Y}}, \mathbf{Z}|\mathbf{Y}, \mathbf{X}) \log \frac{P(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{Z}|\mathbf{X};\theta)}{q(\tilde{\mathbf{Y}}, \mathbf{Z}|\mathbf{Y}, \mathbf{X})}. \tag{7.4}$$

Using the EM algorithm, the E-step involves computing the posterior of the latent variables using the current parameters $\theta^{(t)}$,

$$P(\tilde{\mathbf{Y}}, \mathbf{Z}|\mathbf{Y}, \mathbf{X};\theta^{(t)}) = \frac{P(\mathbf{Y}|\tilde{\mathbf{Y}}, \mathbf{Z};\theta^{(t)})P(\tilde{\mathbf{Y}}|\mathbf{X};\theta^{(t)})P(\mathbf{Z}|\mathbf{X};\theta^{(t)})}{\sum_{\tilde{\mathbf{Y}}',\mathbf{Z}'} P(\mathbf{Y}|\tilde{\mathbf{Y}}', \mathbf{Z}';\theta^{(t)})P(\tilde{\mathbf{Y}}'|\mathbf{X};\theta^{(t)})P(\mathbf{Z}'|\mathbf{X};\theta^{(t)})} \tag{7.5}$$

and then the expected complete log-likelihood can be written as

$$Q(\theta;\theta^{(t)}) = \sum_{\tilde{\mathbf{Y}},\mathbf{Z}} P(\tilde{\mathbf{Y}}, \mathbf{Z}|\mathbf{Y}, \mathbf{X};\theta^{(t)}) \log P(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{Z}|\mathbf{X};\theta). \tag{7.6}$$

For the M-step, we exploit two deep networks to model the probability $P(\tilde{\mathbf{Y}}|\mathbf{X};\theta_1)$ and $P(\mathbf{Z}|\mathbf{X};\theta_2')$. Recall that $\theta_2' = \theta_2 \bigcup \alpha$. The gradient of Q with respect to θ can be decoupled into two parts, which are implemented via

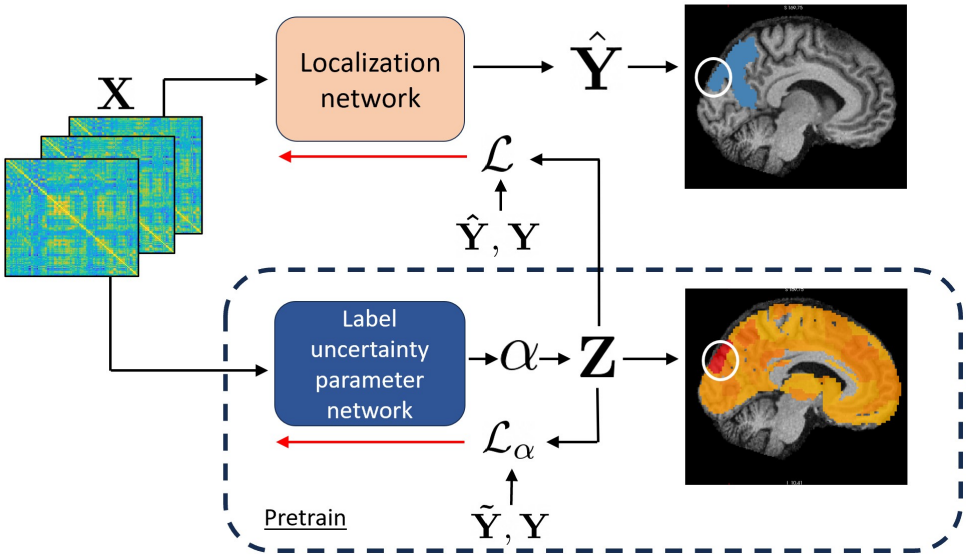

Figure 7.3: Overall block diagram for our setup. **Bottom**: The label uncertainty parameter network is pretrained in a fully supervised manner to learn α from $\mathbf{X}$, as shown by both $\mathbf{Y}$ and $\tilde{\mathbf{Y}}$ being backpropagated through $\mathcal{L}_\alpha$. **Top**: after pretraining, the localization network is trained only with the obseved labels $\mathbf{Y}$.

backpropagation of the two separate neural networks:

$$\frac{\partial Q}{\partial \theta} = \sum_{\tilde{\mathbf{Y}},\mathbf{Z}} P(\tilde{\mathbf{Y}},\mathbf{Z}|\mathbf{Y},\mathbf{X};\theta^{(t)}) \frac{\partial}{\partial \theta} \log P(\mathbf{Y},\tilde{\mathbf{Y}},\mathbf{Z}|\mathbf{X};\theta) \tag{7.7}$$

$$= \sum_{\tilde{\mathbf{Y}}} P(\tilde{\mathbf{Y}}|\mathbf{Y},\mathbf{X};\theta^{(t)}) \frac{\partial}{\partial \theta_1} P(\tilde{\mathbf{Y}}|\mathbf{X};\theta_1) + \sum_{\mathbf{Z}} P(\mathbf{Z}|\mathbf{Y},\mathbf{X};\theta^{(t)}) \frac{\partial}{\partial \theta_2'} P(\mathbf{Z}|\mathbf{X};\theta_2')$$

7.2.2 Deep learning network architecture

7.2.2.1 Overall workflow

An overall diagram for our workflow is shown in Fig. 7.3, where we show the different variables and networks associated with our model. The dynamic connectivity matrices $\mathbf{X}$ are used as input to both the localization network and the label uncertainty parameter network. The localization network is

tasked with producing localization, or classification outputs $\hat{\mathbf{Y}}$ based on the connectivity inputs $\mathbf{X}$. The models presented thus far, such as the one in Fig. 6.8, are what the localization network will be. As shown in the top right, the model outputs $\hat{\mathbf{Y}}$ correspond to class labels (blue).

The label uncertainty parameter network is tasked with predicting α from the connectivity profiles, which is then used to sample from and generate the random variable $\mathbf{Z}$ according to eq. 7.3. This label uncertainty parameter network is strongly supervised during pre-training, where both the true label $\tilde{\mathbf{Y}}$ and the noisy observed label $\mathbf{Y}$ are seen and backpropagated (red arrow) through $\mathcal{L}_\alpha$. However, once the parameter network is pretrained, we do not include information of $\tilde{\mathbf{Y}}$ in the training of the localization network, as shown by block diagram arrows pointing to $\mathcal{L}$, the localization network loss. As shown on the bottom right, the continuous random variable $\mathbf{Z}$ represents a probability of each label being corrupted, based on the connectivity profile. The white circle shows an example where the network had a high uncertainty (shown in red on the bottom right heatmap) of the boundary parcel that it also classified as belonging to the EZ class. Providing the clinician with both of these outputs can be useful in trying to determine where the actual EZ is.

7.2.2.2 Label uncertainty parameter network

Our strategy to identify α_n involves a pre-training phase with full supervision. During this pre-training phase, we have access to both $\mathbf{Y}$ and $\tilde{\mathbf{Y}}$. Fig. 7.4 shows the label uncertainty parameter network (blue box in Fig. 7.3) used for pretraining. As shown in pink, we use the edge-to-edge convolutional

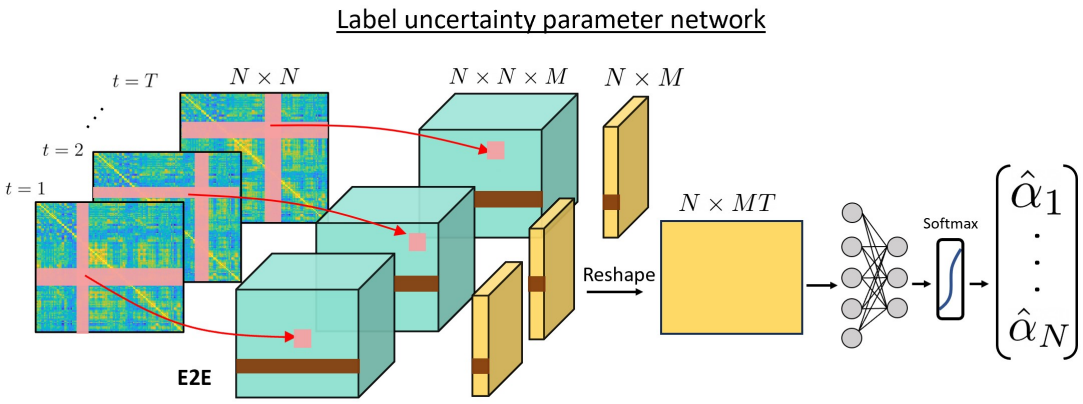

Figure 7.4: We use the edge-to-edge convolutional neural network architecture with an ANN to predict α in a fully supervised manner. This network learns the connectivity patterns associated with mislabeled nodes.

filters designed in [169] to extract intermediate feature maps (green) from the dynamic connectivity inputs. Mathematically, let $m \in \{1, \cdots, M\}$ be the E2E filter index, $\mathbf{r}^m \in \mathbb{R}^{1\times N}$ be the row filter m, $\mathbf{c}^m \in \mathbb{R}^{N\times 1}$ be the column filter m, $\mathbf{b} \in \mathbb{R}^{M\times 1}$ be the E2E bias, and $\boldsymbol{\phi}(.)$ be the activation function. For each time point t the feature map $\mathbf{H}^{m,t} \in \mathbb{R}^{N\times N}$ is computed as follows:

$$\mathbf{H}^{m,t}_{i,j} = \boldsymbol{\phi}\left(\sum_{n=1}^{N} \mathbf{r}^m_n \mathbf{X}^t_{i,n} + \mathbf{c}^m_n \mathbf{X}^t_{n,j} + \mathbf{b}_m\right). \tag{7.8}$$

The edge-to-node (E2N) filter is a 1D convolution along the columns of each feature map. Mathematically, let $\mathbf{g}^m \in \mathbb{R}^{N\times 1}$ be E2N filter m and $\mathbf{p} \in \mathbb{R}^{M\times 1}$ be the E2N bias. The E2N output $\mathbf{h}^{m,t} \in \mathbb{R}^{N\times 1}$ from input $\mathbf{H}^{m,t}$ is computed as

$$\mathbf{h}^{m,t}_i = \boldsymbol{\phi}\left(\sum_{n=1}^{N} \mathbf{g}^m_n \mathbf{H}^{m,t}_{i,n} + \mathbf{p}_m\right). \tag{7.9}$$

We then reshape the representation into a $N \times MT$ matrix to then feed into a 3 layer ANN and obtain predictions α_n. To ensure $0 \leq \alpha_n \leq 1$, we use a softmax layer.

The full supervision of our pretraining phase is reflected in our loss function $\mathcal{L}_\alpha$. Recall that we have access to both $\tilde{\mathbf{Y}}_n, \mathbf{Y}$ using a simulated dataset. We sample $\mathbf{Z}$ and our loss function is

$$\mathcal{L}_\alpha = -\frac{1}{N}\sum_{n=1}^{N} \epsilon_1(1-\Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n))\log(\mathbf{Z}_n) + \epsilon_2\Delta(\mathbf{Y}_n, \tilde{\mathbf{Y}}_n)\log(1-\mathbf{Z}_n). \quad (7.10)$$

The loss function $\mathcal{L}_\alpha$ is a weighted cross entropy function that encourages α_n to be high when the labels $\mathbf{Y}_n, \tilde{\mathbf{Y}}_n$ are different (first term) and encourages α_n to be low when the labels are the same (second term). We introduce the weights ϵ_1, ϵ_2 to handle the class-imbalance, as only a relatively small subset of the nodes will be mislabeled.

7.2.2.3 Combined network training

Once pre-trained, the label uncertainty neural network parameters are set. For this work, we explore the context of noisy labels using the EZ localization network. Therefore, the localization network block in Fig. 7.3 is replaced with our ISBI model shown in Fig. 6.8 during the entire network training. Let $y_n \in \mathbb{R}^{N\times 2}$ be the one-hot encoded version of $\mathbf{Y}_n$ where for nodes belonging to the healthy class, $y_{n,1} = 1, y_{n,2} = 0$ and for nodes belonging to the EZ class, $y_{n,1}0, y_{n,2} = 1$. Let $\hat{y}_n \in \mathbb{R}^{N\times 2}$ be the network outputs for node n. After pretraining is complete, we train the entire network with the following semi-supervised loss function, which only uses knowledge of $\mathbf{Y}$ into account:

$$\mathcal{L} = -\frac{1}{N}\Bigg[\underbrace{(1-\mathbf{Z}_n)\left[y_{n,1}\delta_1 \log \hat{y}_{n,1} + \delta_2 y_{n,2} \log \hat{y}_{n,2}\right]}_{\text{Certain term}} + \underbrace{\mathbf{Z}_n\left(\frac{\log \hat{y}_{n,1} + \log \hat{y}_{n,2}}{2}\right)}_{\text{Uncertain term}}\Bigg]$$

$$+\lambda_1 \underbrace{\left|(1-\mathbf{Z}_n) - \frac{|\sum_{i\in ne(n)} \mathbf{Y}_i - \frac{M}{2}|}{\frac{M}{2}}\right|}_{\text{Neighborhood smoothing term}}. \tag{7.11}$$

As shown, our loss function is broken down into three main terms. The certain term reflects when we are confident about the observed label, i.e., the network believes the label is not corrupted. Recall that since α_n reflects the probability of label corruption, then when $(1-\mathbf{Z}_n)$ is high, we have high confidence in the label being correct. Therefore, we backpropagate the original weighted cross entropy loss as the certain term. The uncertain term reflects the case when we believe label n is incorrect, and therefore we backpropagate the average of the two prediction terms $\log \hat{y}_{n,1}$ and $\log \hat{y}_{n,2}$. The hyperparameters δ_1, δ_2 are the cross entropy weights which help mitigate the class imbalance problem we have, as the majority of nodes considered will belong to the healthy class.

The neighborhood smoothing term in our loss function acts as a biologically inspired regularization term that takes direct spatial neighbors of node n into account, where M is the number of neighbors considered ($M = 6$ in this work). The term $\phi(\mathbf{Y}_n) = \frac{|\sum_{i\in ne(n)} \mathbf{Y}_i - \frac{M}{2}|}{\frac{M}{2}}$ is close to 0 when the neighbors of node n are an even mix of labeled healthy ($\mathbf{Y}_n = 0$) or labeled EZ ($\mathbf{Y}_n = 1$). Given ($\lambda_1 > 0$), when $\phi(\mathbf{Y}_n) = 0$, the term $\lambda_1(1-\mathbf{Z}_n)$ survives,

which encourages α_n to be big, or close to 1 when we minimze the loss during backpropagation. So when the neighbors of n belong to both classes, we are on the boundary of a resection and are more unsure of its label. Likewise, when $\phi(\mathbf{Y}_n) = 1$, our loss encourages α_n to be close to 0, reflecting that we are more sure of the label for nodes with homogeneous neighbors.

