Modelling and Prediction of Neurodevelopment in Preterm Infants using Structural Connectome Data

by

Colin Joseph Brown

B.Math, University of Waterloo, 2009

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
School of Computing Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Spring 2017

Approval

Name:	**Colin Joseph Brown**
Degree:	**Doctor of Philosophy (Computing Science)**
Title:	***Modelling and Prediction of Neurodevelopment in Preterm Infants using Structural Connectome Data***
Examining Committee:	**Chair:** Mark Drew Professor

Ghassan Hamarneh
Senior Supervisor
Professor _____

Greg Mori
Supervisor
Professor _____

Martin Ester
Internal Examiner
Professor
School of Computing Science _____

Maxime Descoteaux
External Examiner
Professor
Centre de Recherche CHUS
Computer Science department
Sherbrooke University _____

Date Defended: April 4th, 2017

Abstract

Each year worldwide, millions of babies are born very preterm (before 32 weeks postmenstral age). Very preterm birth puts infants at higher risk for delayed or altered neurodevelopment. While the mechanisms causing these alterations are not fully understood, it has been shown that image-based biomarkers of the fragile connective white matter brain tissue are correlated with neurodevelopmental outcomes. Diffusion MRI (dMRI) is a non-invasive imaging modality that allows *in-vivo* analysis of an infant's white matter brain network (known as a structural connectome, representing the neuronal wiring of the brain) and can be used to better understand neurodevelopment. The purpose of this thesis is to study how the structural connectome can be used for analysis of development and early prediction of outcomes for better informed care. The thesis begins with a thorough examination of the literature on studies that have applied machine learning to brain network data from MRI. It proceeds with a connectome based analysis of the early neurodevelopment of normative preterm infants. Finally, this thesis tackles the problem of early prediction of cognitive and motor neurodevelopmental outcomes using machine learning on connectome data. Three novel prediction methods are proposed for this task, which are found to be able to accurately predict the 18-month neurodevelopmental outcomes of a cohort of preterm infants from the BC Childrens' Hospital. The thesis concludes with a discussion of how the proposed models may be applicable to a broader set prediction problems and of important future directions for research.

Keywords: Machine Learning, Brain Network, Prediction, Preterm Infants, Connectome Data

Dedication

This thesis is dedicated to my parents who taught me about the beauty and the power of science and to my grandmother, Lucile, who inspired me to go down this path.

Acknowledgements

First and foremost, I'd like to thank my senior supervisor Prof. Ghassan Hamarneh for his guidence, wisdom, attention to detail and passion for doing great research. Without him, I couldn't have done any of this. I would also like to thank my former MIAL labmate, Dr. Brian Booth who helped me immeasurably in my first years and taught me about diffusion MRI, how to write a paper and so much more. Thanks to Jeremy Kawahara for excellent chats about wild ways to process data. As well, thanks to all of the rest of my labmates who all helped me and each other get through grad school. Thanks to Dr. Steven Miller and the whole team of clinical collaborators whose work on the preterm infant dataset gave me a topic for a PhD and who also provided critical knowledge, insight and editing to each of my papers. Thank you to my family for their endless love and support. Finally, thank you Emily, for supporting me and sticking by me the whole way.

Table of Contents

List of Tables

List of Figures

List of Common Mathematical Symbols

Symbol	Definition
N	number of scans in a dataset
M	number of features
$G(E, V)$	graph (i.e., brain network or connectome)
E	set of edges
V	set of vertices
A	adjacency matrix of a connectome
$L(\cdot)$	loss function (i.e., data term)
$R(\cdot)$	regularization function
$f(\cdot)$	learned prediction function
$\mathbf{y}$	vector of instance labels (can be continuous or categorical)
$\hat{\mathbf{y}}$	vector of predicted labels
X	matrix of feature row vectors
Θ	set of model parameters (to be learned)
$\mathbf{w}$	vector of linear model weights
w_{ij}	feature at edge between nodes i and j
λ	hyper-parameter
$\mathcal{L}$	graph Laplacian
$\mathcal{N}(i)$	set of nodes neighbouring node i
d_i	degree of node i
$P_{j,k}$	path (sequence) of edges between nodes j and k
$\mathcal{O}(\cdot)$	big-O notation
$\mathbb{R}$	set of real numbers

Chapter 1

Introduction

1.1 Background and Motivation

The brain is an organ comprising billions of neurons, connected into a network (or graph) called the connectome [203, 259]. The structure and activitiy of the brain network can be imaged non-invasively using diffusion MRI (dMRI) and functional MRI (fMRI), respectively. Use of these technologies has enabled faster and better diagnoses and treatments of neurological disorders (e.g., in dMRI and fMRI guided neurosurgury and in pharmacokinetics) and a deeper understanding of the human brain [23, 178, 225].

Infants born very preterm are one clinical group in particular that has been the focus of many brain imaging studies [158]. The early infant brain develops rapidly and is sensitive to infection and insult, especially within the connective white matter tissue. Very preterm birth (i.e., birth before 32 weeks post-menstral age) has been shown in dMRI studies to affect brain white matter maturation and puts infants at a higher risk for altered or delayed neurodevelopment [182]. Preterm birth is a world-wide health challenge, affecting millions of children every year. Despite decreasing mortality rates for preterm infants due to improving care, the rate of preterm birth is increasing in nearly every country, world-wide (where birth statistics are available) [201].

In groups at risk like preterm infants, analysis of connectome data (i.e., network representations of the connectivity of the brain derived from neuroimaging data) can help unveil both the connective topology of the brain white matter network and how the connections between brain regions change over time. This perspective can enable a better understanding of the aetiology of different conditions and can allow the formulation of new biomarkers for improved diagnosis and care.

Recently, researchers have been exploring the application of machine learning models to connectome data in order to predict clinical outcomes and analyze the importance of subnetworks in the brain. In the case of young infants, given accurate predictions of neurodevelopmental trajectories and outcomes, it may be possible to improve those outcomes

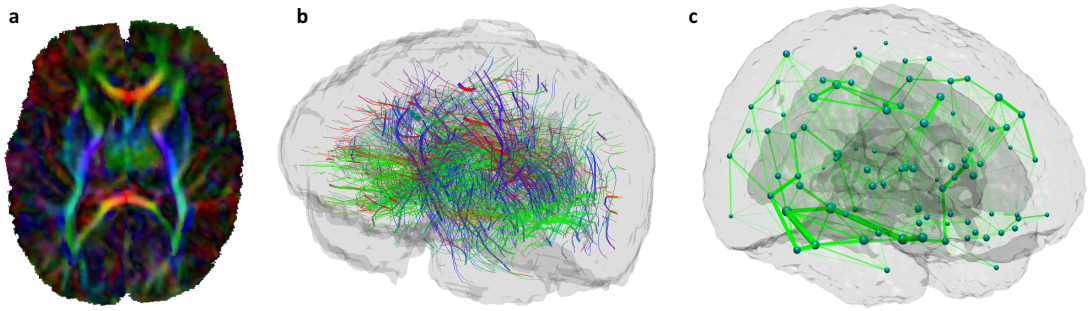

Figure 1.1: Examples of dMRI based data. a) Axial slice of a diffusion tensor image of an infant brain. Colors represent the primary direction of diffusion. b) Selected white matter fibers reconstructed from dMRI via tractography and rendered on brain volume. c) Visualization of a brain network where nodes represent brain region centroids and edge strengths (thickness) represent number of fibers between the pair of regions.

through targeted early interventions [14, 27]. However, connectome data has unique properties that present both special challenges and opportunities when used for machine learning.

With these points in mind, three overarching goals motivate the work in this thesis:

1. To improve the understanding of infant brain structure and development using advanced image analysis techniques.

2. To enable improved care of infants born very preterm by developing methods to predict neurodevelopment outcomes as early as possible.

3. To advance the state-of-the-art of machine learning techniques, especially on network data, by proposing new learning models and data augmentation techniques.

1.2 Thesis Contributions

In this thesis, we first review the literature on the topic of applying machine learning models to MRI-based connectome data (Chapter 2). We then derive connectomes from dMRI images of a cohort of preterm infants and use this data to examine the structure and early development of the preterm-born infant brain (Chapter 3). Fig. 1.1 shows some samples of dMRI data after different stages of processing. Finally, we propose four novel connectome-based machine learning techniques aimed at predicting the neurodevelopmental outcomes of those infants, assessed 18 months after birth (Chapters 4-6). Throughout these latter chapters, the driving motivation is to enable earlier detection of abnormality and ultimately better informed care. The thesis concludes by discussing how the proposed models may be applicable to othe types of network data and by identifying important directions for future research on preterm infant connectome analysis, modelling and prediction (Chapter 7). Here, we present a brief summary of each of the research contributions in this thesis.

1.2.1 Machine Learning on Human Connectome Data from MRI

The field of machine learning on connectome data is growing rapidly and now encompasses a large body of research (Fig. 2.1). In Chapter 2, we survey the literature on the topic of applying machine learning models to MRI-based connectome data. To summarize the research done to date, we provide a comparative, structured summary of 83 relevant works, tabulated according to different criteria (e.g., modality, disease, learning model) that represent a cross-section of the literature on this topic (Table 2.1). After giving an overview of how connectomes are constructed from dMRI and fMRI data, we discuss the variety of machine learning tasks that have been explored with connectome data. We then compare the advantages and drawbacks of different machine learning approaches that have been employed, discussing different feature selection and feature extraction schemes, as well as the learning models and regularization penalties themselves. Throughout this discussion, we focus particularly on how the methods are adapted to the unique graphical nature of connectome data. Finally, we conclude by summarizing the current state of the art and by outlining what we believe are strategic directions for future research.