7.2.2.4 Prediction on testing data

During a forward pass in the testing phase for subject i, our model recovers both localization predictions $\hat{y}_n^i$ and α_n^i for $n \in \{1, \cdots, N\}$. For our final localization prediction, we want to consolidate both pieces of information. Let $\bar{\alpha}^i = \frac{1}{N}\sum_{n=1}^{N} \alpha_n^i$. During testing, we fuse the predictions and alpha parameters per subject i in the following fashion

$$\hat{y}_{n\,test}^i = \hat{y}_n^i * (1 - \alpha_n^i) + \bar{\alpha}^i. \tag{7.12}$$

Since $\hat{y} > 0.5$ results in a prediction of EZ, equation 7.12 encourages $\hat{y}_{n\,test}^i$ to be smaller, and more likely to be classified as healthy when α_n^i is relatively large. This is due to the fact that the majority of recovered α values are relatively low $(0.1 - 0.2)$, reflecting that we are usually confident about the label in this setup.

7.2.2.5 Implementation details

We implement our network using the PyTorch [250] deep learning library. We pre-train the label uncertainty parameter network using the Adam optimizer for 250 epochs with a learning rate of 0.001 that decays by a factor of 0.8 every 20 epochs using a learning rate scheduler. We set the feature map number

$M = 35$ and set the weighted cross entropy hyperparameters to $\epsilon_1 = 1.1$ and $\epsilon_2 = 0.15$ to handle the class imbalance. Once pretrained, we train the entire network using the Adam optimizer for 150 epochs with a learning rate of 0.0005 that decays by a factor of 0.8 every 10 epochs using a learning rate scheduler. It is important to note that we set the initial learning rate for the label parameter network to 0.00005 during the entire network training to prevent the label uncertainty network parameters from moving too far from the strongly supervised pretraining stage. We set the hyperparameters $\delta_1 = 0.16$, $\delta_2 = 1.2$ and $\lambda_1 = 0.08$ in the overall loss function.

7.3 Experimental results

7.3.1 Simulated dataset

Following our work from [41], we use noise models to simulate the EZ region to obtain $\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}}$. We use the noise model approach outlined in [142] to simulate the EZ in healthy rs-fMRI from 400 HCP subjects. The details associated with the HCP dataset are presented in Chapter 2 of this thesis. To simulate the true EZ signal used for our $\tilde{\mathbf{Y}}$ label, we replace the fMRI time series with one of six noise models: (1) adding normally distributed noise, (2) adding uniformly distributed noise, (3) adding power-law noise, (4) adding Brownian noise, (5) adding noise generated by a Levy walk process, and (6) randomly permuting the time series. For the entire network training and testing, we took 200 subjects and created three separate samples per subject. To prevent bias, we do the pretraining phase with an entirely separate 200 subjects. Fig. 7.5 shows a carton illustration of various potential mislabeling schemes we captured in

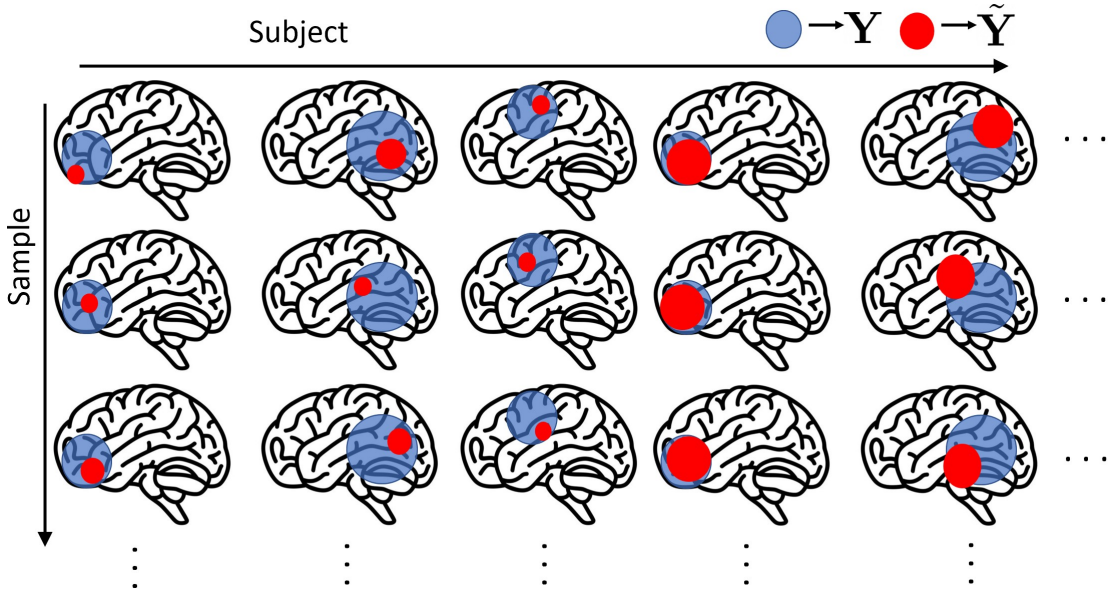

Figure 7.5: A cartoon to illustrate our sampling procedure for creating our simulated dataset. Blue refers to the observed label $\mathbf{Y}$ and red refers to the true label $\tilde{\mathbf{Y}}$. Following what is expected during EZ resections, we make the true label a subset of the observed label, and only modulate the fMRI signal in the true label regions (red). We create a comprehensive training set by including different types of samples, like where $\tilde{\mathbf{Y}}$ is relatively small, large, or on the boundary of the resection.

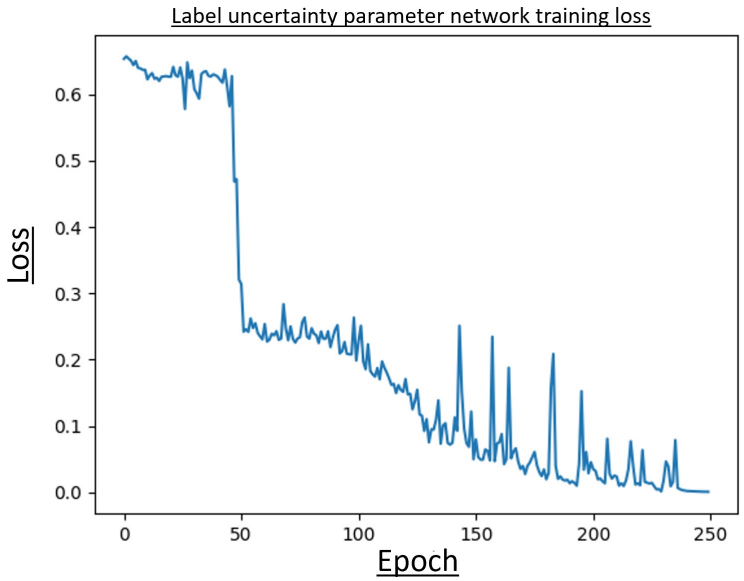

Figure 7.6: The training loss for the label uncertainty parameter network. We see that the network approaches near 0 in the training loss, thus showing a strongly pretrained network.

our dataset, where the red region corresponds to the true label $\tilde{\mathbf{Y}}$ for which the fMRI signal is augmented and the blue corresponds to the observed label $\mathbf{Y}$ where the fMRI signal remains healthy but the region is labeled as belonging to the EZ class. We create a comprehensive training set by including different types of samples, like where $\tilde{\mathbf{Y}}$ is relatively small, large, or on the boundary of $\mathbf{Y}$ and the healthy nodes. While some samples may have $\mathbf{Y}$ be very close to $\tilde{\mathbf{Y}}$, it is important to note that the entire dataset has noisy labels.

7.3.2 Pretraining results

Fig. 7.6 shows the training loss after pre-training the label uncertainty parameter network. As shown, the network eventually approaches a training loss of 0, indicating that there is enough capacity in the network to identify the

mislabeled nodes in a strongly supervised setting. Since we are not introducing knowledge of $\tilde{\mathbf{Y}}$ into the training of the entire network, it is important to note that the effectiveness of our proposed approach is heavily contingent on how well the label uncertainty parameter network can predict α based on the connectivity patterns.

7.3.3 Localization results

Our main experiment involves assessing how well the entire network does with localization, and observing how well the label uncertainty network identifies mislabeled nodes using the simulated noisy dataset. Similar to our previous works, the reported statistics were determined using repeated 10-fold CV, where each run has a different fold membership. We report the mean and standard deviation of the metrics. To demonstrate statistically significant improvement, we perform a t-test (Eq. 4.6) on the repeated 10-fold CV runs, which corrects for the independence assumption between samples [151]. The p-value calculated is with respect to the AUC metric of the proposed method in table 7.1 and with respect to the only localization method in table 7.2. To recall, only the observed $\mathbf{Y}$ label is introduced during the entire network training. We observe the performance of the entire network with the label uncertainty parameter network (proposed) and without (only localization). As a baseline method to evaluate against, we compared with the DeepEZ method proposed in chapter 6 (Fig. 6.1) on static connectivity data. We report test metrics using both $\tilde{\mathbf{Y}}$ (table 7.1) and $\mathbf{Y}$ (table 7.2) to observe how well the network does with the true labels, even when entirely trained on the noisy

Method	Sens	Spec	Acc	AUC	p-value
Proposed	0.45 ± 0.11	0.90 ± 0.07	0.83 ± 0.12	0.71 ± 0.05	
Only localization	0.56 ± 0.14	0.67 ± 0.12	0.71 ± 0.13	0.61 ± 0.05	0.016
DeepEZ	0.34 ± 0.16	0.69 ± 0.13	0.66 ± 0.14	0.56 ± 0.09	0.0012

Table 7.1: Testing metrics using the true labels $\tilde{\mathbf{Y}}$ for metric calculation.

Method	Sens	Spec	Acc	AUC	p-value
Proposed	0.39 ± 0.14	0.88 ± 0.08	0.77 ± 0.11	0.65 ± 0.02	0.40
Only localization	0.46 ± 0.15	0.85 ± 0.09	0.83 ± 0.1	0.68 ± 0.04	
DeepEZ	0.29 ± 0.16	0.8 ± 0.08	0.72 ± 0.11	0.61 ± 0.03	0.041

Table 7.2: Testing metrics using the noisy labels $\mathbf{Y}$ for metric calculation.

labels. To prevent data leakeage and biased results, for the proposed method we pretrain the label uncertainty parameter network with a different subset of subjects than the entire network training and testing.

Table 7.1 shows the testing results when using $\tilde{\mathbf{Y}}$ as the ground truth label. The proposed method including the label uncertainty parameter network outperforms only using the localization network, as shown by a higher accuracy and AUC metric and a p-value of < 0.05. The proposed method maintains a good sensitivity while having a much higher specificity, so it is not suffering from over predictions where the only localization method is. This result highlights that, compared to the only localization method, the proposed method is correctly omitting nodes that are labeled as belonging to the EZ but have a healthy rs-fMRI signature, even though the proposed method is trained on noisy labels. Both methods outperform DeepEZ, which is expected due to the lower parameterization and predictive power of DeepEZ compared to the dynamic transformer model.

The first trend we notice in table 7.2 is that each method performs worse

than our previous results for EZ localization, from the DeepEZ paper and the ISBI 2023 paper. This observation is intuitive because many samples created in the dataset have a large disparity between $\mathbf{Y}$ and $\tilde{\mathbf{Y}}$, and therefore the training is much more poisoned than for our previous work, especially in our ISBI paper where we simulated perfect training labels. Secondly, we notice that the only localization method outperforms the proposed method. Even though the performance difference is not significant ($p = 0.37$), this result is intuitive, as the proposed method is geared to be more conservative with prediction (shown by the smaller sensitivity), and was pre-trained with the true labels.

7.3.3.1 Performance on real data

We evaluate the performance of the proposed, localization only, and DeepEZ method on the 14 subjects from the UW dataset while these three methods are trained entirely on the noisy dataset. Recall that the proposed method is pre-trained using information including the true label as well. We use De Long's test on AUC to show statistical significance in our results. Table 7.3 shows the testing performance, where we observe that the proposed method outperforms the only localization method in AUC and specificity for localization metrics on the real dataset. Even though the proposed method does not outperform the results published in our ISBI paper (Table 6.6), this experiment highlights an interesting result. If trained entirely on a noisy dataset, which is possible in a real world setting, the method from our ISBI paper generalizes much worse to real subjects compared to the proposed method that involves learning the label uncertainty parameter and effectively discarding mislabeled nodes during training. Fig. 7.7-7.8 shows the ground truth (red) labels and predicted (blue)

Method	Sens	Spec	Acc	AUC	p-value
Proposed	0.41	0.91	0.88	0.73	
Only localization	0.51	0.77	0.82	0.68	0.08
DeepEZ	0.25	0.82	0.83	0.59	< 0.01

Table 7.3: Testing metrics using the 14 subjects from the UW dataset.

labels among the three methods considered as well as the recovered α_n values (heat map) for each test subject in the UW dataset. The proposed method accurately localizes regions within the ground truth label while having less false positives than the only localization method. Generally, the proposed method localizes at least one region in the resection, so even though it has a low sensitivity, these predictions are valuable in a clinical setting.

7.3.3.2 Alpha parameter analysis

Contained within Fig.7.7-Fig.7.8 are the recovered α_n parameters shown in a heat map where transparent corresponds to values closer to 0 and solid red to orange corresponds to values closer to 1(column three). We notice some interesting trends in the recovered parameters, such as in subjects 1, 2, 4, 7, 10, 11, 12 where we can see relatively larger α values within parts of the annotated resection zone. This observation captures the phenomenon we are trying to observe, which is parts of the resection zone has healthy tissue. Furthermore, we see some trends where the α_n value is high in regions that results in false positives in the only localization method, such as in subjects 1,4,11,12,13. This observation makes sense, as during training, the α values in synergy with the network predictions combine to encourage less false positives.

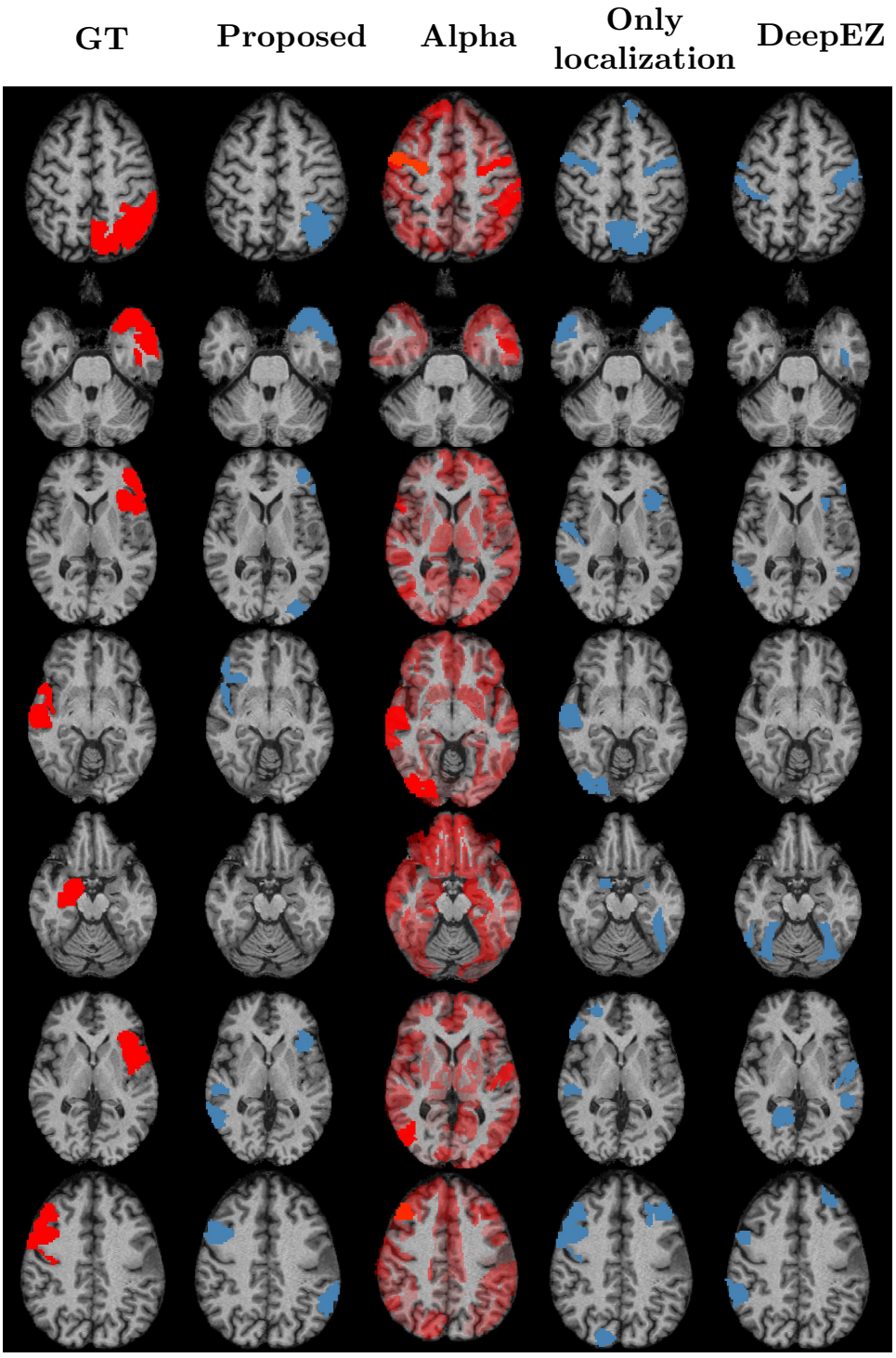

Figure 7.7: Ground truth (red) and predictions (blue) for the first seven subjects in the UW testing dataset. We observe the proposed method suffers from less false positives than the only localization method.

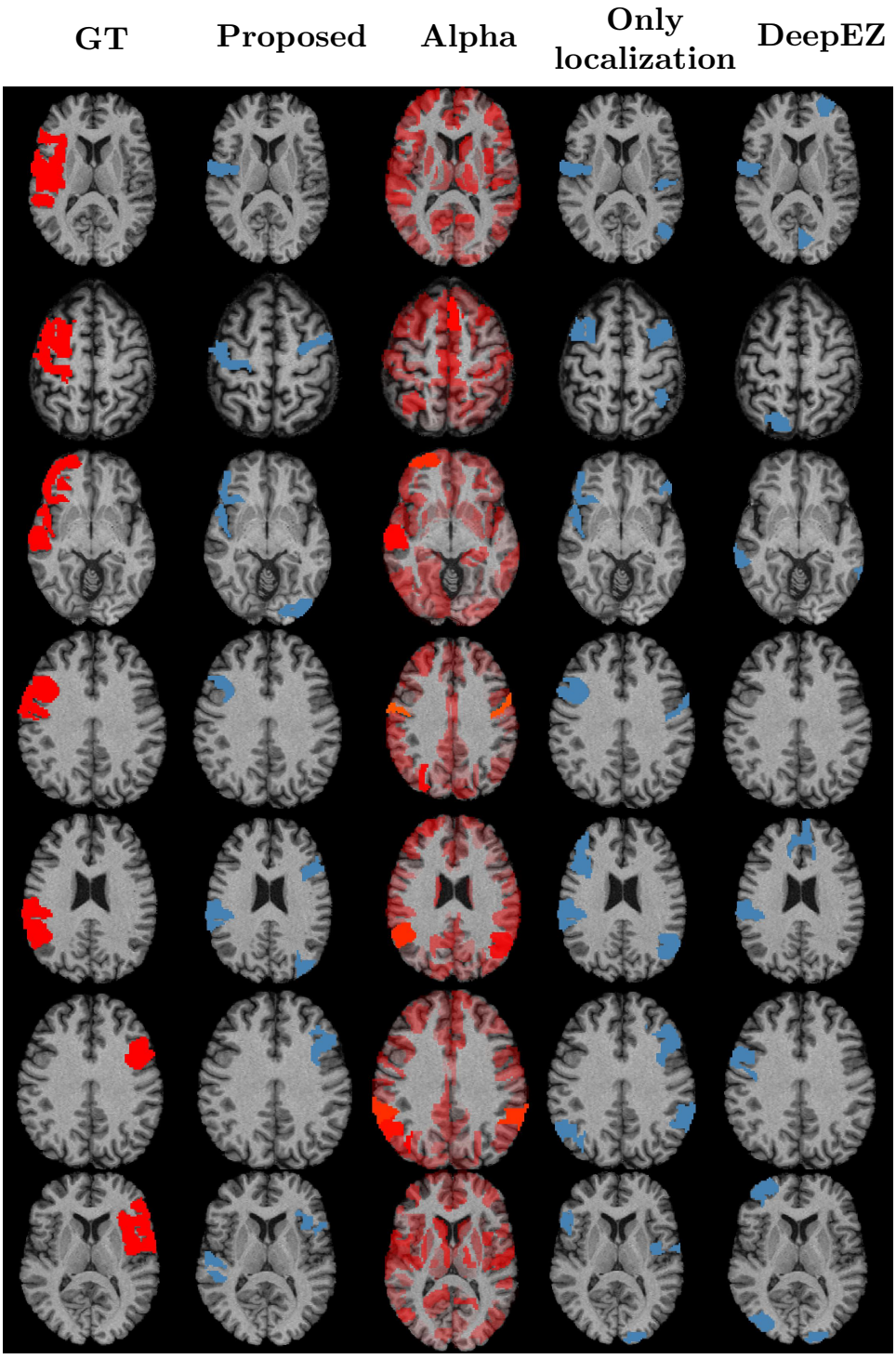

Figure 7.8: Ground truth (red) and predictions (blue) for second seven subjects in the UW testing dataset. We observe the proposed method suffers from less false positives than the only localization method.

7.4 Conclusion

In this section of the thesis, approached the noisy label problem through the context of EZ localization. We assume the probability of mislabeling follows a concrete relaxation of the Bernoulli distribution. We developed a graphical model and derived the EM equations for our setup, and use backpropagation from neural networks to achieve parameter updates. We introduce a deep learning framework that is trained in separate parts to achieve our goal, where we strongly pre-train the label uncertainty parameter network to be able to learn the Bernoulli parameters from a separate dataset. We create an artificial noisy dataset using EZ simulation methods that spans various types of mislabeling schemes we expect to see in the context of EZ localization (i.e. overlapping with the resection, the true label being a small fraction of the resection, etc.). Highlighted by Fig. 7.6, we observe strong pre-training. We show promising results, where even when trained on the noisy labels, our proposed method outperforms the ISBI model when testing with the true labels. We show that the proposed method outperforms the ISBI model on the 14 real focal epilepsy subjects from the UW dataset when trained with the noisy dataset, highlighting the effectiveness of the proposed approach. Finally, we show the predictions on the test subjects for which the proposed method provides a better, more clear localization map than the baselines, as highlighted and guided by the recovered α values.

Chapter 8

Discussion and conclusion

In this chapter, we summarize the main ideas, models, and findings presented in this thesis. The common input data structure to each of our methods is a connectivity graph which summarizes the whole-brain connectivity of a person via their rs-fMRI scan. Overall, we introduce novel deep learning based approaches to perform parcellation refinement and to analyze rs-fMRI connectivity of atypical populations, such as those with brain tumors or having focal epilepsy. We emphasize the application of localization to improve the planning of neurosurgery resection procedures. We test the relatively new hypothesis surrounding dynamic connectivity, that different cognitive systems of interest phase in and out of each other in an interconnected way throughout the course of a rs-fMRI scan, and show localizatiom improvement using models developed for dynamic connectivity. We conclude the thesis with a new exploratory direction in tackling the noisy label problem present in the EZ localization work.

This final chapter of the thesis is organized as follows. We first summarize the main findings from each of the five main chapters in this thesis. We will

do a brief overview of the models developed and the results that we obtained. We will discuss potential drawbacks of the methods developed. We conclude with a discussion of future work and potential extensions of our work to be more valuable from a performance standpoint and in a clinical setting.

8.1 Overview

In chapter 3 we built two separate models that are capable of performing subject-specific parcellation refinement. Our goal in this chapter was to develop these methods to characterize the individual nuances and differences in subjects' neural organization, especially for pathological cases. Our first model (Fig. 3.1) employs a MAP procedure using a markov random fields prior and a correlation coefficient based likelihood term for reassignment. We showed that our MRF model can provide better motor task concordance in tumor patients than existing methods. We then developed RefineNet (Fig. 3.6), the first neural network inspired module to perform parcellation refinement that is also jointly trained on maximizing performance on common rs-fMRI connectivity analysis downstream tasks. We showed that RefineNet can improve downstream task performance for ASD vs NC classification, language localization, and fluid intelligence prediction. We show that a subject-specific approach to parcellation construction via refinement from an established parcellation is a potential avenue for better performance on rs-fMRI analysis tasks.

In chapter 4 we develop our first models applied to connectivity data from brain tumor patients that are tasked with localizing the eloquent cortex.

We leverage specialized convolutional filters designed to extract informative connectivity based features from graphs and developed our original GNN (Fig. 4.2), which was tasked with individually localizing the language and various motor networks in tumor subjects. We extend this work to our multi-task GNN (4.6), which includes a multi-task learning branch and a much larger capacity model that is capable of simultaneously classifying the language and various motor regions of the brain tumor cohort. This improved model is capable of handling missing data due to the loss function, and we performed a vast number of experiments to show its robustness and generalizability, such as including performance on a simulated dataset, varying the parcellation used, performing hyperparameter sweeps and degrading the tumor segmentations. We show clinically promising results, as our methods are able to even capture bilateral language areas during testing even when these cases were not present during the training phase. However, our performance using static connectivity graphs as the input was capped at around $0.76 - 0.83$ AUC, which is not desirable in a clinical setting.

We extend the work from chapter 4 to the dynamic connectivity case in chapter 5, where we introduce novel attention based models to improve eloquent cortex localization. A In our MLCN model (Fig. 5.1), we combine our MTGNN applied to dynamic connectivity inputs with a temporal attention mechanism drawn from an LSTM module. The model extracts two temporal attention vectors, one for language and one for motor. The attention vectors are tasked with identifying which time points are more salient or informative in the downstream node classification. We then extend this work by developing

a spatiotemporal attention model. This model includes a multi-scale spatial attention method that uses 2D convolutions on the intermediate convolutional features and refines them for maximal class separation. The IPMI model (Fig. 5.4), has a temporal attention mechanism as well, and achieves $0.87 - 0.91$ AUC, with a very notable increase in the language AUC compared to the static models. We demonstrated combining neurobiological hypotheses with novel learning architectures to capture the nuances in temporal evolution to ultimately improve eloquent cortex localization in tumor patients from rs-fMRI connectivity.

In chapter 6, we develop models that are able to localize the epileptogenic zone (EZ) in focal epilepsy subjects. We note that for our work, this is a considerably more challenging problem compared to eloquent cortex localization in tumor patients for a few key reasons. First, the rs-fMRI of EZ regions doesn't necessarily have an identifiable signature, or a common connectivity pattern with the rest of the brain. It has been though to be an isolated network with abnormal connectivity patterns. Second, the EZ can span the entire brain, and is not always present in the same region. Third, the dataset available is much smaller and may contain noisy labels due to using the resection area acting as the ground truth. To this end, we developed a graph convolutional network with an anatomical connectivity graph structure to localize the EZ from connectivity graphs. DeepEZ (Fig. 6.1) contains a biologically inspired loss function to account for the intrinsic symmetry found within rs-fMRI connectivity analysis and a subject-specific bias term to help mitigate learning from a small dataset. We observe that DeepEZ outperforms the CNN

architecture used for eloquent cortex localization, and observe that various ablations of DeepEZ do not perform as well as using the combination of aforementioned modelling choices. Later in chapter 6, we improve upon DeepEZ by introducing dynamic connectivity analysis to our modelling choices. We use DeepEZ applied to many input graphs at once and a transformer network to identify temporal attention and consolidate information along the time axis. To handle the small dataset problem, we train our network entirely on an artificially created dataset, where we simulate the EZ in healthy rs-fMRI from the HCP dataset. Therefore, we have a comprehensive large training set to learn from. We see a marked improvement in localization using this combination of data augmentation for training and a dynamic connectivity based model with temporal attention.

Finally, in chapter 7, we conclude the thesis with our last exploratory research endeavor involving learning from datasets with noisy labels, specifically in the context of EZ localization. Following the assumption that the true EZ is most likely only a fraction of the entire resection zone, we create an entire dataset of artifically created noisy labels, where the true label is a fraction of or on the boundary of the observed label, and we only augmented the true label region's fMRI. We assume the probability of mislabeling follows a concrete relaxation of the Bernoulli distribution and develop a graphical model approach to lay out our mathematical assumptions. We then use deep networks to do two tasks: (1) predict the Bernoulli parameter of mislabeling and (2) perform localization. Using a pre-training fully supervised strategy, we train a parameter learning network to learn the probability of a noisy label.

We then only include the observed labels for training our entire setup, which is tasked with performing localization as well as identifying highly uncertain labels. We observe that our proposed method performs better on the true labels than the baseline in the AUC and specificity metrics, suggesting that our pre-training strategy can identify the mislabeled nodes and accurately label them as healthy while maintaining a good sensitivity. We show that our proposed method performs better than the only localization method when testing on the 14 UW subjects, when both methods are trained on the noisy dataset. This results shows promise in our framework, as real EZ datasets are more likely to be noisy than contain perfect labels.