Chapter 2 is adapted from our 2016 survey paper posted to arXiv [43],

- C J Brown, G Hamarneh, *"Machine Learning on Human Connectome Data from MRI"*, arXiv preprint arXiv:1611.08699 (2016).

This work has been submitted to the journal Medical Image Analysis for publication. Additionally, the structured summary of papers in this survey has been posted online at http://connectomelearning.cs.sfu.ca/, in a living table that the community can contribute to as the field continues to grow.

1.2.2 Structural Network Analysis of Brain Development in Young Preterm Neonates

As was mentioned above, preterm infants develop differently than those born at term and are at higher risk of brain pathology. Thus, an understanding of their development is of particular importance. dMRI of preterm infants' brains offers a window into neurodevelopment at a very early age, an age at which that development is not yet fully understood.

In Chapter 3, we examine a cohort of 47 normal preterm neonates (i.e., without brain injury and with normal neurodevelopmental outcomes at 18 months of age) scanned between 27 and 45 weeks post-menstrual age (PMA). The goal of this study is to further the understanding of how the early structural connectome develops, especially in the case of preterm birth. We use full-brain tractography to reconstruct white matter tracts between the 90 cortical and sub-cortical regions defined in the University of North Carolina Chapel Hill neonatal atlas [247]. We then analyze the resulting connectomes and explore the differences

between measuring edges connectivity by tract count or by mean fractional anisotropy. We also create a linear mixed effects model of the connectomes over time to observe how the structure of the brain develops. Of specific interest was to discover if the topology of the preterm infant brains exhibit small-world properties, the hallmarks of well organized and efficient communication networks. Such properties have been previously observed in the brain networks of adults and infants born at term.

We observed that the brain networks in preterm infants, much like infants born at term, show high efficiency and clustering measures across a range of network scales. Further, the development of many individual region-pair connections, particularly in the frontal and occipital lobes, is significantly correlated with age. Finally, we observe that the preterm infant connectome remains highly efficient yet becomes more clustered across this age range, indicating increasing small-worldness with developmental age (Fig. 3.6).

Chapter 3 is adapted from our 2014 paper published in NeuroImage (Impact Factor: 6.132, ranked among top 100 scientific journals) [44],

- C J Brown, S P Miller, B G Booth, S Andrews, V Chau, K J Poskitt, G Hamarneh, *"Structural Network Analysis of Brain Development in Young Preterm Neonates"*, NeuroImage, 101:667-680 (2014), Elsevier.

1.2.3 Prediction of Motor Function in Very Preterm Infants using Connectome Features and Local Synthetic Instances

Abnormal neuromotor developmental outcomes, including conditions of severe motor function deficit like cerebral palsy, is one kind of adverse outcome that infants born preterm are at particularly high risk for [325]. However, early indication of abnormal neuromotor development could better inform care givers and potentially enable improved, targeted interventions.

In Chapter 4, we propose a method to identify preterm infants at highest risk of adverse neuromotor developmental outcomes (identified at around 18 months of age) using connectome features from dMRIs acquired shortly after birth (Fig. 1.2). For each full-brain dMRI, a connectome is constructed and high-level network features are extracted that characterize the topology of the brain network. After reducing the dimensionality of the feature vector via principal component analysis (PCA), a support vector machine (SVM) is used to predict a normal or an abnormally low expected motor function outcome for each infant. This work is the first image analysis approach for predicting impairment of motor function in preterm-born infants. Additionally, in order to mitigate the effects of class imbalance, a novel method is proposed to produce realistic synthetic training data for a particular class. Our method is validated on a dataset of 168 dMRIs of 115 very preterm infants (which includes those 47 normal infants examined in Chapter 3), all scanned between 27 and 45 weeks PMA with neuromotor development outcomes assessed using the Bayley scales of

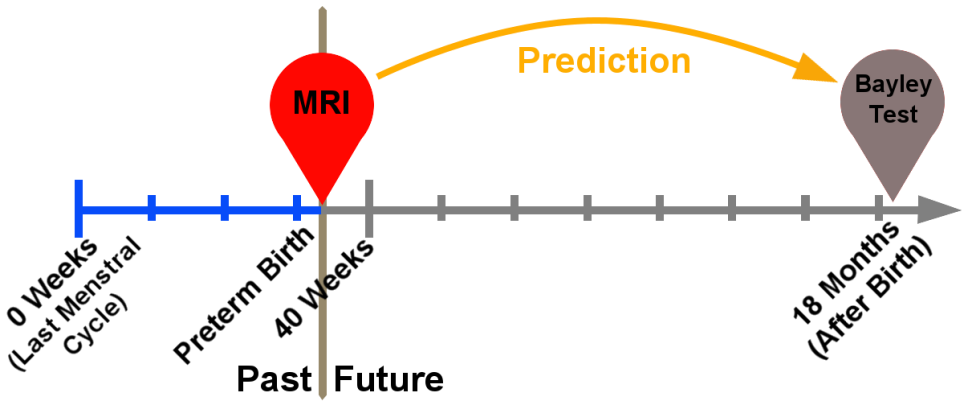

Figure 1.2: Timeline showing the time of the MRI scan, shortly after preterm birth, and the time of the outcome assessment. We developed machine learning models to predict the outcome scores using only information at the time of the scan.

infant and toddler development III (Bayley-III). We show that using synthesized training data can consistently improve classification accuracy while setting a baseline performance for this challenging prediction problem.

Chapter 4 is adapted from a paper published in the proceedings of the 2015 international conference on medical image computing and computer assisted intervention (MICCAI) (30% acceptance rate) [45],

- Colin J Brown, Steven P Miller, Brian G Booth, Kenneth J Poskitt, Vann Chau, Anne R Synnes, Jill G Zwicker, Ruth E Grunau, Ghassan Hamarneh, *"Prediction of Motor Function in Very Preterm Infants using Connectome Features and Local Synthetic Instances"*, International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 69-67, Springer International Publishing, 2015.

1.2.4 Predictive Subnetwork Extraction with Structural Priors for Infant Connectomes

While prediction of outcomes can help to inform the risks of abnormal neurodevelopment, it does not elucidate which of the many connections or regions in the brain are affected by such abnormal development or are most important for predicting the outcome. Furthermore, from prior knowledge about the structure of the brain, it is expected that certain sets of connections (e.g., subnetworks) should be more likely to accurately predictic outcomes than others.

Accordingly, in Chapter 5 we present a new method to identify anatomical subnetworks of the human white matter connectome that are predictive of neurodevelopmental outcomes. We employ our method on the same dataset of 168 preterm infant connectomes, generated from dMRIs acquired shortly after birth, to discover subnetworks that predict neuromotor

development scores assesed at 18 months after birth. Additionally, our approach is able perform the more challenging task of predicting cognitive neurodevelopmental outcome scores for the first time.

Predictive subnetworks are extracted via sparse linear regression with weights on each connectome edge. By enforcing novel backbone network and connectivity based priors, along with a non-negativity constraint, the learned subnetworks are simultaneously anatomically plausible, well connected, positively weighted and reasonably sparse. Compared to other state-of-the-art subnetwork extraction methods, we found that our approach extracts subnetworks that are more integrated, have fewer noisy edges and that are also better predictive of neurodevelopmental outcomes.

Chapter 5 is adapted from a paper published in the proceedings of the MICCAI 2016 (30% acceptance rate), where it was presented as an oral presentation (5% acceptance rate) [46],

- Colin J Brown, Steven P Miller, Brian G Booth, Jill G Zwicker, Ruth E Grunau, Anne R Synnes, Vann Chau, Ghassan Hamarneh, *"Predictive Subnetwork Extraction with Structural Priors for Infant Connectomes"*, International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pages 175-183, Springer International Publishing, 2016.

Additionally, an extended version of this work has now been submitted to a special, invitational edition of Medical Image Analysis that comprises top MICCAI papers of the previous year.

1.2.5 BrainNetCNN: Convolutional Neural Networks for Brain Networks

Given the complexity of the brain, its rapid development during infancy and the multitude of ways that different etiologies might impact neurodevelopment, we expect that the relationship between the connectome after birth and the neurodevelopmental outcome scores at 18 months to be complex and likely highly non-linear. Thus, while linear prediction models performed well, it is reasonable to assume that a stacked, non-linear model might perform as well or better, espcially given the success of models such as convolutional neural networks (CNN) for similar prediction tasks on other kinds of medical images [41, 69, 232]. CNNs are a good choice of stacked predictive model for cases where data is limited since a limited set of learned parameters can be reused in convolutional filters that are applied repeatedly across an image. However, traditional CNNs are only designed for standard grid-like images and are not trivially applicable to brain network data.

Thus, in Chapter 6, we propose BrainNetCNN, a CNN framework designed specifically to predict clinical neurodevelopmental outcomes from brain networks. In contrast to the *spatially* local convolutions done in traditional image-based CNNs, our BrainNetCNN is

composed of novel edge-to-edge, edge-to-node and node-to-graph convolutional filters that leverage the *topological locality* of structural brain networks (Fig. 6.1).