8.1.1 Scope

The models and findings presented in this thesis lay the foundation for continued exploration of rs-fMRI connectivity analysis to localize key regions in tumor and epileptic patients, as well as to provide subject-specific parcellation construction / refinement approaches. We introduce methods to obtain more accurate subject specific parcellations that capture informative differences between subjects for general rs-fMRI analysis. Regarding eloquent cortex localization in brain tumor patients, we developed the first model to perform localization from using the connectivity graph as an input, which contributes to the foundation of using rs-fMRI to aid preoperative planning procedures for tumor removal surgeries. The scope of our EZ work includes broadening the field of non-invasive EZ localization for preoperative mapping via novel

graph convolutional and transformer based networks applied to rs-fMRI connectivity. Finally, we lay the foundational groundwork for identifying noisy labels in epileptic datasets, which can be helpful for future models.

8.2 Limitations and future work

We acknowledge some limitations associated with our work, models and datasets presented in this thesis. The first limitation involves the data presented. Larger datasets will have better predictive power and models trained on larger datasets will be able to generalize better to unseen cases. Generally, our models suffered from low performance, potentially due to having smaller datasets. For example, we only had access to 62 subjects for the tumor dataset and 14 real subjects for the EZ localization work. Even though we included simulated datasets using the publicly available HCP dataset, our datasets did not exceed a few hundred subjects, which is still considered small in a general machine learning sense. Future work could include collecting and curating more subjects' data and observing if our models performance increases.

Another limitation is how our models either lacked available additional data modalities or did not show improvements with using existing additional modalities. For example, we had access to the subject-specific DTI of the focal epilepsy patients, but showed that using their structural connectivity profiles decreased performance in our DeepEZ GCN. Future work involves developing a more intricate model to meaningfully extract information from the additional modalities present, such as DTI for the EZ localization work or the structural MRI for both EZ and eloquent cortex localization. Following

the work of [275], we can improve our parcellation refinement models by incorporating a multi-modal approach as well, where information from DTI or structural T1 images can improve the resulting refined parcellations.

In general, deep learning methods are considered "black box" models that lack interpretability. This presents a large limitation of the models presented in this thesis, as a barrier of including fully deep learning based models in the clinic / hospital include lack of interpretability. If the engineer cannot explain in detail how or why a certain method works, it will not be seen as reliable through the clinician's viewpoint. Furthermore, regarding both eloquent cortex and EZ localization, our models achieved relatively lower AUC (< 0.92), specifically for EZ localization, highlighting how foundational this research is. We would not necessarily be confident in directly implementing our models into the hospital as is. Future work includes training deeper models with larger, more comprehensive datasets to eventually achieve clinically reliable accuracies.

One limitation and future work regarding the parcellation refinement techniques presented is the lack of model exploration. Since the main focus of this thesis has been eloquent cortex and EZ localization, we did not write or publish a journal article on the refinement work. For example, RefineNet has many possible extensions, such as a neighbor analysis, adding layers to the network, or training a similarity network to identify similarity between time series rather than just using the correlation coefficient. Models such as the one in [276] have been developed to train deep networks to identify similarity between time series to replace the standard correlation coefficient.

Adding this module to RefineNet could improve the refinement accuracy by incorporating a data-driven approach for the likelihood term. Furthermore, we did not explore incorporating dynamic connectivity into refinement, where one could observe how the refinement changes over time based on which segment of the rs-fMRI is input into the models and ensemble or average the final predictions.

One limitation and future work of the eloquent cortex localization work is lack of exploration in modelling the tumor. We simply zeroed out the tumor regions in our connectivity graphs. Future work involves modeling the effect that the tumor has on surrounding functionality. Moving towards voxel-level localization is one direction of future work as well, as the parcellations suffer from partial volume effects and a less fine resolution of analysis.

Finally, we acknowledge current limitations associated with our ongoing work on EZ localization using noisy labels. First, we acknowledge that the results presented are only significant in table 7.1, highlighted by a p-value of < 0.05. Even though the proposed method outperforms the localization only method when testing on the 14 real subjects from the UW dataset, the result is not significant and it still does not outperform the localization method when the localization method is trained on perfect labels (our ISBI paper uses perfect labels for training). There are many avenues for future work with this method. The first is experimenting with the various terms in Eq. 7.11, where the uncertain term could be a different penalty to potentially improve optimization and training. Furthermore, we can increase the capacity of the network and increase the number of samples in the dataset. We could include

different variations of mislabeled nodes that could be present in a real world dataset, such as including the contralateral region. Finally, we can extend this framework to the eloquent cortex localization work, which would require different biological assumptions but an overall similar approach.

8.3 Conclusion

To conclude, we've explored four main branches of applying machine and deep learning models to rs-fMRI connectivity analysis: (1) parcellation refinement techniques, (2) eloquent cortex localization in brain tumor patients, (3) EZ localization in focal epilepsy subjects and (4) EZ localization in the presence of noisy labels. We tackled relatively new and unexplored problems using rs-fMRI connectivity graphs, mainly focusing on atypical rs-fMRI and developing subject-specific approaches. We developed models that work well given how small and heterogenous the datasets are. We explored dynamic connectivity models to improve our localization models via leveraging the hypothesis that even at rest, various cognitive systems phase in and out of inter-connected synchrony. Our models capture this phenomenon using specialized deep learning modules. Finally, we conclude via exploring a new and important sub-field in rs-fMRI analysis, and that is the modeling of noisy labels, specifically through the context of EZ localization. The unifying theme of this thesis is that no two brains are the same, and group-level derived methods may not apply well to every case, especially when pathology is present. The work presented in this thesis increases promise of using rs-fMRI as a valuable preoperative mapping tool for eloquent cortex and EZ localization, as well

as a valuable research and clinical tool to investigate the organization of the brain.

References

[1] Albert S Chang and Jeffrey S Ross. "Diagnostic Neuroradiology: CT, MRI, fMRI, MRS, PET, and Octreotide SPECT". In: *Meningiomas* (2009), pp. 55–65.

[2] LP Clarke, RP Velthuizen, MA Camacho, JJ Heine, M Vaidyanathan, LO Hall, RW Thatcher, and ML Silbiger. "MRI segmentation: methods and applications". In: *Magnetic resonance imaging* 13.3 (1995), pp. 343–368.

[3] Karl J Friston, Andrew P Holmes, Keith J Worsley, J-P Poline, Chris D Frith, and Richard SJ Frackowiak. "Statistical parametric maps in functional imaging: a general linear approach". In: *Human brain mapping* 2.4 (1994), pp. 189–210.

[4] Alexander Bowring, Camille Maumet, and Thomas E Nichols. "Exploring the impact of analysis software on task fMRI results". In: *Human brain mapping* 40.11 (2019), pp. 3362–3384.

[5] Bharat Biswal, F Zerrin Yetkin, Victor M Haughton, and James S Hyde. "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI". In: *Magnetic resonance in medicine* 34.4 (1995), pp. 537–541.

[6] Kaiming Li, Lei Guo, Jingxin Nie, Gang Li, and Tianming Liu. "Review of methods for functional brain connectivity detection using fMRI". In: *Computerized medical imaging and graphics* 33.2 (2009), pp. 131–139.

[7] Caio Pinheiro Santana, Emerson Assis de Carvalho, Igor Duarte Rodrigues, Guilherme Sousa Bastos, Adler Diniz de Souza, and Lucelmo Lacerda de Brito. "rs-fMRI and machine learning for ASD diagnosis: A systematic review and meta-analysis". In: *Scientific reports* 12.1 (2022), p. 6030.

[8] Megan H Lee, Christopher D Smyser, and Joshua S Shimony. "Resting-state fMRI: a review of methods and clinical applications". In: *American Journal of neuroradiology* 34.10 (2013), pp. 1866–1872.

[9] Jaroslav Hlinka, Milan Paluš, Martin Vejmelka, Dante Mantini, and Maurizio Corbetta. "Functional connectivity in resting-state fMRI: is linear correlation sufficient?" In: *Neuroimage* 54.3 (2011), pp. 2218–2225.

[10] Dong Wen, Zhenhao Wei, Yanhong Zhou, Guolin Li, Xu Zhang, and Wei Han. "Deep learning methods to process fmri data and their application in the diagnosis of cognitive impairment: a brief overview and our opinion". In: *Frontiers in neuroinformatics* 12 (2018), p. 23.

[11] Wutao Yin, Longhai Li, and Fang-Xiang Wu. "Deep learning for brain disorder diagnosis based on fMRI images". In: *Neurocomputing* 469 (2022), pp. 332–345.

[12] Weibin Feng, Guangyuan Liu, Kelong Zeng, Minchen Zeng, and Ying Liu. "A review of methods for classification and recognition of ASD using fMRI data". In: *Journal of neuroscience methods* 368 (2022), p. 109456.

[13] Abdulaziz Alorf and Muhammad Usman Ghani Khan. "Multi-label classification of Alzheimer's disease stages from resting-state fMRI-based correlation connectivity data and deep learning". In: *Computers in Biology and Medicine* 151 (2022), p. 106240.

[14] Julien Dubois and Ralph Adolphs. "Building a science of individual differences from fMRI". In: *Trends in cognitive sciences* 20.6 (2016), pp. 425–443.

[15] Asif Iqbal, Abd-Krim Seghouane, and Tülay Adalı. "Shared and subject-specific dictionary learning (ShSSDL) algorithm for multisubject fMRI data analysis". In: *IEEE Transactions on Biomedical Engineering* 65.11 (2018), pp. 2519–2528.

[16] Sunil Vasu Kalmady, Russell Greiner, Rimjhim Agrawal, Venkataram Shivakumar, Janardhanan C Narayanaswamy, Matthew RG Brown, Andrew J Greenshaw, Serdar M Dursun, and Ganesan Venkatasubramanian. "Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning". In: *npj Schizophrenia* 5.1 (2019), pp. 1–11.

[17] R Matthew Hutchison, Thilo Womelsdorf, Elena A Allen, Peter A Bandettini, Vince D Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H Duyn, Gary H Glover, Javier Gonzalez-Castillo, et al. "Dynamic functional connectivity: promise, issues, and interpretations". In: *Neuroimage* 80 (2013), pp. 360–378.

[18] Mitchel S Berger, Joseph Kincaid, George A Ojemann, and Ettore Lettich. "Brain mapping techniques to maximize resection, safety, and seizure control in children with brain tumors". In: *Neurosurgery* 25.5 (1989), pp. 786–792.

[19] C Fadul, J Wood, H Thaler, J Galicich, RH Patterson, and JB Posner. "Morbidity and mortality of craniotomy for excision of supratentorial gliomas". In: *Neurology* 38.9 (1988), pp. 1374–1374.

[20] George A Ojemann and Harry A Whitaker. "Language localization and variability". In: *Brain and language* 6.2 (1978), pp. 239–260.

[21] Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain". In: *Neuroimage* 15.1 (2002), pp. 273–289.

[22] Hugues Duffau, Laurent Capelle, Dominique Denvil, Nicole Sichez, Peggy Gatignol, Luc Taillandier, Manuel Lopes, Mary-Christine Mitchell, Sabine Roche, Jean-Charles Muller, et al. "Usefulness of intraoperative electrical subcortical mapping during surgery for low-grade gliomas located within eloquent brain regions: functional results in a consecutive series of 103 patients". In: *Journal of neurosurgery* 98.4 (2003), pp. 764–778.

[23] Alberto Bizzi, Valeria Blasi, Andrea Falini, Paolo Ferroli, Marcello Cadioli, Ugo Danesi, Domenico Aquino, Carlo Marras, Dario Caldiroli, and Giovanni Broggi. "Presurgical functional MR imaging of language and motor functions: validation with intraoperative electrocortical mapping". In: *Radiology* 248.2 (2008), pp. 579–589.

[24] Carlo Giussani, Frank-Emmanuel Roux, Jeffrey Ojemann, Erik Pietro Sganzerla, David Pirillo, and Costanza Papagno. "Is preoperative functional magnetic resonance imaging reliable for language areas mapping in brain tumor surgery? Review of language functional magnetic resonance imaging and direct cortical stimulation correlation studies". In: *Neurosurgery* 66.1 (2010), pp. 113–120.

[25] Salla-Maarit Kokkonen, Juha Nikkinen, Jukka Remes, Jussi Kantola, Tuomo Starck, Marianne Haapea, Juho Tuominen, Osmo Tervonen, and Vesa Kiviniemi. "Preoperative localization of the sensorimotor area using independent component analysis of resting-state fMRI". In: *Magnetic resonance imaging* 27.6 (2009), pp. 733–740.

[26] Diana C Ghinda, Jin-Song Wu, Niall W Duncan, and Georg Northoff. "How much is enough—Can resting state fMRI provide a demarcation for neurosurgical resection in glioma?" In: *Neuroscience & Biobehavioral Reviews* 84 (2018), pp. 245–261.

[27] Warren T Blume, Hans O Lüders, Eli Mizrahi, Carlo Tassinari, Walter van Emde Boas, and Jerome Engel Jr Ex-officio. "Glossary of descriptive terminology for ictal semiology: report of the ILAE task force on classification and terminology". In: *Epilepsia* 42.9 (2001), pp. 1212–1218.

[28] M Scott Perry and Michael Duchowny. "Surgical versus medical treatment for refractory epilepsy: outcomes beyond seizure control". In: *Epilepsia* 54.12 (2013), pp. 2060–2070.

[29] Nitin Tandon et al. "Analysis of morbidity and outcomes associated with use of subdural grids vs stereoelectroencephalography in patients with intractable epilepsy". In: *JAMA neurology* 76.6 (2019), pp. 672–681.

[30] Felix Rosenow and Hans Lüders. "Presurgical evaluation of epilepsy". In: *Brain* 124.9 (2001), pp. 1683–1700.

[31] Borbála Hunyadi, Simon Tousseyn, Patrick Dupont, Sabine Van Huffel, Maarten De Vos, and Wim Van Paesschen. "A prospective fMRI-based technique for localising the epileptogenic zone in presurgical evaluation of epilepsy". In: *Neuroimage* 113 (2015), pp. 329–339.

[32] Davood Karimi, Haoran Dou, Simon K Warfield, and Ali Gholipour. "Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis". In: *Medical image analysis* 65 (2020), p. 101759.

[33] Haris I Sair, Noushin Yahyavi-Firouz-Abadi, Vince D Calhoun, Raag D Airan, Shruti Agarwal, Jarunee Intrapiromkul, Ann S Choe, Sachin K Gujar, Brian Caffo, Martin A Lindquist, et al. "Presurgical brain mapping of the language network in patients with brain tumors using resting-state f MRI: Comparison with task f MRI". In: *Human brain mapping* 37.3 (2016), pp. 913–923.