We first demonstrate the predictive capabilities of BrainNetCNN on synthetic phantom networks with simulated injury patterns and added noise. BrainNetCNN outperforms a fully connected neural-network with the same number of model parameters on both phantoms with focal and diffuse injury patterns. We then apply this model method to the task of joint prediction of Bayley-III cognitive and motor scores, using structural brain networks of the same set of preterm born infants. We show that our BrainNetCNN framework outperforms a variety of other methods on the same data. Furthermore, BrainNetCNN is able to identify an infant's post menstrual age to within about 2 weeks. Finally, we explore the high-level features learned by BrainNetCNN by visualizing the importance of each connection in the brain with respect to predicting the outcome scores. These findings are then discussed in the context of the anatomy and function of the developing preterm infant brain.

Chapter 6 is adapted from a 2016 paper published in NeuroImage [149],

- Colin J Brown*, Jeremy Kawahara*, Steven P Miller, Brian G Booth, Vann Chau, Ruth E Grunau, Jill G Zwicker, Ghassan Hamarneh, *"BrainNetCNN: Convolutional Neural Networks for Brain Networks; Towards Predicting Neurodevelopment"*, NeuroImage, (In press). *Co-first authors

Note that this work was co-first authored by myself and Jeremy Kawahara. The idea and formulation of the convolutional filters were proposed and developed together along with our supervisor, Dr. Ghassan Hamarneh. Jeremy performed the majority of the experiments while I wrote the majority of the manuscript, although we each wrote some code, performed some experiments, and wrote some sections of the text. Thus, it was a strong collaborative effort that cannot be attributed more to either Jeremy or myself.

1.2.6 Additional Contributions Omitted From Thesis

In addition to the publications listed above that have been included as chapters, a number of related works have been completed during work on this thesis that have been omitted to maintain a clear focus. These works are listed below in reverse chronological order, but with first authorship papers before co-authored papers:

- C J Brown, B G Booth, G Hamarneh, *"Uncertainty in Tractography via Tract Confidence Regions"*, Springer Mathematics and Visualization, 129-138 (2014), Springer, Computational Diffusion MRI Workshop at MICCAI 2013, Japan, Nagoya.

- C J Brown, B G Booth, G Hamarneh, *"K-Confidence: Assessing Uncertainty in Tractography Using K Optimal Paths"*, ISBI 2013, 250 - 253 (2013), IEEE Xplore, United States, California IEEE International Symposium on Biomedical Imaging (IEEE ISBI), San Francisco, CA.

- R Zakariaee, G Hamarneh, C J Brown, M Gaudet, C Aquino-Parsons, I Spadinger, *"Bladder accumulated dose in image-guided high-dose-rate brachytherapy for locally advanced cervical cancer and its relation to urinary toxicity"*, Physics in Medicine and Biology, 2016, (In Press).

- R Zakariaee, G Hamarneh, C J Brown , I. Spadinger, *"Validation of non-rigid point-set registration methods using a porcine bladder pelvic phantom"*, Physics in Medicine and Biology 61:2 (2016).

- B G Booth, S P Miller, C J Brown, K J Poskitt, V Chau, R E Grunau, A R Synnes, G Hamarneh, *"STEAM - Statistical Template Estimation for Abnormality Mapping: A personalized DTI analysis technique with applications to the screening of preterm infants"*, NeuroImage 125, 705-723 (2016).

- J Smith-Forrester, N Rotem-Kohavi, C J Brown, G Hamarneh, and N Virji-Babul, *"Changes in Structural Brain Connectivity Following Concussion"*, In The 9th Annual Canadian Neuroscience Meeting (CAN-ACN), Symposium: Vision and Movement Order and Disorder: From Bench to Bedside, 2015.

- R Zakariaee, C J Brown, G Hamarneh, C A Parsons, I Spadinger, *"Estimation of Bladder-Wall Cumulative Dose in Multi-Fraction Image-Based Gynaecological Brachytherapy Using Deformable Point Set Registration"*, Medical Physics, 41, 8, 3-3 (2014), AIP, COMP Annual Scientific Meeting 2014, Banff, AB.

- R Zakariaee, M Gaudet, R Dubash, C Parsons, C J Brown, I Spadinger, *"A Method to Estimate the Spatial Distribution of Bladder Hot Spots in Image-Guided Intracavitary Brachytherapy of Cervical Cancer"*, Brachytherapy, Vol 13, Supplement 1, S88, Elsevier, 2014 ABS Annual Meeting, San Diego, CA.

- A Top, C J Brown , M R Rezaee, G Hamarneh. *"Method And Apparatus For Detecting Anatomical Elements"*. Patent Pending (2013). Docket No. 050704/432915.

1.3 Thesis Outline

Together, the work in this thesis contributes to the fields of image analysis, machine learning, neuroscience and the study of the care of preterm infants. This thesis:

1. Presents the first connectome analysis of infants as young as 27 weeks PMA,

2. Presents the first work to look at whole-brain network integration and segregation in a large, normative cohort of preterm infants,

3. Discovers small-worldness in preterm infant brain networks increasing with age,

4. Proposes the local synthetic instances (LSI) method for generating realistic synthetic data in a single class,

5. Presents the first image analysis approach to predicting impairment in motor function in preterm-born infants,

6. Presents the first work to predict cognitive neurodevelopmental outcomes using the structural connectome,

7. Proposes two novel structural subnetwork priors for learning models over brain network data,

8. Proposes BrainNetCNN, the first CNN designed to be applied to brain network data with novel layer types that extend convolutions to the domain of networks.

It is the hope of the author that this work has a positive impact on the care for infants born preterm, the understanding of the brain and the field of machine learning.

Chapter 2

Machine Learning on Human Connectome Data from MRI

2.1 Introduction

The human brain is a complex network of neurons [259]. Advances in imaging technologies such as functional MRI (fMRI) and diffusion MRI (dMRI) that allow non-invasive, *in-vivo* analysis of a patient's brain network (known as a connectome), along with increased access to these technologies by researchers and clinicians, have substantially enhanced our understanding of the human brain, and are beginning to contribute to improved diagnoses and treatments of neurological disorders (e.g., for pre-surgical applications) [21, 123, 196, 178].

As was mentioned in Chapter 1, a connectome is a graphical representation of structural or functional connections in the brain, inferred from imaging data (e.g., fMRI or dMRI). Deriving a connectome from an image requires a pipeline of processing steps, which from a high level involves parcellating the brain into localized regions and then computing the connectivity between each pair of regions. While much work has been done to determine the best approaches for inferring connectome data from imaging data (Section 2.2), it is far from a solved problem [22]. Despite this ongoing exploration of how best to construct a connectome, researchers have been starting to apply the tools of machine learning to connectome data in order to perform tasks such as prediction of clinical outcomes and analysis of subnetworks in the brain. In this chapter, we present a review of machine learning techniques that have been used for connectome data. We first outline the unique characteristics of connectome data (Section 2.1.1) and then outline the purpose, scope and layout of the survey presented in this chapter (Section 2.1.2).

2.1.1 Special Challenges and Opportunities of Connectome Data

Connectome data is unique in a variety of ways which both pose unique challenges and provide unique opportunities when used in the context of machine learning. Whereas standard medical images are grid-like with each pixel neighbouring only other pixels that are spatially near-by, connectomes have a more general topology. Each brain region (represented as a node in the network) may be connected, either structurally and/or functionally, to any other brain region (with each connection between two regions respresented as a binary or weighted edge in the network). This structure conveys intrinsic information about the connectivity of the brain that is not explicitly represented in other modalities or image formats. This topological information has been found to be useful as a discriminative biomarker of a variety of neurological conditions [18, 21, 120, 131, 197, 235, 263]. Nevertheless, many types of features that have successfully been applied to extract information from grid-like image data (e.g., scale invariant feature transform features [175], histogram of oriented gradient features [75] and learned convolutional neural network (CNN) features [155]) are inapplicable or would need to be modified for use on network structured data. The adoption of connectome data has required new kinds of features to be explored (Section 2.4.2).

Another aspect of connectome data that differs from grid-like image data is that, in most cases, connectome nodes have intrinsic correspondence between subjects (and between multiple scans of the same subject). Thus, once the connectomes are constructed, each edge and node has a particular biological interpretation and comparison between nodes or edges of multiple connectomes is trivial. In contrast, the biological interpretation of individual voxels in an image depend completely on the frame of reference and two images must undergo image registration in order to establish correspondence. Nonetheless, constructing a connectome from raw image data is typically a multi-stage pipeline (Section 2.2), which often includes image registration and other processes that can introduce error and noise at each step [255, 318].

The long acquisition time and high cost of acquiring dMRI and fMRI scans means that many studies are performed and validated over relatively few scans. For instance, examining the studies listed in Table 2.1 shows that, while there is a trend of connectome datasets becoming larger (see Fig. 2.5), the median number of scans used in each study is only $N = 69$. Though connectomes are typically lower dimensional than grid-like images (e.g., thousands of connections versus millions of voxels), this dimensionality is nearly always larger than the number of scans. Of the papers covered in this review, the median number of basic edge features in each connectome was $M = 2850$, giving rise to the cannonical problem of $N \ll M$ [190], known as the high dimensional small sample size (HDSSS) problem. In order to prevent model learning from being an ill-posed problem [30], feature selection, data augmentation, dimensionality reduction (Section 2.4.3) and model regularization (Sec-

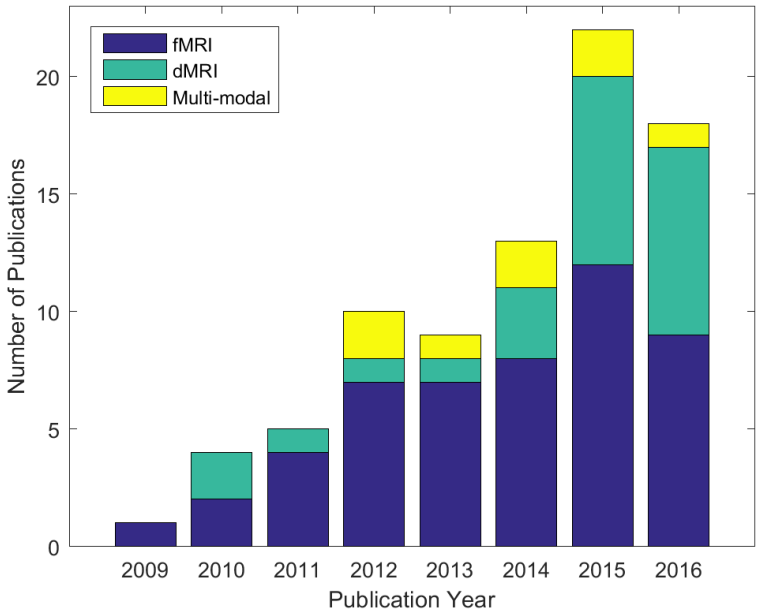

Figure 2.1: Number of studies (found by this survey) on the topic of machine learning for MRI-based connectome data versus year of publication. Colors represent the number of fMRI (blue), dMRI (green) and mutli-modality (yellow) studies by year. Note that for papers which studied multiple datasets independently, each dataset is counted separately in this chart.

tion 2.4.4) are often essential components of a machine learning pipeline for connectome data.

Another unique aspect of machine learning on connectome data is the biology of the brain itself: The brain is a highly complex organ comprising billions of neurons [203] and the structural arrangement of these neurons changes greatly over a persons lifetime [70]. Also, the connectivity of the brain is altered by learning and expierences [82], as well as injury and pathology, so the exact structure of each person's brain is inherently unique, at least on fine scales. Thus, modeling variation in connectivity of the brain over a population may be especially challenging, perhaps requiring flexible and highly non-linear models [184]. Furthermore, ground truth class labels for images of the brain often rely on measurements or tests of ability which may introduce subjective bias from the rater. Thus, connectome data often have noisy labels as well as noisy features.

2.1.2 Purpose, Related Works and Scope

The purpose of this chapter is to provide a thorough review of studies that have applied machine learning models to MRI-based connectome data. To do this, we have created a table of 83 works (Table 2.1) that represent the majority of the literature on this topic. The field is growing rapidly and now encompasses a large body of research. Fig. 2.1 shows the number of studies on machine learning with connectome data published per year (as listed

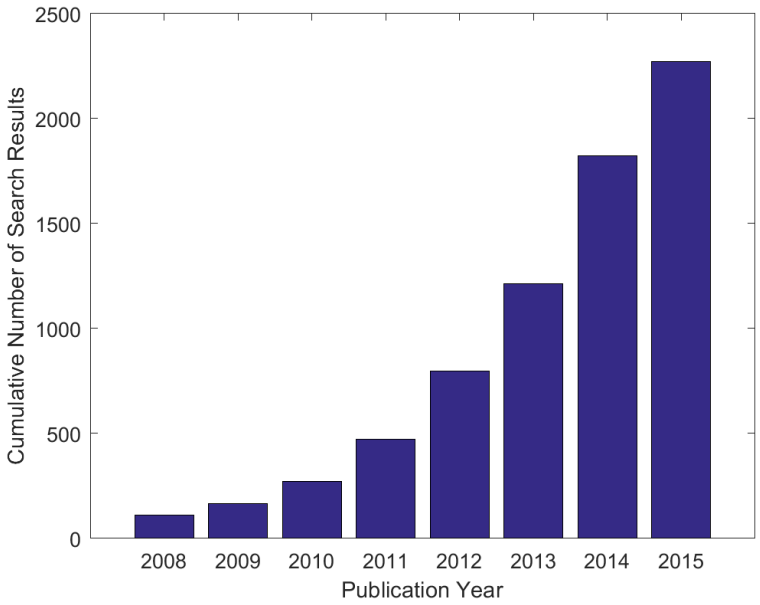

Figure 2.2: Number of search results found by Google Scholar for search pattern "machine learning" + (connectome OR "brain network") + (fMRI OR dMRI OR DTI OR "diffusion MRI" OR "functional MRI" OR "diffusion tensor") versus maximum publication year. Note that the results are cumulative and that these searches likely greatly overestimate the actual number of publications on the topic but do suggest a relative increase in scientific interest on the topic over time.

in Table 2.1). No papers found were published prior to 2009 but the average number of publications has increased nearly every year since.

Performing a search using Google Scholar[1] with the search pattern "machine learning" + (connectome OR "brain network") + (fMRI OR dMRI OR DTI OR "diffusion MRI" OR "functional MRI" OR "diffusion tensor") returns 2270 results but, when restricted to only return results from before 2009, the same search only returns 110 results (as of September, 2016). Fig. 2.2 shows this trend over time. While the results of these searches likely overestimate the true number of studies on this topic (e.g., due to finding the keywords out of context or in references to other works), it suggests that the relative popularity of studying machine learning on connectome data is increasing. Thus, there is a present need for surveys on machine learning with connectome data that summarize the methods and results of the research that has been done in the past decade or so.

To date, few surveys on applying machine learning to connectome data have been written. Richiardi et al. wrote a survey on this topic in 2013 [225] however it only covered machine learning on fMRI based connectome data. Furthermore, since this paper was published a great deal of work has been done (i.e., Fig. 2.1 shows that more than half of the papers listed in the present review were published after 2013). In 2014, Varoquaux and Thirion published a review that was also focused on functional connectome data (including

[1] `scholar.google.com`

EEG) but did not perform an in-depth comparitive analysis of relevant works as we do here [286]. Mwangi et al.'s review (2014) examined papers across all neuroimaging modalities but focused only on feature selection and dimensionality reduction techniques [190]. Bassett and Lynall's survey (2013) covered a variety of network based approaches for analysis of the brain but didn't focus specifically on machine learning [23].

The scope of the papers included in Table 2.1 is limited specifically to those that applied machine learning to MRI derived connectome data. While many studies have performed statistical analsysis studies (e.g., group difference studies) on functional and structural connectome data (e.g., [21, 44, 235, 270]), papers were only included in this review if they employed supervised or unsupervised machine learning models. In the case of supervised learning, we only examined studies where the models were trained and tested on distinct datasets or studies that used a cross-validation scheme. Because the focus of this review is on connectome data from MRI images, papers that primarily examined electro-encephalography (EEG) data (e.g., Micheloyannis et al. [181]), magneto-encephalography (MEG) data (e.g., Plis et al. [214]) or metabolic brain networks from FDG-PET data (e.g., Zhou et al. [319]) were not included. Similarly, because the focus here is on either structural or functional connectivity, this review does not include studies which primarily examine relationships between cortical thicknesses in different regions (e.g., Cuingnet et al. [73], Liu et al. [171]). Finally, papers that used machine learning to construct connectomes only and did not perform any further learning tasks on those connectomes (e.g., Jiang et al. [137]) were not included in Table 2.1.

The remainder of this chapter is organized as follows: We first breifly discuss how connectomes are constructed from fMRI or dMRI data (Section 2.2) before presenting a thorough survey of different machine learning tasks that have been explored (Section 2.3), different machine learning models that have been proposed or employed (Section 2.4) and finally a summary of the current state of the art and a discussion about what we believe are strategic directions for future research (Section 2.5).

2.2 Connectome Construction

Constructing a connectome from a dMRI or fMRI scan requires a pipeline of processing steps [51]. Here, we only briefly describe the MRI acquisition process (Sections 2.2.1 and 2.2.2), as it has been discussed in great detail in previous works [36, 187, 279]. Also, since the focus of this survey is on the methods that take connectomes as input, we only outline the most common methodologies for connectome construction and we don't discuss the details of each step. However, it should be noted that changing parameter settings or using different methods at each pipeline stage has been shown to have a significant impact on results [50, 71, 169, 220, 308, 318]. For in-depth explorations of how different construction methods affect resulting functional connectomes see Liang et al. [169] and for structural connectomes see Zhong et al. [318], Qi et al. [220] or Buchanan et al. [50]. Furthermore, Yao

et al. examined specifically how atlas choice can affect resulting structural and functional connectomes [308].

Formally, a connectome is a graph, $G(E, V)$, representing the structural or functional connectivity between pairs of brain regions of interest (ROI) or landmarks represented as a set of nodes, V. The connectivity between pairs of nodes is represented as the set of edges, E, which may be binary or weighted (and either directed or undirected). Each connectome is typically encoded as an adjacency matrix, $A \in \mathbb{R}^{|V| \times |V|}$, in which each entry represents either the existence of an edge (in a binary network) or the weight of an edge (in a weighted network). In certain cases, each node may also be assigned a weight (e.g., Ng et al. [194] used fMRI signals to define the weight at each node and dMRI connecitivity to define the weight of edges between nodes).

The number of nodes, $|V|$, defines the scale of the brain network with large scale networks being defined by a few nodes from large ROIs (or spatially sparse landmarks) and fine scale networks being defined by many nodes from small ROIs (or densely distributed landmarks) [3]. In some extreme cases, each voxel in the image is associated with a node [7]. Some works have also constructed connectome from multiple scales [142] or instead constructed a fine scale connectome and then employed multi-scale connectome features [151, 152].