[34] Naresh Nandakumar, Niharika S D'Souza, Jeff Craley, Komal Manzoor, Jay J Pillai, Sachin K Gujar, Haris I Sair, and Archana Venkataraman. "Defining patient specific functional parcellations in lesional cohorts via Markov random fields". In: *Connectomics in NeuroImaging: Second International Workshop, CNI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 2*. Springer. 2018, pp. 88–98.

[35] Naresh Nandakumar, Komal Manzoor, Shruti Agarwal, Haris I Sair, and Archana Venkataraman. "RefineNet: An Automated Framework to Generate Task and Subject-Specific Brain Parcellations for Resting-State fMRI Analysis". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part I*. Springer. 2022, pp. 315–325.

[36] Naresh Nandakumar, Komal Manzoor, Jay J Pillai, Sachin K Gujar, Haris I Sair, and Archana Venkataraman. "A novel graph neural network to localize eloquent cortex in brain tumor patients from resting-state fmri connectivity". In: *Connectomics in NeuroImaging: Third International Workshop, CNI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 3*. Springer. 2019, pp. 10–20.

[37] Naresh Nandakumar, Komal Manzoor, Shruti Agarwal, Jay J Pillai, Sachin K Gujar, Haris I Sair, and Archana Venkataraman. "Automated eloquent cortex localization in brain tumor patients using multi-task graph neural networks". In: *Medical image analysis* 74 (2021), p. 102203.

[38] Naresh Nandakumar, Niharika Shimona D'souza, Komal Manzoor, Jay J Pillai, Sachin K Gujar, Haris I Sair, and Archana Venkataraman. "A multi-task deep learning framework to localize the eloquent cortex in brain tumor patients using dynamic functional connectivity". In: *Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-oncology: Third International Workshop, MLCN 2020, and Second International Workshop, RNO-AI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings 3*. Springer. 2020, pp. 34–44.

[39] Naresh Nandakumar, Komal Manzoor, Shruti Agarwal, Jay J Pillai, Sachin K Gujar, Haris I Sair, and Archana Venkataraman. "A multi-scale spatial and temporal attention network on dynamic connectivity to localize the eloquent cortex in brain tumor patients". In: *Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings 27*. Springer. 2021, pp. 241–252.

[40] Naresh Nandakumar, David Hsu, Raheel Ahmed, and Archana Venkataraman. "DeepEZ: A Graph Convolutional Network for Automated Epileptogenic Zone Localization from Resting-State fMRI Connectivity". In: *IEEE Transactions on Biomedical Engineering* 70.1 (2022), pp. 216–227.

[41] Naresh Nandakumar et al. “A Deep Learning Framework To Localize the Epileptogenic Zone From Dynamic Functional Connectivity Using A Combined Graph Convolutional and Transformer Network”. In: *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. 2023.

[42] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L Rubin, and Bradley J Erickson. “Deep learning for brain MRI segmentation: state of the art and future directions”. In: *Journal of digital imaging* 30 (2017), pp. 449–459.

[43] Jeff H Duyn, Chrit TW Moonen, Gert H van Yperen, Ruud W de Boer, and Peter R Luyten. “Inflow versus deoxyhemoglobin effects in BOLD functional MRI using gradient echoes at 1.5 T”. In: *NMR in Biomedicine* 7.1-2 (1994), pp. 83–88.

[44] D Rangaprakash, Reza Tadayonnejad, Gopikrishna Deshpande, Joseph O’Neill, and Jamie D Feusner. “FMRI hemodynamic response function (HRF) as a novel marker of brain function: applications for understanding obsessive-compulsive disorder pathology and treatment response”. In: *Brain Imaging and Behavior* 15 (2021), pp. 1622–1640.

[45] Geoffrey Karl Aguirre, Eric Zarahn, and Mark D’Esposito. “The variability of human, BOLD hemodynamic responses”. In: *Neuroimage* 8.4 (1998), pp. 360–369.

[46] Martin J McKeown, Scott Makeig, Greg G Brown, Tzyy-Ping Jung, Sandra S Kindermann, Anthony J Bell, and Terrence J Sejnowski. “Analysis of fMRI data by blind separation into independent spatial components”. In: *Human brain mapping* 6.3 (1998), pp. 160–188.

[47] Joshua S Shimony, Dongyang Zhang, James M Johnston, Michael D Fox, Abhik Roy, and Eric C Leuthardt. “Resting-state spontaneous fluctuations in brain activity: a new paradigm for presurgical planning using fMRI”. In: *Academic radiology* 16.5 (2009), pp. 578–583.

[48] Michael D Fox and Marcus E Raichle. “Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging”. In: *Nature reviews neuroscience* 8.9 (2007), pp. 700–711.

[49] KA Smitha, K Akhil Raja, KM Arun, PG Rajesh, Bejoy Thomas, TR Kapilamoorthy, and Chandrasekharan Kesavadas. “Resting state fMRI: A review on methods in resting state connectivity analysis and resting state networks”. In: *The neuroradiology journal* 30.4 (2017), pp. 305–317.

[50] Nancy J Minshew and Timothy A Keller. "The nature of brain dysfunction in autism: functional brain imaging studies". In: *Current opinion in neurology* 23.2 (2010), pp. 124–130.

[51] Buhari Ibrahim, Subapriya Suppiah, Normala Ibrahim, Mazlyfarina Mohamad, Hasyma Abu Hassan, Nisha Syed Nasser, and M Iqbal Saripan. "Diagnostic power of resting-state fMRI for detection of network connectivity in Alzheimer's disease and mild cognitive impairment: A systematic review". In: *Human Brain Mapping* 42.9 (2021), pp. 2941–2968.

[52] Emilia Sbardella, Francesca Tona, Nikolaos Petsas, and Patrizia Pantano. "DTI measurements in multiple sclerosis: evaluation of brain damage and clinical implications". In: *Multiple sclerosis international* 2013 (2013).

[53] Kenichi Oishi, Michelle M Mielke, Marilyn Albert, Constantine G Lyketsos, and Susumu Mori. "DTI analyses and clinical applications in Alzheimer's disease". In: *Journal of Alzheimer's Disease* 26.s3 (2011), pp. 287–296.

[54] Niharika Shimona D'Souza, Mary Beth Nebel, Deana Crocetti, Joshua Robinson, Stewart Mostofsky, and Archana Venkataraman. "A matrix autoencoder framework to align the functional and structural connectivity manifolds as guided by behavioral phenotypes". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII 24*. Springer. 2021, pp. 625–636.

[55] Klaus Hahn, Nicholas Myers, Sergei Prigarin, Karsten Rodenacker, Alexander Kurz, Hans Förstl, Claus Zimmer, Afra M Wohlschläger, and Christian Sorg. "Selectively and progressively disrupted structural connectivity of functional brain networks in Alzheimer's disease—revealed by a novel framework to analyze edge distributions of networks detecting disruptions with strong statistical evidence". In: *Neuroimage* 81 (2013), pp. 96–109.

[56] Erik SB van Oort, Maarten Mennes, Tobias Navarro Schröder, Vinod J Kumar, Nestor I Zaragoza Jimenez, Wolfgang Grodd, Christian F Doeller, and Christian F Beckmann. "Functional parcellation using time courses of instantaneous connectivity". In: *Neuroimage* 170 (2018), pp. 31–40.

[57] Ricarda I Schubotz, Alfred Anwander, Thomas R Knösche, D Yves von Cramon, and Marc Tittgemeyer. "Anatomical and functional parcellation of the human lateral premotor cortex". In: *Neuroimage* 50.2 (2010), pp. 396–408.

[58] Pantea Moghimi, Anh The Dang, Quan Do, Theoden I Netoff, Kelvin O Lim, and Gowtham Atluri. "Evaluation of functional MRI-based human brain parcellation: a review". In: *Journal of neurophysiology* 128.1 (2022), pp. 197–217.

[59] R Cameron Craddock, G Andrew James, Paul E Holtzheimer III, Xiaoping P Hu, and Helen S Mayberg. "A whole brain fMRI atlas generated via spatially constrained spectral clustering". In: *Human brain mapping* 33.8 (2012), pp. 1914–1928.

[60] Lingzhong Fan, Hai Li, Junjie Zhuo, Yu Zhang, Jiaojian Wang, Liangfu Chen, Zhengyi Yang, Congying Chu, Sangma Xie, Angela R Laird, et al. "The human brainnetome atlas: a new brain atlas based on connectional architecture". In: *Cerebral cortex* 26.8 (2016), pp. 3508–3526.

[61] BT Thomas Yeo, Fenna M Krienen, Jorge Sepulcre, Mert R Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L Roffman, Jordan W Smoller, Lilla Zöllei, Jonathan R Polimeni, et al. "The organization of the human cerebral cortex estimated by intrinsic functional connectivity". In: *Journal of neurophysiology* (2011).

[62] Minqi Chong, Chitresh Bhushan, Anand A Joshi, Soyoung Choi, Justin P Haldar, David W Shattuck, R Nathan Spreng, and Richard M Leahy. "Individual parcellation of resting fMRI with a group functional connectivity prior". In: *NeuroImage* 156 (2017), pp. 87–100.

[63] Thomas Blumensath, Saad Jbabdi, Matthew F Glasser, David C Van Essen, Kamil Ugurbil, Timothy EJ Behrens, and Stephen M Smith. "Spatially constrained hierarchical parcellation of the brain with resting-state fMRI". In: *Neuroimage* 76 (2013), pp. 313–324.

[64] Yinzhi Li, Ningkai Wang, Hao Wang, Yating Lv, Qihong Zou, and Jinhui Wang. "Surface-based single-subject morphological brain networks: Effects of morphological index, brain parcellation and similarity measure, sample size-varying stability and test-retest reliability". In: *NeuroImage* 235 (2021), p. 118018.

[65] Danhong Wang et al. "Parcellating cortical functional networks in individuals". In: *Nature neuroscience* 18.12 (2015), p. 1853.

[66] AM Aertsen, GL Gerstein, MK Habib, and GwtcoPGK Palm. “Dynamics of neuronal firing correlation: modulation of" effective connectivity"”. In: *Journal of neurophysiology* 61.5 (1989), pp. 900–917.

[67] Martijn P Van Den Heuvel and Hilleke E Hulshoff Pol. “Exploring the brain network: a review on resting-state fMRI functional connectivity”. In: *European neuropsychopharmacology* 20.8 (2010), pp. 519–534.

[68] Yao Yu, Dong-Yi Lan, Li-Ying Tang, Ting Su, Biao Li, Nan Jiang, Rong-Bin Liang, Qian-Min Ge, Qiu-Yu Li, and Yi Shao. “Intrinsic functional connectivity alterations of the primary visual cortex in patients with proliferative diabetic retinopathy: a seed-based resting-state fMRI study”. In: *Therapeutic Advances in Endocrinology and Metabolism* 11 (2020), p. 2042018820960296.

[69] Wei Shen, Yiheng Tu, Randy L Gollub, Ana Ortiz, Vitaly Napadow, Siyi Yu, Georgia Wilson, Joel Park, Courtney Lang, Minyoung Jung, et al. “Visual network alterations in brain functional connectivity in chronic low back pain: A resting state functional connectivity and machine learning study”. In: *NeuroImage: Clinical* 22 (2019), p. 101775.

[70] Gert Kristo, Geert-Jan Rutten, Mathijs Raemaekers, Bea de Gelder, Serge ARB Rombouts, and Nick F Ramsey. “Task and task-free FMRI reproducibility comparison for motor network identification”. In: *Human brain mapping* 35.1 (2014), pp. 340–352.

[71] Stephanie Noble, Dustin Scheinost, Emily S Finn, Xilin Shen, Xenophon Papademetris, Sarah C McEwen, Carrie E Bearden, Jean Addington, Bradley Goodyear, Kristin S Cadenhead, et al. “Multisite reliability of MR-based functional connectivity”. In: *Neuroimage* 146 (2017), pp. 959–970.

[72] Christian F Beckmann, Clare E Mackay, Nicola Filippini, Stephen M Smith, et al. “Group comparison of resting-state FMRI data using multi-subject ICA and dual regression”. In: *Neuroimage* 47.Suppl 1 (2009), S148.

[73] Fabrizio Esposito, Adriana Aragri, Ilaria Pesaresi, Sossio Cirillo, Gioacchino Tedeschi, Elio Marciano, Rainer Goebel, and Francesco Di Salle. “Independent component model of the default-mode brain function: combining individual-level and population-level analyses in resting-state fMRI”. In: *Magnetic resonance imaging* 26.7 (2008), pp. 905–913.

[74] Arezoo Alizadeh, Emad Fatemizadeh, and Mohammad Reza Deevband. "Investigation of Brain Default Network's activation in autism spectrum disorders using Group Independent Component Analysis". In: *2014 21th Iranian Conference on Biomedical Engineering (ICBME)*. IEEE. 2014, pp. 177–180.

[75] Maja AA Binnewijzend, Menno M Schoonheim, Ernesto Sanz-Arigita, Alle Meije Wink, Wiesje M van der Flier, Nelleke Tolboom, Sofie M Adriaanse, Jessica S Damoiseaux, Philip Scheltens, Bart NM van Berckel, et al. "Resting-state fMRI changes in Alzheimer's disease and mild cognitive impairment". In: *Neurobiology of aging* 33.9 (2012), pp. 2018–2028.

[76] J Wongsripuemtet, AE Tyan, A Carass, S Agarwal, SK Gujar, JJ Pillai, and HI Sair. "Preoperative mapping of the supplementary motor area in patients with brain tumor using resting-state fMRI with seed-based analysis". In: *American Journal of Neuroradiology* 39.8 (2018), pp. 1493–1498.

[77] Yanmei Tie, Laura Rigolo, Isaiah H Norton, Raymond Y Huang, Wentao Wu, Daniel Orringer, Srinivasan Mukundan Jr, and Alexandra J Golby. "Defining language networks from resting-state fMRI for surgical planning—a feasibility study". In: *Human brain mapping* 35.3 (2014), pp. 1018–1030.

[78] Steven M Stufflebeam, Hesheng Liu, Jorge Sepulcre, Naoaki Tanaka, Randy L Buckner, and Joseph R Madsen. "Localization of focal epileptic discharges using functional connectivity magnetic resonance imaging". In: *Journal of neurosurgery* 114.6 (2011), pp. 1693–1697.

[79] Andrew Sweet, Archana Venkataraman, Steven M Stufflebeam, Hesheng Liu, Naoro Tanaka, Joseph Madsen, and Polina Golland. "Detecting epileptic regions based on global brain connectivity patterns". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part I 16*. Springer. 2013, pp. 98–105.

[80] Clara Huishi Zhang et al. "Lateralization and localization of epilepsy related hemodynamic foci using presurgical fMRI". In: *Clinical Neurophysiology* 126.1 (2015), pp. 27–38.