A common way to define ROIs is to register (i.e., align, either linearly or non-linearly) each MRI brain scan to a template image with an associated atlas of segmented brain regions that can be used for parcellation. Many machine learning studies on both functional and structural connectomes used atlas based approaches to identify ROIs. In Table 2.1, of the 83 listed papers, 54 used atlases, such as the automated anatomical labeling (AAL) atlas [277], versus only 6 that used landmarks, such as the dense individualized common connectivity-based cortical landmarks (DICCCOL) [320]. The main advantage of using a labeled atlas of the brain to define ROIs is that it facilitates an objective anatomical interpretation of learned models which can easily be compared to other results in the literature. However, it has been noted in the literature that atlas based approaches are susceptable to imperfect segmentations of functional regions which can considerably impact sensitivity [255]. DICCCOL uses the white matter structure of the brain in attempt to ensure the correct placement of each landmark but only defines a single position, rather than a region, for every landmark which may limit the number of voxels that can be reliably associated with each connectome node.

Fifteen papers, the majority of which focused on fMRI, used independent component analyis (ICA) [162] or other unsupervised learning techniques such as Ward hierarchial clustering [194, 294] in order to identify regions with self-similar features. The remaining papers either assigned nodes to hand-delineated ROIs, ROIs from uniform tile-like parcellations of the brain or to each voxel. One drawback of these latter approaches (including the learning approaches), is that the nodes are not *apriori* associated with anatomically defined regions

and so interpretation of results may be more difficult. Furthermore, when these kinds of approaches are applied independently to each scan, there may not be correspondence between ROIs in different scans [269]. This will also be the case when comparing connectomes constructed with different pipelines. However, there exist methods, such as graph kernels (Section 2.4.4), that can compare connectomes and be used to build learning models even when correspondence between nodes is not defined.

Edge definitions vary by modality so in the next two subsections we discuss how connectivity is respectively defined for dMRI and fMRI based connectomes.

2.2.1 Structural Connectomes

Structural connectomes are derived from dMRI, which can non-invasively measure water diffusion in living brain tissue (and other tissues). Briefly, to acquire a dMRI of the brain, the patient's head is positioned inside a very strong (e.g., typically 1.5 to 7 tesla), tube-shaped magnet [241]. The strong, uniform magnetic field has the effect of aligning the spin of some proportion of the hydrogen atoms in the water molecules in the tissue. A radio frequency (RF) pulse is then applied, causing the spin of the hydrogen atoms to precess around the initial axis. Then after a specified period of delay, a second identical but 180°-phase-shifted RF pulse is applied. The initial precession is predominantly canceled but this cancelling effect is ineffective where diffusion has occurred, causing a measurable magnetic signal as the precession of the atoms in the diffusing molecules decay. Diffusion can then be measured in specific directions by inducing a magnetic gradient using a gradient coil [19].

Water diffuses more freely along fiberous tissue (like the connective white matter tissue in the brain) compared to diffusion across more uniformly structured tissue (or to diffusion across the fiberous tissue, but perpendicular to the fibers). Thus, by measuring diffusion in different directions at each voxel of the brain image, the structure of the white matter can be inferred (though there remains debate about exactly how to perform this inference) [13, 36, 144].

Edges between each pair of nodes in a structural connectome are assigned (or are assumed to exist and are weighted) by measuring some degree of white matter connectivity between the associated pair of ROIs. This is typically done by first fitting a diffusion model, such as the diffusion tensor imaging (DTI) model to each voxel, and then reconstructing the white matter fibers via tractography [36]. The degree of connectivity between a pair of ROIs is then commonly defined as the number of reconstructed tracts with end-points in (or which simply pass through) both ROIs. In the case of proabilistic tractography, the degree of connectivity may represent the probability of a given tract connecting two ROIs [98, 209]. Alternatively, connectivity can be defined as the mean diffusion fractional anisotropy (FA) of each voxel intersecting the tract and then averaged over all tracts between the two ROIs [184, 228]. Rather than choosing only one definition, some studies analyse features from multiple connectomes, constructed with different definitions of struc-

tural connectivity [45, 299]. Finally, while tractography is the most popular way to define structural connectivity, other approaches have been explored including connectivity based on the fast marching algorithm over the set of voxels with diffusion defining edges between neighbouring voxels [35, 217]. For instance, Prasad et al. examined a measure of maximum flow between pairs of ROIs across a lattice defined by the diffusion tensor at each voxel [217]. Because dMRI does not measure directionality of diffusion flow, edges of structural connectomes are necessarily undirected [36]. For further discussion on constructing structural connectomes, see Hagmann et al. [118] and de Reus and Heuvel [77].

2.2.2 Functional Connectomes

Rather than measuring diffusion, fMRI data measures a blood-oxygen-level dependent (BOLD) signal that is indicative of neural activity over time, at each voxel. This signal is typically averaged within each ROI. Functional connectivity between a pair of ROIs is computed using some measure of correlation or dependence between the two averaged signals. Functional connectivity can then be interpreted as communication between pairs of brain regions. The most standard measure of correlation is the Pearson's correlation coefficient. However, in the case of $N \ll M$, Pearson's correlation overestimates the number of correlated pairs of regions [310], especially when the data is noisy as is the case with fMRI [265]. Gellerup et al. performed a comparison of twelve different measures of connectivity and found Pearson's correlation with Bonferroni multiple comparison correction, which reduces the number of false positives by imposing a very strict significance threshold, worked best [105]. Instead, some studies defined the weight of each edge as the partial correlation between two ROIs. To compute the partial correlation between two BOLD time signals, the effect of every other time signal is first removed (via regression) and then Pearson's correlation is computed. Rosa et al. showed that using partial correlation based functional connectomes lead to higher prediction accuracy [231]. Another approach is to estimate a sparse inverse covariance matrix for the set of ROI time signals. One commonly used method of this type is called the graphical least absolute shrinkage and selection operator (GLASSO) which uses ℓ_1 regularization to promote a sparse solution [96]. Smith et al. and Rosa et al. found that functional connectome construction methods based on sparse estimation of covariance work best in terms of accurate estimation of true connections and in terms of classification accuracy, respectively [231, 255].

After computing the correlations between ROIs, many methods (e.g., [88, 141]) apply Fischer's r-to-z transform [251] which converts correlation values into Z-scores in order to prevent bias from being introduced in subsequent steps. Furthermore, since correlations (and Z-scores) can be negative but certain feature descriptors may assume non-negative inputs, some approaches applied an absolute value function to correlation values, but extracted features from edges with positive and negative correlations separately [6].

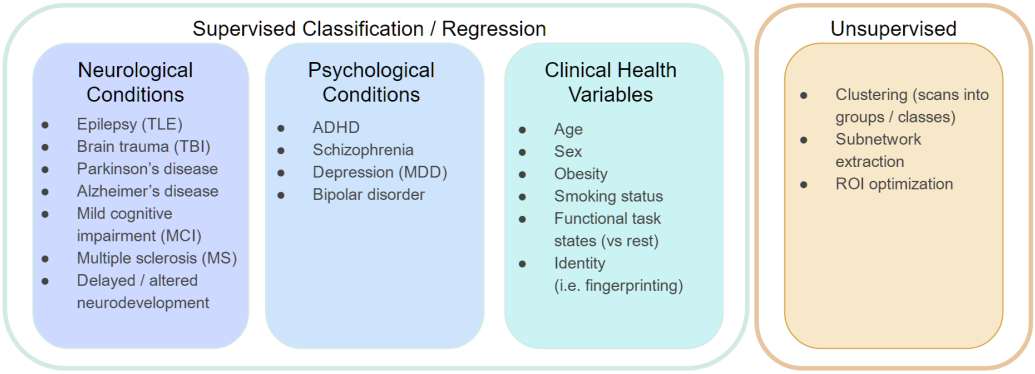

Figure 2.3: Machine learning tasks that have been performed using MRI based connectome data.

Edges in functional connectomes are typically undirected as they are in structural connectoms. However, because fMRI has a temporal component, it is possible (but less common) to construct functional connectomes with directed edges [242, 297]. These connectomes represent what is known as effective connectivity. Both Shahnazian et al. [242] and Wee et al. [297] trained learning models on directed, Granger causality networks [113]. To infer this causality, Shahnazian et al. used conditional Granger causality analysis (CGCA) whereas Wee et al. used multivariate autoregressive (MAR) modelling[242, 297]. Smith et al. found that using Granger causality and other temporal lag based methods performed poorly when trying to infer directional structure, whereas Patel's τ measure [210] (which examines the imbalance between conditional probablities of each direction of connectivity), performed best [255]. Effective connectivity networks better model the true dynamics of the brain [99] and have been shown to enable higher classification accuracy [297], but come with the drawback of requiring twice as many edge weights to represent them.

2.2.3 Signal Noise and Bias

There exist many sources of noise and bias in the construction process for both structural and functional connectomes. Inaccurate placement of ROIs [255] and patient motion during scan, due to the long scan times required for both dMRI and fMRI [216, 229, 282], can affect connectomes of both modalities. Though, whereas structural connectomes tend to contain more false negative connections, due to tractography algorithms terminating too early in crossing and heavily curved white-matter regions, functional connectomes tend to contain more false positive connections, in part due to patient breathing and pulse (which affect blood oxygen levels across the brain) [310]. Never-the-less, many tractography algorithms also produce false positives [176].

2.3 Machine Learning Tasks

The majority of studies that apply machine learning to connectome data have focused on predicting outcomes via classification or regression (71 of 83 of the papers in Table 2.1). However, other works have explored applying machine learning to unsupervised clustering of patient connectomes into groups, identification of important subnetworks and optimization of ROIs.

2.3.1 Outcome Prediction

Table 2.1 reveals that features of structural and functional connectome data can be used to identify or predict a wide variety of pyschological disorders, neurological disorders, patient attributes and mental states (see Fig. 2.3). For instance, connectome features have been used to identify patients with schizophrenia [10, 24, 79, 244, 322], major depressive disorder (MDD) [72, 116, 230, 231, 238], attention deficit hyperactivity disorder (ADHD) [65] and bipolar disorder [186]. Clinical diagnoses of these psychological disorders can be challenging and accurate connectome based identification could allow earlier diagnosis and treatment [10]. While the majority of these works used functional connectomes to train their models, structural connectomes were used by Sacchet et al. [238] and Moyer et al. [186] for identifying MDD and Bipolar disorder, resepectively.

The use of machine learning on connectome data for early diagnosis could especially be beneficial for neurodevelopmental disorders such as austism spectrum disorder (ASD) [142, 166, 300] and delayed motor and cognitive development in preterm infants [45, 46, 149] since early intervention of these conditions can improve patient outcomes [27, 300]. The use of connectome data and machine learning models for identification of neurodegenerative diseases such as Parkinson's disease [100, 105], Alzheimer's disease (AD) [12, 61, 151, 152, 186, 217, 218, 284, 297, 305, 315] and early and late mild cognitive impariment (MCI) [88, 141, 138, 140, 298, 301, 299, 322, 321] may also be important as differential diagnoses of these conditions using traditional image analysis techniques can be challenging [223]. Furthermore, the trained machine learning models may reveal important (i.e., discriminative) connections and subnetworks in the brain which could help researchers better understand disease etiology.

Connectome features have also been used to train models that can discriminate normal control (NC) infants from preterm born infants [15] and infants with neonatal encephalopathy [323] as well as adult NCs from multiple sclerosis (MS) patients [227], patients who have had traumatic brain injury (TBI) [53, 184, 284] and patients with temporal lobe epilepsy (TLE) [146, 188]. As well, learned models have been used to predict the location of epileptic regions [265]. Machine learning models trained on connectome features can also accurately identify patient age [81, 106, 149, 219, 221, 228, 257], sex [273], obesity [208], status as a smoker or non-smoker [207] and restedness [148]. Age, in particular, is an interesting

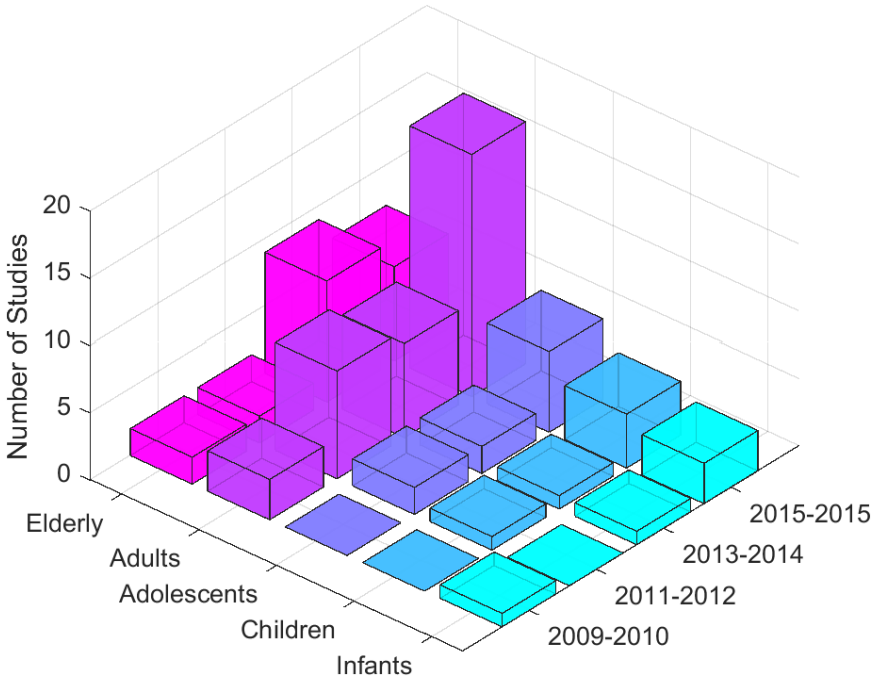

Figure 2.4: 2D histogram of number of studies grouped by age group of study and biennially by year of publication. Note that there are an increasing number of studies not only on adults but also on adolescents, children and infants.

application since a person's physiological or developmental age may differ from their actual age [183]. A discrepancy between predicted and true age may indicate delayed or accellerated brain development or aging.

Finally, functional connectome data, in particular, has been used to train models to distinguish between different functional task loads (including resting) [3, 185, 242], between rest and stimulated states (e.g., watching a movie [226], viewing a face [194], hearing auditory stimuli [269, 287]) and before, during and after memory tasks [195].

A task similar to predicting clinical variables is the task of unique identification or "fingerprinting". Finn et al. found that the functional connectomes of healthy adults best corresponded to their own functional connectomes acquired on the following day, as opposed to the connectomes of others [90].

Table 2.1 shows that the majority of studies in this field have focused on adult and elderly age groups. This is reflected in the large number of studies on conditions like AD and MCI, which primarily affect aging populations. Fig. 2.4 shows, however, that the number of studies focused on infant, child and adolesent age groups is increasing.

Of the 71 studies that performed prediction, 68 performed classification and only 3 performed regression. Of those 68 classification studies, 57 performed binary (2-class) classification and the remaining 11 studies performed multi-class classification. Different tasks

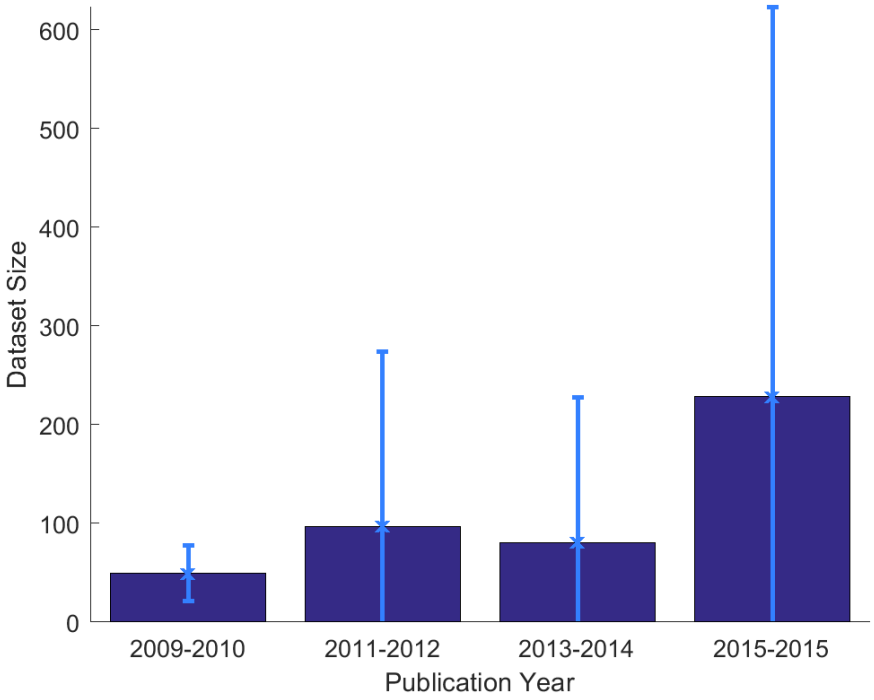

Figure 2.5: Average number of MRI scans per study versus publication year, grouped biennially.

require different machine learning models, which we discuss below in Section 2.4.4.

As was mentioned above, most studies that have applied machine learning models to connectome data have been exploratory (rather than large scale studies), and have trained and validated their models on small datasets only. However, Fig. 2.5 shows that the mean number of scans used per study is growing, as is the standard deviation. This may be in part due to the recent emergence of large public databases such as the human connectome project (HCP) [283], the alzheimer's disease neuroimaging initiative 2 (ADNI-2)[2](which only began collecting dMRI and fMRI data in 2011), the autism brain imaging data exchange (ABIDE) [78] and the ADHD-200 dataset[3]. Pre-processed connectome data is available for both of the latter two datasets from the pre-processed connectome project[4], also initiated in 2011.