[81] Mikail Rubinov and Olaf Sporns. "Complex network measures of brain connectivity: uses and interpretations". In: *Neuroimage* 52.3 (2010), pp. 1059–1069.

[82] Amirali Kazeminejad, Soroosh Golbabaei, and Hamid Soltanian-Zadeh. "Graph theoretical metrics and machine learning for diagnosis of Parkinson's disease using rs-fMRI". In: *2017 Artificial Intelligence and Signal Processing Conference (AISP)*. IEEE. 2017, pp. 134–139.

[83] Barnaly Rashid, Eswar Damaraju, Godfrey D Pearlson, and Vince D Calhoun. "Dynamic connectivity states estimated from resting fMRI Identify differences among Schizophrenia, bipolar disorder, and healthy control subjects". In: *Frontiers in human neuroscience* 8 (2014), p. 897.

[84] True Price, Chong-Yaw Wee, Wei Gao, and Dinggang Shen. "Multiple-network classification of childhood autism using functional connectivity dynamics". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part III 17*. Springer. 2014, pp. 177–184.

[85] Martin A Lindquist, Yuting Xu, Mary Beth Nebel, and Brain S Caffo. "Evaluating dynamic bivariate correlations in resting-state fMRI: a comparison study and a new approach". In: *NeuroImage* 101 (2014), pp. 531–546.

[86] Biao Cai, Pascal Zille, Julia M Stephen, Tony W Wilson, Vince D Calhoun, and Yu Ping Wang. "Estimation of dynamic sparse connectivity patterns from resting state fMRI". In: *IEEE transactions on medical imaging* 37.5 (2017), pp. 1224–1234.

[87] Frank Rosenblatt. "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6 (1958), p. 386.

[88] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Communications of the ACM* 60.6 (2017), pp. 84–90.

[89] Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. "Deep learning–based text classification: a comprehensive review". In: *ACM computing surveys (CSUR)* 54.3 (2021), pp. 1–40.

[90] João PS Rosa, Daniel JD Guerra, Nuno CG Horta, Ricardo MF Martins, Nuno CC Lourenço, João PS Rosa, Daniel JD Guerra, Nuno CG Horta, Ricardo MF Martins, and Nuno CC Lourenço. "Overview of artificial neural networks". In: *Using artificial neural networks for analog integrated circuit design automation* (2020), pp. 21–44.

[91] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators". In: *Neural networks* 2.5 (1989), pp. 359–366.

[92] Xavier Glorot and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks". In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings. 2010, pp. 249–256.

[93] Anna Rakitianskaia and Andries Engelbrecht. "Measuring saturation in neural networks". In: *2015 IEEE symposium series on computational intelligence*. IEEE. 2015, pp. 1423–1430.

[94] Si Zhang, Hanghang Tong, Jiejun Xu, and Ross Maciejewski. "Graph convolutional networks: a comprehensive review". In: *Computational Social Networks* 6.1 (2019), pp. 1–23.

[95] Alex Graves and Alex Graves. "Long short-term memory". In: *Supervised sequence labelling with recurrent neural networks* (2012), pp. 37–45.

[96] Barbara Hammer. "On the approximation capability of recurrent neural networks". In: *Neurocomputing* 31.1-4 (2000), pp. 107–123.

[97] Nicha C Dvornek, Pamela Ventola, Kevin A Pelphrey, and James S Duncan. "Identifying autism from resting-state fMRI using long short-term memory networks". In: *Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings 8*. Springer. 2017, pp. 362–370.

[98] Felix Weninger, Hakan Erdogan, Shinji Watanabe, Emmanuel Vincent, Jonathan Le Roux, John R Hershey, and Björn Schuller. "Speech enhancement with LSTM recurrent neural networks and its application to noise-robust ASR". In: *Latent Variable Analysis and Signal Separation: 12th International Conference, LVA/ICA 2015, Liberec, Czech Republic, August 25-28, 2015, Proceedings 12*. Springer. 2015, pp. 91–99.

[99] Adil Moghar and Mhamed Hamiche. "Stock market prediction using LSTM recurrent neural network". In: *Procedia Computer Science* 170 (2020), pp. 1168–1173.

[100] Ralf C Staudemeyer and Eric Rothstein Morris. "Understanding LSTM–a tutorial into long short-term memory recurrent neural networks". In: *arXiv preprint arXiv:1909.09586* (2019).

[101] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).

[102] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate". In: *arXiv preprint arXiv:1409.0473* (2014).

[103] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. "Show, attend and tell: Neural image caption generation with visual attention". In: *International conference on machine learning*. PMLR. 2015, pp. 2048–2057.

[104] Jan K Chorowski, Dzmitry Bahdanau, Dmitriy Serdyuk, Kyunghyun Cho, and Yoshua Bengio. "Attention-based models for speech recognition". In: *Advances in neural information processing systems* 28 (2015).

[105] Som S Biswas. "Potential use of chat gpt in global warming". In: *Annals of biomedical engineering* (2023), pp. 1–2.

[106] Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Noam Shazeer, Alexander Ku, and Dustin Tran. "Image transformer". In: *International conference on machine learning*. PMLR. 2018, pp. 4055–4064.

[107] Yali Qiu, Shuangzhi Yu, Yanhong Zhou, Dongdong Liu, Xuegang Song, Tianfu Wang, and Baiying Lei. "Multi-channel sparse graph transformer network for early alzheimer's disease identification". In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2021, pp. 1794–1797.

[108] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. "Learning representations by back-propagating errors". In: *nature* 323.6088 (1986), pp. 533–536.

[109] Shun-ichi Amari. "Backpropagation and stochastic gradient descent method". In: *Neurocomputing* 5.4-5 (1993), pp. 185–196.

[110] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[111] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting". In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.

[112] Connor Shorten and Taghi M Khoshgoftaar. "A survey on image data augmentation for deep learning". In: *Journal of big data* 6.1 (2019), pp. 1–48.

[113] Rongjian Zhao, Buyue Qian, Xianli Zhang, Yang Li, Rong Wei, Yang Liu, and Yinggang Pan. "Rethinking dice loss for medical image segmentation". In: *2020 IEEE International Conference on Data Mining (ICDM)*. IEEE. 2020, pp. 851–860.

[114] Katarzyna Janocha and Wojciech Marian Czarnecki. "On loss functions for deep neural networks in classification". In: *arXiv preprint arXiv:1702.05659* (2017).

[115] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. "Focal loss for dense object detection". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2980–2988.

[116] Qinwei Fan, Jacek M Zurada, and Wei Wu. "Convergence of online gradient method for feedforward neural networks with smoothing L1/2 regularization penalty". In: *Neurocomputing* 131 (2014), pp. 208–216.

[117] Meenakshi Khosla, Keith Jamison, Gia H Ngo, Amy Kuceyeski, and Mert R Sabuncu. "Machine learning in resting-state fMRI analysis". In: *Magnetic resonance imaging* 64 (2019), pp. 101–121.

[118] Anibal Sólon Heinsfeld, Alexandre Rosa Franco, R Cameron Craddock, Augusto Buchweitz, and Felipe Meneguzzi. "Identification of autism spectrum disorder using deep learning and the ABIDE dataset". In: *NeuroImage: Clinical* 17 (2018), pp. 16–23.

[119] Junghoe Kim, Vince D Calhoun, Eunsoo Shim, and Jong-Hwan Lee. "Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia". In: *Neuroimage* 124 (2016), pp. 127–146.

[120] Meenakshi Khosla, Keith Jamison, Amy Kuceyeski, and Mert R Sabuncu. "Ensemble learning with 3D convolutional neural networks for functional connectome-based prediction". In: *NeuroImage* 199 (2019), pp. 651–662.

[121] Muhammad Naveed Iqbal Qureshi, Jooyoung Oh, and Boreom Lee. "3D-CNN based discrimination of schizophrenia using resting-state fMRI". In: *Artificial intelligence in medicine* 98 (2019), pp. 10–17.

[122] Lebo Wang, Kaiming Li, and Xiaoping P Hu. "Graph convolutional network for fMRI analysis based on connectivity neighborhood". In: *Network Neuroscience* 5.1 (2021), pp. 83–95.

[123] Soham Gadgil, Qingyu Zhao, Adolf Pfefferbaum, Edith V Sullivan, Ehsan Adeli, and Kilian M Pohl. "Spatio-temporal graph convolution for resting-state fMRI analysis". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23*. Springer. 2020, pp. 528–538.

[124] Ahmed El-Gazzar, Mirjam Quaak, Leonardo Cerliani, Peter Bloem, Guido van Wingen, and Rajat Mani Thomas. "A hybrid 3DCNN and 3DC-LSTM based model for 4D spatio-temporal fMRI data: an ABIDE autism classification study". In: *OR 2.0 Context-Aware Operating Theaters and Machine Learning in Clinical Neuroimaging: Second International Workshop, OR 2.0 2019, and Second International Workshop, MLCN 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13 and 17, 2019, Proceedings 2*. Springer. 2019, pp. 95–102.

[125] Jinlong Hu, Yangmin Huang, Nan Wang, and Shoubin Dong. "Brain-NPT: Pre-training of Transformer networks for brain network classification". In: *arXiv preprint arXiv:2305.01666* (2023).

[126] Adriana Di Martino, David O'connor, Bosi Chen, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Joshua H Balsters, Leslie Baxter, Anita Beggiato, Sylvie Bernaerts, et al. "Enhancing studies of the connectome in autism using the autism brain imaging data exchange II". In: *Scientific data* 4.1 (2017), pp. 1–15.

[127] Cameron Craddock, Sharad Sikka, Brian Cheung, Ranjeet Khanuja, Satrajit S Ghosh, Chaogan Yan, Qingyang Li, Daniel Lurie, Joshua Vogelstein, Randal Burns, et al. "Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac)". In: *Front Neuroinform* 42 (2013), pp. 10–3389.

[128] Matthew J McAuliffe, Francois M Lalonde, Delia McGarry, William Gandler, Karl Csaky, and Benes L Trus. "Medical image processing, analysis and visualization in clinical research". In: *Proceedings 14th IEEE Symposium on Computer-Based Medical Systems. CBMS 2001*. IEEE. 2001, pp. 381–386.

[129] William D Penny, Karl J Friston, John T Ashburner, Stefan J Kiebel, and Thomas E Nichols. *Statistical parametric mapping: the analysis of functional brain images*. Elsevier, 2011.

[130] Clifford R Jack Jr, Richard M Thompson, R Kim Butts, Frank W Sharbrough, Patrick J Kelly, Dennis P Hanson, Stephen J Riederer, Richard L Ehman, Nicholas J Hangiandreou, and Gregory D Cascino. "Sensory motor cortex: correlation of presurgical mapping with functional MR imaging and invasive cortical mapping." In: *Radiology* 190.1 (1994), pp. 85–92.

[131] Christopher FA Benjamin, Isha Dhingra, Alexa X Li, Hal Blumenfeld, Rafeed Alkawadri, Stephan Bickel, Christoph Helmstaedter, Stefano Meletti, Richard A Bronen, Simon K Warfield, et al. "Presurgical language fMRI: Technical practices in epilepsy surgical planning". In: *Human brain mapping* 39.10 (2018), pp. 4032–4042.

[132] Jay J Pillai and Domenico Zaca. "Relative utility for hemispheric lateralization of different clinical fMRI activation tasks within a comprehensive language paradigm battery in brain tumor patients as assessed by both threshold-dependent and threshold-independent analysis methods". In: *Neuroimage* 54 (2011), S136–S145.

[133] Yashar Behzadi, Khaled Restom, Joy Liau, and Thomas T Liu. "A component based noise correction method (CompCor) for BOLD and perfusion based fMRI". In: *Neuroimage* 37.1 (2007), pp. 90–101.

[134] Paul K Mazaika, Fumiko Hoeft, Gary H Glover, Allan L Reiss, et al. "Methods and software for fMRI analysis of clinical subjects". In: *Neuroimage* 47.Suppl 1 (2009), S58.

[135] JVNP Engel Jr. "Outcome with respect to epileptic seizures." In: *Surgical treatment of the epilepsies* (1993), pp. 609–621.

[136] Colin D Binnie and Charles E Polkey. "Commission on neurosurgery of the international league against epilepsy (ILAE) 1993–1997: recommended standards". In: *Epilepsia* 41.10 (2000), pp. 1346–1349.

[137] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. "The WU-Minn human connectome project: an overview". In: *Neuroimage* 80 (2013), pp. 62–79.

[138] Stephen M Smith, Christian F Beckmann, Jesper Andersson, Edward J Auerbach, Janine Bijsterbosch, Gwenaëlle Douaud, Eugene Duff, David A Feinberg, Ludovica Griffanti, Michael P Harms, et al. "Resting-state fMRI in the human connectome project". In: *Neuroimage* 80 (2013), pp. 144–168.

[139] Jeffrey R Binder, William L Gross, Jane B Allendorfer, Leonardo Bonilha, Jessica Chapin, Jonathan C Edwards, Thomas J Grabowski, John T Langfitt, David W Loring, Mark J Lowe, et al. "Mapping anterior temporal lobe language areas with fMRI: a multicenter normative study". In: *Neuroimage* 54.2 (2011), pp. 1465–1475.

[140] Randy L Buckner, Fenna M Krienen, Angela Castellanos, Julio C Diaz, and BT Thomas Yeo. "The organization of the human cerebellum estimated by intrinsic functional connectivity". In: *Journal of neurophysiology* 106.5 (2011), pp. 2322–2345.

[141] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. "Fsl". In: *Neuroimage* 62.2 (2012), pp. 782–790.

[142] Patrick H Luckett, Luigi Maccotta, John J Lee, Ki Yun Park, Nico UF Dosenbach, Beau M Ances, Robert Edward Hogan, Joshua S Shimony, and Eric C Leuthardt. "Deep learning resting state functional magnetic resonance imaging lateralization of temporal lobe epilepsy". In: *Epilepsia* 63.6 (2022), pp. 1542–1552.

[143] Gregory Kiar et al. "A comprehensive cloud framework for accurate and reliable human connectome estimation and meganalysis". In: *bioRxiv* (2017), p. 188706.

[144] Archana Venkataraman et al. "Bayesian community detection in the space of group-level functional differences". In: *IEEE transactions on medical imaging* 35.8 (2016), pp. 1866–1882.

[145] Srikanth Ryali et al. "A parcellation scheme based on von Mises-Fisher distributions and Markov random fields for segmenting brain regions using resting-state fMRI". In: *NeuroImage* 65 (2013), pp. 83–96.

[146] Martin J Wainwright, Michael I Jordan, et al. "Graphical models, exponential families, and variational inference". In: *Foundations and Trends® in Machine Learning* 1.1–2 (2008), pp. 1–305.

[147] Frank R Kschischang, Brendan J Frey, and H-A Loeliger. "Factor graphs and the sum-product algorithm". In: *IEEE Transactions on information theory* 47.2 (2001), pp. 498–519.