The average best reported prediction accuracy across all datasets in all classification studies was 81%. Note however that the difficulty of the prediction task and a variety of other factors can greatly impact prediction accuracy. For instance, Iidaka found that the number of folds in the cross-validation scheme can significantly affect the prediction accuracy (Section 2.4.5) [130]. We found that prediction accuracies were moderately but significantly

[2]http://www.adni-info.org/
[3]http://fcon_1000.projects.nitrc.org/indi/adhd200/
[4]http://preprocessed-connectomes-project.org/

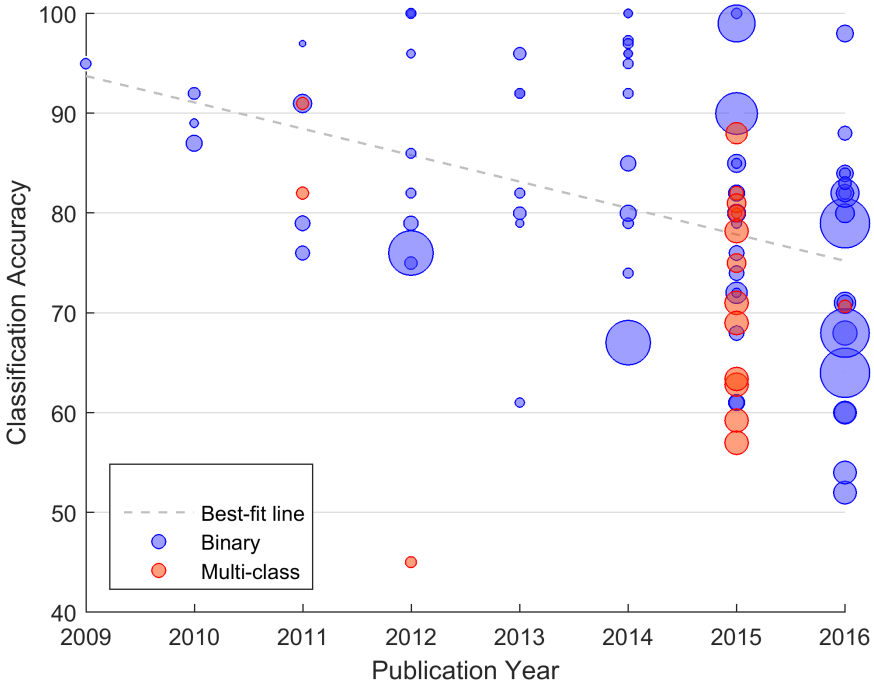

Figure 2.6: Best reported accuracy of each dataset in each classification study vs year of publication. Size of marker represents the relative size of the dataset. Binary and multi-class classification tasks are denoted by blue and red markers respectively.

negatively correlated with publication year, $r = -0.388, p = 1 \times 10^{-4}$ (Fig. 2.6). This may be partially due to the effect of increasing dataset sizes, as the prediction accuracies are weakly but significantly negatively correlated with the number of scans, $r = -0.267, p = 0.01$. However, it is likely also due to researchers exploring increasingly challenging problems such as multi-class classification (e.g., Prasad et al. [217], Takerkart et al. [269]) and prediction of future outcomes (i.e., cases where there is a temporal delay between acquiring the scan and labelling the scan) [45, 46, 149].

2.3.2 Clustering, Subnetwork Extraction and ROI Optimization

A minority of machine learning papers employing connectome data did not perform outcome prediction. In place of supervised learning models, Ashikh et al. [12] used an unsupervised approach to group connectomes into clusters and found that these clusters matched the class labels (NC, early MCI, late MCI and AD). Other studies used connectome data with unsupervised techniques to learn subnetworks containing regions that co-varied across subjects [311] and subgroups of subjects [102] or to learn subnetworks discriminative of TBI [62], post-traumatic stress disorder (PTSD) [168], ADHD [6] or ASD [106]. In the work by Ghanbari et al. different subnetworks were found that were predictive of age and ASD, simultanenously [106]. Yang et al.'s method finds subnetworks that are common across the entire dataset but identifies those connections that are specific to individual groups [305].

Chen et al. used connectome based models to optimize the ROIs for group-wise consistency of connectivity patterns [63].

2.4 Machine Learning Models for Connectome Data

A wide variety of machine learning methods have been explored for use on structural and functional connectome data. Given the unique characteristics of network type data (discussed in Section 2.1.1), many approaches present ways to both reduce dimensionality and leverage the data's special topology. In this section we discuss methods to deal with class imbalance, connectome features, methods of feature selection and dimensionality reduction, learning models, kernels, regularization terms and validation procedures that have been proposed for use on connectome data.

2.4.1 Class Imbalance

When training a learning model, it is often ideal to have a balanced number of examples in each class (or a balanced histogram of labels in the case of continuous labels) in order to not introduce bias into the model. However, of the papers in Table 2.1, 19 validated their approach on datasets where the largest class was more than twice as large as the smallest class (which can be considered very imbalanced). Some of these studies employed strategies to mitigate the effects of class imbalance. For instance, Ball et al. used only a subset of the scans in their majority class, along with all of the scans in their minority class when using ICA to determine ROIs for connectome construction [15]. In contrast, Kaufmann et al. noted that despite their imbalanced classes, they performed ICA on the whole dataset, preferring to include all scans in the definition of ROIs [148]. (Neither study validated their choice experimentally.)

When training their prediction model, Ball et al. also weighted training instances from the minority class proportionally higher, which is similar to replicating minority class instances until the number of instances in each class is equal. Instead, Khazaee et al. used the hold-out method [89], selecting a balanced, random subset of training data in each round of cross validation [151].

Kawahara and Brown et al. used the synthetic minority oversampling technique (SMOTE) [60] to generate synthetic instances in their training set [46, 149]. They generated more synthetic instances of the minority class than those in the majority class in order to balance their training set before training their classifier. Similarly, we used our proposed local synthetic instances (LSI) to augment their dataset, generating only instances of the minority class [45] (Chapter 4). LSI assumes the case of HDSSS. In contrast to SMOTE, which attempts to sample from a manifold of training data that may be too sparsely defined, LSI generates new synthetic instances which are guaranteed to be near to real instances and that are perturbed in a realistic way.

Class imbalance can also affect validation as well, biasing traditional measures such as prediction accuracy, sensitivity and specificity. Jie et al., Wee et al., Galvais et al. and Ball et al. all validated their models using balanced accuracy, defined as the mean of the sensitivity and specificity, which weights the accuracy in each class equally giving a better sense of how the prediction model actually performed on imbalanced test data [15, 100, 141, 140, 301]. In the approaches presented in Chapters 4 and 5, we instead selected only balanced test sets for each round of validation [45, 46].

Class imbalance is not a unique problem to machine learning on connectome data, but the problem is exacerbated when combined with the HDSSS problem, which is common for connectome datasets. More research is required to find new methods (such as LSI) that can balance and augment connectome training sets in a principled and effective way.

2.4.2 Connectome Features

In order to train a model, useful features are first extracted from the connectome (or the connectome as a whole can be used, given a proper representation). Similar to features from grid-like image data, connectome features may range in scale and complexity.

At the finest scale, the feature set may comprise a vector embedding of the edge-wise features in the connectome adjacency matrix. Of the papers in Table 2.1, more than half (42 of 83) trained machine learning models on basic edge-wise features of either structural connectivity (e.g., ROI-pair tract count or mean FA) or functional connectivity (e.g., correlation between mean fMRI signals in pairs of ROIs). Park et al. employed a multi-modal approach and defined connectivity between two ROIs in terms of both structural fiber density and mean nodal functional connectivity of the two ROIs [208]. Similarly, Chen et al. constructed both structural and functional connectomes for each subject and used both connectome types as input to their model [62]. By passing edge features into the learning model, no *a priori* assumptions need to be made about which aspects of the connectome are important. This puts more responsibility on the learning model to extract relevant cues from the data. Additionally, each individual edge feature may be noisy [225, 265] and since the number of edges is often large (i.e., in this case $M = \mathcal{O}(N^2)$), feature selection becomes especially important.

One way of reducing the effect of noise and uncertainty on connectivity measurments (i.e., tract count, FA) is to threshold the value of each edge at an assumed or calculated noise amplitude, resulting in a binary value for each edge weight [140, 287, 291]. Extending this idea, Jie et al. used features of binary connectomes generated from multiple thresholds [141]. Yang et al. generated multiple instances of each connectome with random thresholding at each ege and then fused the instances together via similarity network fusion [290, 306].

Some works on fMRI used edge features but also introduced temporal information. Wee et al. used edge features from regional fMRI signal correlations computed in a set of overlapping temporal windows [298]. They used a fused multiple GLASSO (FMGL) approach

to jointly estimate the signal covariance matrices in all temporal windows simultaneously and edge features from each window were concatened into one large feature vector. In their recent follow-up [300], rather than using a sliding window, they first clustered the fMRI signals temporally in order to find distinct brain states before estimating the functional connectivity networks. Ashikh et al. used a sliding temporal window but extracted the temporal variance in functional connectivity at each edge as the input features to their learning model [12]. While these approaches can better capture dynamic patterns in brain activity, they require estimation of functional connectivitiy from fewer time slices, since a window or temporal cluster is a strict subset of the entire set of time slices, and so may fit the data less well.