[148] Won Hee Lee and Sophia Frangou. "Linking functional connectivity and dynamic properties of resting-state networks". In: *Scientific reports* 7.1 (2017), p. 16610.

[149] Niharika Shimona Dsouza, Mary Beth Nebel, Deana Crocetti, Joshua Robinson, Stewart Mostofsky, and Archana Venkataraman. "M-gcn: A multimodal graph convolutional network to integrate functional and structural connectomics data to predict multidimensional phenotypic characterizations". In: *Medical Imaging with Deep Learning*. PMLR. 2021, pp. 119–130.

[150] Jin Zhang, Fan Feng, Tianyi Han, Xiaoli Gong, and Feng Duan. "Detection of Autism Spectrum Disorder using fMRI Functional Connectivity with Feature Selection and Deep Learning". In: *Cognitive Computation* (2022), pp. 1–12.

[151] Remco R Bouckaert and Eibe Frank. "Evaluating the replicability of significance tests for comparing learning algorithms". In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer. 2004, pp. 3–12.

[152] Hugues Duffau. "Lessons from brain mapping in surgery for low-grade glioma: insights into associations between tumour and brain plasticity". In: *The Lancet Neurology* 4.8 (2005), pp. 476–486.

[153] Raymond Sawaya, Maarouf Hammoud, Derek Schoppa, Kenneth R Hess, Shu Z Wu, Wei-Ming Shi, and David M WiIdrick. "Neurosurgical outcomes in a modern series of 400 craniotomies for treatment of parenchymal tumors". In: *Neurosurgery* 42.5 (1998), pp. 1044–1055.

[154] Deepak Kumar Gupta, PS Chandra, BK Ojha, BS Sharma, AK Mahapatra, and VS Mehta. "Awake craniotomy versus surgery under general anesthesia for resection of intrinsic lesions of eloquent cortex—a prospective randomised study". In: *Clinical neurology and neurosurgery* 109.4 (2007), pp. 335–343.

[155] Isaac Yang and Giyarpuram N Prashant. "Advances in the surgical resection of temporo-parieto-occipital junction gliomas". In: *New Techniques for Management of'Inoperable'Gliomas*. Elsevier, 2019, pp. 73–87.

[156] Hussein Kekhia, Laura Rigolo, Isaiah Norton, and Alexandra J Golby. "Special surgical considerations for functional brain mapping". In: *Neurosurgery Clinics* 22.2 (2011), pp. 111–132.

[157] Cristina Rosazza, Domenico Aquino, Ludovico D'Incerti, Roberto Cordella, Adrian Andronache, Domenico Zacà, Maria Grazia Bruzzone, Giovanni Tringali, and Ludovico Minati. "Preoperative mapping of the sensorimotor cortex: comparative assessment of task-based and resting-state FMRI". In: *PLoS One* 9.6 (2014), e98860.

[158] Jeffrey R Binder, Sara J Swanson, Thomas A Hammeke, George L Morris, Wade M Mueller, Mariellen Fischer, S Benbadis, Julie A Frost, Stephen M Rao, and Victor M Haughton. "Determination of language dominance using functional MRI: a comparison with the Wada test". In: *Neurology* 46.4 (1996), pp. 978–984.

[159] Ralph O Suarez, Stephen Whalen, Aaron P Nelson, Yanmei Tie, Mary-Ellen Meadows, Alireza Radmanesh, and Alexandra J Golby. "Threshold-independent functional MRI determination of language dominance: a validation study against clinical gold standards". In: *Epilepsy & Behavior* 16.2 (2009), pp. 288–297.

[160] Megan H Lee, Michelle M Miller-Thomas, Tammie L Benzinger, Daniel S Marcus, Carl D Hacker, Eric C Leuthardt, and Joshua S Shimony. "Clinical resting-state fMRI in the preoperative setting: are we ready for prime time?" In: *Topics in magnetic resonance imaging: TMRI* 25.1 (2016), p. 11.

[161] Archana Venkataraman, Thomas J Whitford, Carl-Fredrik Westin, Polina Golland, and Marek Kubicki. "Whole brain resting state functional connectivity abnormalities in schizophrenia". In: *Schizophrenia research* 139.1-3 (2012), pp. 7–12.

[162] Eric C Leuthardt, Gloria Guzman, S Kathleen Bandt, Carl Hacker, Ananth K Vellimana, David Limbrick, Mikhail Milchenko, Pamela Lamontagne, Benjamin Speidel, Jarod Roland, et al. "Integration of resting state functional MRI into clinical practice-A large single institution experience". In: *PloS one* 13.6 (2018), e0198349.

[163] Reinhard J Tomczak, Arthur P Wunderlich, Yang Wang, Veit Braun, Gregor Antoniadis, Johannes Görich, Hans-Peter Richter, and Hans-Jürgen Brambs. "fMRI for preoperative neurosurgical mapping of motor cortex and language in a clinical setting". In: *Journal of computer assisted tomography* 24.6 (2000), pp. 927–934.

[164] Georg Langs, Yanmei Tie, Laura Rigolo, Alexandra Golby, and Polina Golland. "Functional geometry alignment and localization of brain areas". In: *Advances in neural information processing systems*. 2010, pp. 1225–1233.

[165] Georg Langs, Andrew Sweet, Danial Lashkari, Yanmei Tie, Laura Rigolo, Alexandra J Golby, and Polina Golland. "Decoupling function and anatomy in atlases of functional connectivity patterns: Language mapping in tumor patients". In: *Neuroimage* 103 (2014), pp. 462–475.

[166] Carl D Hacker, Timothy O Laumann, Nicholas P Szrama, Antonello Baldassarre, Abraham Z Snyder, Eric C Leuthardt, and Maurizio Corbetta. "Resting state network estimation in individual subjects". In: *Neuroimage* 82 (2013), pp. 616–633.

[167] Konstantinos Kamnitsas et al. "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation". In: *MedIA* 36 (2017), pp. 61–78.

[168] Meenakshi Khosla et al. "3D Convolutional Neural Networks for Classification of Functional Connectomes". In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 137–145.

[169] Jeremy Kawahara, Colin J Brown, Steven P Miller, Brian G Booth, Vann Chau, Ruth E Grunau, Jill G Zwicker, and Ghassan Hamarneh. "BrainNetCNN: convolutional neural networks for brain networks; towards predicting neurodevelopment". In: *NeuroImage* 146 (2017), pp. 1038–1049.

[170] Sebastian Ruder. "An overview of multi-task learning in deep neural networks". In: *arXiv preprint arXiv:1706.05098* (2017).

[171] Cha Zhang and Zhengyou Zhang. "Improving multiview face detection with multi-task deep convolutional neural networks". In: *IEEE Winter Conference on Applications of Computer Vision*. IEEE. 2014, pp. 1036–1041.

[172] Juan Martino, Susanne M Honma, Anne M Findlay, Adrian G Guggisberg, Julia P Owen, Heidi E Kirsch, Mitchel S Berger, and Srikantan S Nagarajan. "Resting functional connectivity in patients with brain tumors in eloquent areas". In: *Annals of neurology* 69.3 (2011), pp. 521–532.

[173] Rich Caruana. "Multitask learning". In: *Machine learning* 28.1 (1997), pp. 41–75.

[174] Meredith Gabriel, Nicole P Brennan, Kyung K Peck, and Andrei I Holodny. "Blood oxygen level dependent functional magnetic resonance imaging for presurgical planning". In: *Neuroimaging Clinics* 24.4 (2014), pp. 557–571.

[175] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. "Brain tumor segmentation with deep neural networks". In: *Medical image analysis* 35 (2017), pp. 18–31.

[176] Ali Işın, Cem Direkoğlu, and Melike Şah. "Review of MRI-based brain tumor image segmentation using deep learning methods". In: *Procedia Computer Science* 102 (2016), pp. 317–324.

[177] Nathalie Tzourio-Mazoyer, G Josse, Fabrice Crivello, and Bernard Mazoyer. "Interindividual variability in the hemispheric organization for speech". In: *Neuroimage* 21.1 (2004), pp. 422–435.

[178] Tore Opsahl et al. "Node centrality in weighted networks: Generalizing degree and shortest paths". In: *Social networks* 32.3 (2010), pp. 245–251.

[179] Sundaram Suresh et al. "Risk-sensitive loss functions for sparse multi-category classification problems". In: *Information Sciences* 178.12 (2008), pp. 2621–2638.

[180] Pedro J García-Laencina, Jesús Serrano, Aníbal R Figueiras-Vidal, and José-Luis Sancho-Gómez. "Multi-task neural networks for dealing with missing inputs". In: *International Work-Conference on the Interplay Between Natural and Artificial Computation*. Springer. 2007, pp. 282–291.

[181] Jintao Zhang and Jun Huan. "Inductive multi-task learning with multiple view data". In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2012, pp. 543–551.

[182] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. "Automatic differentiation in pytorch". In: (2017).

[183] Santo Fortunato. "Community detection in graphs". In: *Physics reports* 486.3-5 (2010), pp. 75–174.

[184] Marcel A de Reus and Martijn P Van den Heuvel. "The parcellation-based connectome: limitations and extensions". In: *Neuroimage* 80 (2013), pp. 397–404.

[185] Anton Lord, Stefan Ehrlich, Viola Borchardt, Daniel Geisler, Maria Seidel, Stefanie Huber, Julia Murr, and Martin Walter. "Brain parcellation choice affects disease-related topology differences increasingly from global to local network levels". In: *Psychiatry Research: Neuroimaging* 249 (2016), pp. 12–19.

[186] Luis Perez and Jason Wang. "The effectiveness of data augmentation in image classification using deep learning". In: *arXiv preprint arXiv:1712.04621* (2017).

[187] Khandakar M Rashid and Joseph Louis. "Times-series data augmentation and deep learning for construction equipment activity recognition". In: *Advanced Engineering Informatics* 42 (2019), p. 100944.

[188] Qingyu Zhao, Ehsan Adeli, and Kilian M Pohl. "Training confounder-free deep learning models for medical applications". In: *Nature communications* 11.1 (2020), p. 6010.

[189] Nicha C Dvornek, Xiaoxiao Li, Juntang Zhuang, and James S Duncan. "Jointly discriminative and generative recurrent neural networks for learning from fmri". In: *International Workshop on Machine Learning in Medical Imaging*. Springer. 2019, pp. 382–390.

[190] Weizheng Yan, Han Zhang, Jing Sui, and Dinggang Shen. "Deep chronnectome learning via full bidirectional long short-term memory networks for MCI diagnosis". In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2018, pp. 249–257.

[191] Barnaly Rashid, Mohammad R Arbabshirani, Eswar Damaraju, Mustafa S Cetin, Robyn Miller, Godfrey D Pearlson, and Vince D Calhoun. "Classification of schizophrenia and bipolar patients using static and dynamic resting-state fMRI brain connectivity". In: *Neuroimage* 134 (2016), pp. 645–657.

[192] Sanghyun Woo et al. "Cbam: Convolutional block attention module". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 3–19.

[193] James Michael Kunert-Graf, Kristian Eschenburg, David Galas, J Nathan Kutz, Swati Rane, and Bingni Wen Brunton. "Extracting reproducible time-resolved resting state networks using dynamic mode decomposition". In: *Frontiers in computational neuroscience* 13 (2019), p. 75.

[194] SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. "Convolutional LSTM network: A machine learning approach for precipitation nowcasting". In: *Advances in neural information processing systems*. 2015, pp. 802–810.

[195] Hongming Li and Yong Fan. "Brain decoding from functional MRI using long short-term memory recurrent neural networks". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 320–328.

[196] D Tomasi and ND Volkow. "Language network: segregation, laterality and connectivity". In: *Molecular psychiatry* 17.8 (2012), p. 759.

[197] Jie Chen et al. “Multi-scale spatial and channel-wise attention for improving object detection in remote sensing imagery”. In: *IEEE Geoscience and Remote Sensing Letters* 17.4 (2019), pp. 681–685.

[198] Sergey Zagoruyko and Nikos Komodakis. “Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer”. In: *arXiv preprint arXiv:1612.03928* (2016).

[199] Alexander Schaefer et al. “Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI”. In: *Cerebral cortex* (2018).

[200] W.J. Cadarelli and B.J. Smith. “The Burden of Epilepsy to Patients and Payers”. In: *American Journal of Managed Care* 16.12 (2010), S331–336.

[201] Katharina Allers et al. “The economic impact of epilepsy: a systematic review”. In: *BMC neurology* 15.1 (2015), pp. 1–16.

[202] Rekha Dwivedi et al. “Surgery for drug-resistant epilepsy in children”. In: *New England Journal of Medicine* 377.17 (2017), pp. 1639–1647.

[203] F. Rosenow and H. Luders. “Presurgical Evaluation of Epilepsy”. In: *Brain* 124.9 (2001), pp. 1683–1700.

[204] M. Mohan et al. “The Long-Term Outcomes of Epilepsy Surgery”. In: *PLOS ONE* 5 (2018), pp. 1–16.

[205] M.O. Krucoff et al. “Rates and Predictors of Success and Failure in Repeat Epilepsy Surgery: A Meta-Analysis and Systematic Review”. In: *Epilepsia* 58.12 (2017), pp. 2133–2142.

[206] Abbas Sohrabpour, Zhengxiang Cai, Shuai Ye, Benjamin Brinkmann, Gregory Worrell, and Bin He. “Noninvasive electromagnetic source imaging of spatiotemporally distributed epileptogenic brain sources”. In: *Nature communications* 11.1 (2020), pp. 1–15.

[207] Shuai Ye, Lin Yang, Yunfeng Lu, Michal T Kucewicz, Benjamin Brinkmann, Cindy Nelson, Abbas Sohrabpour, Gregory A Worrell, and Bin He. “Contribution of Ictal Source Imaging for Localizing Seizure Onset Zone in Patients With Focal Epilepsy”. In: *Neurology* 96.3 (2021), e366–e375.

[208] Giovanni Pellegrino, Tanguy Hedrich, Rasheda Chowdhury, Jeffery A Hall, Jean-Marc Lina, Francois Dubeau, Eliane Kobayashi, and Christophe Grova. “Source localization of the seizure onset zone from ictal EEG/MEG data”. In: *Human brain mapping* 37.7 (2016), pp. 2528–2546.

[209] Chris Plummer, Simon J Vogrin, William P Woods, Michael A Murphy, Mark J Cook, and David TJ Liley. "Interictal and ictal source localization for epilepsy surgery using high-density EEG with MEG: a prospective long-term study". In: *Brain* 142.4 (2019), pp. 932–951.

[210] B.N. Cuffin. "EEG Localization Accuracy Improvements using Realistically Shaped Head Models". In: *IEEE Trans Biomedical Engineering* 43 (1996), pp. 299–393.

[211] A. Crouzeix et al. "An Evaluation of Dipole Reconstruction Accuracy with Spherical and Realistic Head Models in MEG". In: *Clinical Neurophysiology* 110 (1999), pp. 2176–2188.

[212] V. Jurcak, D. Tsuzuki, and I. Dan. "10/20, 10/10, and 10/5 systems revisited: Their Validity as Relative Head-Surface-Based Positioning Systems". In: *NeuroImage* 34 (4 2007), pp. 1600–1611.

[213] Anto I. Bagic and Richard C. Burgess. "Utilization of MEG Among the US Epilepsy Centers: A Survey-Based Appraisal". In: *Journal of Clinical Neurophysiology* 37.6 (2020).

[214] David Ahmedt-Aristizabal et al. "Neural memory networks for seizure type classification". In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2020, pp. 569–575.

[215] Subhrajit Roy et al. "Machine learning for seizure type classification: setting the benchmark". In: *arXiv preprint arXiv:1902.01012* (2019).

[216] I.R. Dwi Saputro et al. "Seizure Type Classification on EEG Signal using Support Vector Machine". In: *Journal of Physics: Conference Series* 1201 (2019), p. 012065.

[217] M. El Azami et al. "Detection of Lesions Underlying Intractable Epilepsy on T1-Weighted MRI as an Outlier Detection Problem". In: *PLOS ONE* 11.9 (2016), pp. 1–21.

[218] S.B Antel et al. "Automated Detection of Focal Cortical Dysplasia Lesions using Computational Models of their MRI Characteristics and Texture Analysis". In: *NeuroImage* 19.4 (2003), pp. 1748 –1759. ISSN: 1053-8119.

[219] S. Srivastava et al. "Feature-Based Statistical Analysis of Structural MR Data for Automatic Detection of Focal Cortical Dysplastic Lesions". In: *NeuroImage* 27 (2 2005), pp. 253–266.

[220] O. Colliot et al. "Individual Voxel-Based Analysis of Gray Matter in Focal Cortical Dysplasia". In: *NeuroImage* 29.1 (2006), pp. 162 –171.

[221] S.-J. Hong et al. “Automated Detection of Cortical Dysplasia Type II in MRI-Negative Epilepsy”. In: *Neurology* 83.1 (2014), pp. 48–55.

[222] B. Ahmed et al. “Cortical Feature Analysis and Machine Learning Improves Detection of “MRI-Negative" Focal Cortical Dysplasia”. In: *Epilepsy & Behavior* 48 (2015), pp. 21–28.

[223] R. Rodríguez-Cruces and L. Concha. “White Matter in Temporal Lobe Epilepsy: Clinico-Pathological Correlates of Water Diffusion Abnormalities”. In: *Quantitative Imaging in Medicine and Surgery* 5 (2 2015), pp. 264–278.

[224] F. Deleo et al. “Histological and MRI Markers of White Matter Damage in Focal Epilepsy”. In: *Epilepsy Research* 140 (2018), pp. 29 –38.

[225] D. Ding et al. “Cerebral Arteriovenous Malformations and Epilepsy, Part 1: Predictors of Seizure Presentation”. In: *World Neurosurgery* 84.3 (2015), pp. 645 –652.

[226] F. Turjman et al. “Epilepsy associated with cerebral arteriovenous malformations: a multivariate analysis of angioarchitectural characteristics.” In: *American Journal of Neuroradiology* 16.2 (1995), pp. 345–350.

[227] B. Garcin et al. “Epileptic seizures at initial presentation in patients with brain arteriovenous malformation”. In: *Neurology* 78.9 (2012), pp. 626–631.

[228] Y. Wang et al. “Voxel-based automated detection of focal cortical dysplasia lesions using diffusion tensor imaging and T2-weighted MRI data”. In: *Epilepsy & Behavior* 84 (2018), pp. 127–134.

[229] S. Adler et al. “Novel surface features for automated detection of focal cortical dysplasias in paediatric epilepsy”. In: *NeuroImage: Clinical* 14 (2016), pp. 18–27.

[230] J.-J. Mo et al. “Clinical Value of Machine Learning in the Automated Detection of Focal Cortical Dysplasia Using Quantitative Multimodal Surface-Based Features”. In: *Frontiers in Neuroscience* 12 (2019), p. 1008.

[231] Ingmar Blümcke et al. “The Clinicopathologic Spectrum of Focal Cortical Dysplasias: A Consensus Classification Proposed by an Ad Hoc Task Force of the ILAE Diagnostic Methods Commission”. In: *Epilepsia* 52.1 (2011), pp. 158–174.

[232] C. Shepherd et al. “A Quantitative Study of White Matter Hypomyelination and Oligodendroglial Maturation in Focal Cortical Dysplasia Type II”. In: *Epilepsia* 54.5 (2013), pp. 898–908.

[233] C.J. Cook et al. “Effective Connectivity Within the Default Mode Network in Left Temporal Lobe Epilepsy: Findings from the Epilepsy Connectome Project”. In: *Brain Connectivity* 9.2 (2019), pp. 174–183.

[234] G. Hwang et al. “Using Low-Frequency Oscillations to Detect Temporal Lobe Epilepsy with Machine Learning”. In: *Brain Connectivity* 9.2 (2019), pp. 184–193.

[235] Eric van Diessen et al. “Functional and structural brain networks in epilepsy: what have we learned?” In: *Epilepsia* 54.11 (2013), pp. 1855–1865.

[236] P.J. Basser and C. Pierpaoli. “Microstructural and Physiological Features of Tissues Elucidated by Quantitative-Diffusion-Tensor MRI”. In: *Journal of Magnetic Resonance* 111 (1996), pp. 209–219.

[237] Gaëlle E Doucet et al. “Early and late age of seizure onset have a differential impact on brain resting-state organization in temporal lobe epilepsy”. In: *Brain topography* 28.1 (2015), pp. 113–126.

[238] Victoria L Morgan et al. “Evolution of functional connectivity of brain networks and their dynamic interaction in temporal lobe epilepsy”. In: *Brain connectivity* 5.1 (2015), pp. 35–44.

[239] C Garcia-Ramos et al. “Low functional robustness in mesial temporal lobe epilepsy”. In: *Epilepsy research* 123 (2016), pp. 20–28.

[240] Mohsen Mazrooyisebdani et al. “Graph theory analysis of functional connectivity combined with machine learning approaches demonstrates widespread network differences and predicts clinical variables in temporal lobe epilepsy”. In: *Brain connectivity* 10.1 (2020), pp. 39–50.

[241] Aaron F Struck et al. “Regional and global resting-state functional MR connectivity in temporal lobe epilepsy: Results from the Epilepsy Connectome Project”. In: *Epilepsy & Behavior* 117 (2021), p. 107841.

[242] Eunji Jun et al. “Identifying resting-state effective connectivity abnormalities in drug-naïve major depressive disorder diagnosis via graph convolutional networks”. In: *Human Brain Mapping* 41.17 (2020), pp. 4997–5014.

[243] Dongren Yao et al. “Triplet graph convolutional network for multi-scale analysis of functional connectivity using functional MRI”. In: *International Workshop on Graph Learning in Medical Imaging*. Springer. 2019, pp. 70–78.

[244] Michael D Greicius, Kaustubh Supekar, Vinod Menon, and Robert F Dougherty. “Resting-state functional connectivity reflects structural connectivity in the default mode network”. In: *Cerebral cortex* 19.1 (2009), pp. 72–78.

[245] Zhiqun Wang, Jianli Wang, Han Zhang, Robert Mchugh, Xiaoyu Sun, Kuncheng Li, and Qing X Yang. “Interhemispheric functional and structural disconnection in Alzheimer’s disease: a combined resting-state fMRI and DTI study”. In: *PLoS One* 10.5 (2015), e0126310.

[246] Y. Minowa et al. “Comparing Deep Learning Models for Age Prediction Based on the Resting State fMRI Dataset from the" Brain Dock" Service in Japan”. In: (2022).

[247] R Matthew Hutchison et al. “Functional connectivity of the frontal eye fields in humans and macaque monkeys investigated with resting-state fMRI”. In: *Journal of neurophysiology* 107.9 (2012), pp. 2463–2474.

[248] Gonzalo M Rojas et al. “Study of resting-state functional connectivity networks using EEG electrodes position as seed”. In: *Frontiers in neuroscience* 12 (2018), p. 235.

[249] Maria Centeno and David W Carmichael. “Network connectivity in epilepsy: resting state fMRI and EEG–fMRI contributions”. In: *Frontiers in neurology* 5 (2014), p. 93.

[250] Adam Paszke et al. “Pytorch: An imperative style, high-performance deep learning library”. In: *Advances in neural information processing systems* 32 (2019), pp. 8026–8037.

[251] Arpan R Chakraborty et al. “Resting-state functional magnetic resonance imaging with independent component analysis for presurgical seizure onset zone localization: A systematic review and meta-analysis”. In: *Epilepsia* 61.9 (2020), pp. 1958–1968.

[252] James A Roberts, Alistair Perry, Anton R Lord, Gloria Roberts, Philip B Mitchell, Robert E Smith, Fernando Calamante, and Michael Breakspear. “The contribution of geometry to the human connectome”. In: *Neuroimage* 124 (2016), pp. 379–393.

[253] Varina L Boerwinkle et al. “Correlating resting-state functional magnetic resonance imaging connectivity by independent component analysis-based epileptogenic zones with intracranial electroencephalogram localized seizure onset zones and surgical outcomes in prospective pediatric intractable epilepsy study”. In: *Brain connectivity* 7.7 (2017), pp. 424–442.

[254] Francisco Gil et al. “Beyond the epileptic focus: functional epileptic networks in focal epilepsy”. In: *Cerebral Cortex* 30.4 (2020), pp. 2338–2357.

[255] Borbála Hunyadi et al. “ICA extracts epileptic sources from fMRI in EEG-negative patients: a retrospective validation study”. In: *PloS one* 8.11 (2013), e78796.

[256] Elizabeth R DeLong et al. “Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach”. In: *Biometrics* (1988), pp. 837–845.

[257] Bhim M Adhikari et al. “A resting state fMRI analysis pipeline for pooling inference across diverse cohorts: an ENIGMA rs-fMRI protocol”. In: *Brain imaging and behavior* 13.5 (2019), pp. 1453–1467.

[258] Katherine T Martucci et al. “The posterior medial cortex in urologic chronic pelvic pain syndrome: detachment from default mode network. A resting-state study from the MAPP research network”. In: *Pain* 156.9 (2015), p. 1755.

[259] Jewell B Thomas et al. “Functional connectivity in autosomal dominant and late-onset Alzheimer disease”. In: *JAMA neurology* 71.9 (2014), pp. 1111–1122.

[260] Chandra Sripada et al. “Disrupted network architecture of the resting brain in attention-deficit/hyperactivity disorder”. In: *Human brain mapping* 35.9 (2014), pp. 4693–4705.

[261] Madhura Ingalhalikar et al. “Functional connectivity-based prediction of Autism on site harmonized ABIDE dataset”. In: *IEEE Transactions on Biomedical Engineering* (2021).

[262] Jeremy Howard and Sebastian Ruder. “Universal language model fine-tuning for text classification”. In: *arXiv preprint arXiv:1801.06146* (2018).

[263] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. “Ms-celeb-1m: A dataset and benchmark for large-scale face recognition”. In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*. Springer. 2016, pp. 87–102.

[264] Hwanjun Song, Minseok Kim, Dongmin Park, Yooju Shin, and Jae-Gil Lee. “Learning from noisy labels with deep neural networks: A survey”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2022).

[265] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. "Dermatologist-level classification of skin cancer with deep neural networks". In: *nature* 542.7639 (2017), pp. 115–118.

[266] Danna Gurari, Diane Theriault, Mehrnoosh Sameki, Brett Isenberg, Tuan A Pham, Alberto Purwada, Patricia Solski, Matthew Walker, Chentian Zhang, Joyce Y Wong, et al. "How to collect segmentations for biomedical images? A benchmark evaluating the performance of experts, crowdsourced non-experts, and algorithms". In: *2015 IEEE winter conference on applications of computer vision*. IEEE. 2015, pp. 1169–1176.

[267] Curtis P Langlotz, Bibb Allen, Bradley J Erickson, Jayashree Kalpathy-Cramer, Keith Bigelow, Tessa S Cook, Adam E Flanders, Matthew P Lungren, David S Mendelson, Jeffrey D Rudie, et al. "A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/The Academy Workshop". In: *Radiology* 291.3 (2019), pp. 781–791.

[268] Maxwell L Elliott, Annchen R Knodt, David Ireland, Meriwether L Morris, Richie Poulton, Sandhya Ramrakha, Maria L Sison, Terrie E Moffitt, Avshalom Caspi, and Ahmad R Hariri. "What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis". In: *Psychological science* 31.7 (2020), pp. 792–806.

[269] M Tynan R Stevens, David B Clarke, Gerhard Stroink, Steven D Beyea, and Ryan CN D'Arcy. "Improving fMRI reliability in presurgical mapping for brain tumours". In: *Journal of Neurology, Neurosurgery & Psychiatry* 87.3 (2016), pp. 267–274.

[270] Olesya Grinenko, Jian Li, John C Mosher, Irene Z Wang, Juan C Bulacio, Jorge Gonzalez-Martinez, Dileep Nair, Imad Najm, Richard M Leahy, and Patrick Chauvel. "A fingerprint of the epileptogenic zone in human epilepsies". In: *Brain* 141.1 (2018), pp. 117–131.

[271] Lara Jehi. "The epileptogenic zone: concept and definition". In: *Epilepsy currents* 18.1 (2018), pp. 12–16.

[272] Yarin Gal, Jiri Hron, and Alex Kendall. "Concrete dropout". In: *Advances in neural information processing systems* 30 (2017).

[273] Tong Xiao, Tian Xia, Yi Yang, Chang Huang, and Xiaogang Wang. "Learning from massive noisy labeled data for image classification". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 2691–2699.

[274] Shu Kay Ng, Thriyambakam Krishnan, and Geoffrey J McLachlan. "The EM algorithm". In: *Handbook of computational statistics: concepts and methods* (2012), pp. 139–172.

[275] Sarah Parisot, Ben Glocker, Sofia Ira Ktena, Salim Arslan, Markus D Schirmer, and Daniel Rueckert. "A flexible graphical model for multimodal parcellation of the cortex". In: *Neuroimage* 162 (2017), pp. 226–248.

[276] Atif Riaz, Muhammad Asad, SM Masudur Rahman Al Arif, Eduardo Alonso, Danai Dima, Philip Corr, and Greg Slabaugh. "Deep fMRI: An end-to-end deep network for classification of fMRI data". In: *2018 ieee 15th international symposium on biomedical imaging (isbi 2018)*. IEEE. 2018, pp. 1419–1422.