Other works in the literature have proposed using features of each node, including features of local connectivity [6, 116, 141, 140, 146, 159, 168, 291, 299]. For instance, from each node in each connectome, Guo et al. and Kamiya et al. extracted nodal degree, local efficiency and betweenness centrality [116, 146]. Given a node, i, and its set of neighbouring nodes, $\mathcal{N}(i)$, the (weighted) nodal degree at node i, is defined as,

$$d_i = \sum_{j \in \mathcal{N}(i)} w_{ij} \tag{2.1}$$

where w_{ij} is the weight of the edge between nodes i and j. Nodal degree captures how well connected a node is to the rest of the connectome. Similarly to edge-wise features, nodal degree features have also been computed within a sliding window across time [168]. Local efficiency (LE) involves the concept of a shortest path, along edges and is defined as,

$$LE_i = \frac{1}{N} \sum_{j,k \in \mathcal{N}(i)} \frac{\left(w_{ij} w_{ik} |P^*_{j,k}(\mathcal{N}(i))|^{-1} \right)^{\frac{1}{3}}}{d_i(d_i - 1)} \tag{2.2}$$

and where $P^*_{j,k}(\mathcal{N}(i))$ is the shortest path of edges, $(p, q) \in E$, between nodes j and k, passing only through nodes neighbouring i [159, 234]. $|P^*_{j,k}(\mathcal{N}(i))|$ is then the length of this shortest path, computed as,

$$|P^*_{j,k}(\mathcal{N}(i))| = \sum_{(p,q) \in P^*_{j,k}(\mathcal{N}(i))} \frac{1}{w_{p,q}}. \tag{2.3}$$

Intuitively, local efficiency measures how robust a node's local neighbourhood is. Given the set of shortest paths between every pair of nodes in the entire connectome, betweenness centrality at node i is the fraction of those paths that pass through node i [234]. Another local measure of network topology, called the clustering coefficient (CC), is defined in terms of triangles of edges between a node, i, and two of its neighbours, j and k. It is computed

as,

$$CC_i = \frac{2t_i}{d_i(d_i-1)} \qquad \text{where} \qquad t_i = \frac{1}{2}\sum_{j,k\in\mathcal{N}(i)} (w_{ij}w_{jk}w_{ik})^{\frac{1}{3}}. \tag{2.4}$$

Note that t_i can be interpreted as the average geometric mean of weights of edges forming triangles around i. If the edge weights are binary then the same formula holds and t_i counts the number of triangles around i. CC has been used as a node-wise feature by several studies [141, 140, 146, 291, 299] and indicates the degree of segregation that a node and its direct neighbours have from the rest of the newtork. High values of CC indicates that a node is part of a segregated module. In contrast, the participation coefficient measures the degree of integration a node has with such segregated modules and so is a measure of centrality, similar to betweenness centrality [234]. Nodes with high participation coefficients are then likely to be interconnecting hubs. Anderson et al. extracted the participation coefficient at each node, along with non-connectome features of cortical thickness and non-image-based metadata [6]. Lastly, some fMRI studies simply used the BOLD signal at each node as features [311, 310]. In these cases, the connectome nature of the data was not explicitly encoded in the input features, but instead it was encoded into the learning models as prior information. Note that if a constant number of features is extracted at each node, there will be $M = \mathcal{O}(N)$ features versus $M = \mathcal{O}(N^2)$ edge-wise features.

Somewhat similar to measures of nodal CC and participation coeficient, Tunc et al. used features of connectivity within and between pre-defined functional subnetworks [273]. In contrast to nodal CC and participation coeficient features, these features were only extracted for each single and pair of pre-defined subnetworks, rather than for each node in the entire connectome.

Rather than using pre-defined subnetworks, Fei et al. extracted the most frequently occuring subnetworks in each class using the graph-based substructure pattern mining (gSpan) algorithm [88]. These subnetworks were then extracted across all connectomes of both classes. While these subnetworks are not necessarily discriminative, they each represent stable parts of the network structure of at least one class. From this set of features, only the most discriminative were selected (Section 2.4.3 below).

At the largest scale, single features of the entire connectome can be extracted (i.e., $M = \mathcal{O}(1)$) [44, 66, 217, 238, 323]. Global network measures are often based on local (nodal) network measures, but instead summarize the entire network toplogy and include measures of network integration, segregation, small-worldness and rich-clubness [234, 281]. We (in Chapter 4), along with Prasad et al., Sacchet et al. and Ziv et al. all extracted multiple global network measures as input features to learning models [44, 217, 238, 323]. In addition, Ziv et al. also extracted a different type of global feature of the network based on subgraph enumeration [323]. To compute this feature, they count the number of possible subgraphs that can be generated by a walk of length 8 (i.e., 8 edges) on a connectome with

thresholded, binary edge weights. These counts are binned by isomorphic subgraphs, and the array of bin counts is used as a feature vector.

The concept of heat flow has also been used to extract global, topological connectome features [66]. Chung et al. simulated heat flow across the network domain of the connectome and measured the maximum heat flow rate, the time to achieve maximum heat flow and the time to achieve near equilibrium.

While these global measures capture salient features of the entire connectome topology in a very compact form and have been successfully used for machine learning tasks on connectome data, they may be less sensitive to local cues and so may not be suitable for applications where local effects are important. Furthermore, while each global measure may have a well undestood topological interpretation, these functions are typically non-invertable. Thus, for machine learning models trained on global network features, analysis of the influence of individual edges may not be possible, making anatomical interpretation and localization of important regions more difficult.

Rather than reducing dimensionality with global measures, Ng et al. treated each adjacency matrix, A, (computed from estimation of functional covariance) as a single multivariate feature representing a point on the manifold of positive definite matrices [195]. Dodero et al. applied the same approach but first computed the graph Laplacian, $L = D - A$, of each connectome, where D is a diagonal matrix with the degree, d_i of each node, i, along the diagonal [79]. In contrast to approaches that collapse edge features into a single vector, these methods retain the matrix structure of the original connectome representation enabling them to use the manifold in which they are embedded.

Finally, while all of the above features have been engineered, high level features may also be learned by a feature selection or dimensionality reduction technique or by the learning model itself. Engineered features can be understood as methods for dimensionality reduction or low level feature selection that incorporate some prior knowledge about the expected structure of the data. However, prior knowledge can be incorporated in other ways, such as model regularization. Throughout the next few sections, different methods for learning high level connectome features are discussed, of which many require basic (low level) edge features as input.

2.4.3 Feature Selection and Dimensionality Reduction

In order to train a model that fits well to the data and also generalizes well to unseen instances, it is useful to select or combine features in a way that retains only the most informative (e.g., discriminative) and least noisy features. While a connectome representation of the brain is relatively compact compared to, for instance, a volume of voxels (e.g., typically 10^3 to 10^5 connections in a connectome vs 10^6 to 10^7 voxels in an image volume), feature selection and/or dimensionality reduction remain important aspects of a machine learning approach, especially when many low-level features are being used and when the number of

datasets to train on are limited. Feature selection methods can broadly be split into three categories: Filters, wrappers and embedded selectors (summarized next and examined in detail in Sections 2.4.3, 2.4.3 and 2.4.4).

Wrapper type feature selectors (Section 2.4.3) are based on a learned model (often a classifier) that is fit to some subset of the training data and identifies discriminative features as those that were most important (e.g., for classification). While these methods can be slower, because they require a model to be trained, often multiple times, they have the advantage of identifying interactions between features that might be independently weak. In contrast, filter type feature selectors (Section 2.4.3) do not train a model and typically decide the importance or weight of each input feature based on a heuristic criterion (often independently) [47]. These types of feature selectors are then efficient and can often be run in parallel across the data, but may discard features that are discriminative when combined together. Embedded feature selectors are similar to wrappers except that they are not run as a preprocessing step but instead are integrated into the final learning model being trained, usually as a regularization term that encourages only a sparse subset of the features to be used. We discuss embedded feature selectors in terms of how they regularize different learning models in Section 2.4.4. Some approaches used multiple feature selection techniques, stacked in order to utilize the advantages of each technique [151, 208, 322].

Along with feature selection, a variety of dimensionality reduction techniques (discussed below in Section 2.4.3) have been applied to connectome data in preparation to machine learning tasks. While similar to feature selection in that the number of output features is reduced, dimensionality reduction techniques act as transforms on the input feature space, rather than filters of individual features.

Filter Type Feature Selectors

Of the papers listed in Table 2.1, the most widely used feature selection technique is the application of a non-paired two-sample t-test to each edge feature [81, 141, 184, 208, 226, 322]. A given edge feature is then only selected if the distributions of feature values (i.e., often edge connectivity strength) across connectomes in the two classes are significantly different, indicating that the feature discriminates between classes. The use of this test for brain networks has also been called a brain-wide association study (BWAS) after the similar genome-wide association study (GWAS), commonly used in genomics [134]. The two-sample t-test, however, assumes that the number of instances in each class (i.e., sample sizes) are equal, which is not true in general. Khazaee instead used the Fisher score [84], which takes the size of each class into account [151].

The t-test also assumes normally distributed data which may not be true in practice, especially for features derived from data with a lower and/or upper bound (e.g., edge tract counts, FA and correlation values). In order to remove the assumption of normally distributed data, Chen et al. [61] used the non-parametric Wilcoxon rank sum test [198].

Shen et al. used another non-parametric test, the Kendel tau rank correlation coefficient (KTRCC) [150] which, for each feature, counts the number of samples that co-vary with their label versus the number of samples that do not [244]. Features are then ranked by this measure and only the top ranking features are selected. Similarly, Park et al. [208] employed permutation testing [256] which is also non-parametric but is often computationally intensive. These kinds of tests are adventageous because no assumption needs to be made about the distribution of edge weights across scans.

To achieve a given statistical significance level, it is assumed that only a single test has been run, which is not the case for tests like the t-test or Fisher score which are run independently for each feature (i.e., M tests). False discovery rate (FDR) [29] has been used with the t-test as a correction for multiple comparisons, in order to account for the number of edges which may be selected by random chance and may not be class discriminative [226]. Multiple comparison corrections allows the significance threshold for feature selection to be picked in a more principled way.

However, all of the filter type feature selection tests discussed above, except permutation testing, assume that each feature is distributed independently of other features. This assumption is unlikely to hold for edge features, especially between those edges that share a node [314]. Correlation-based feature selection (CFS) [119], used by Zhu et al., selects those features that are not only group discriminative, but that are also more *weakly* correlated with other features [322]. This strategy helps reduce the number of redundant features that are selected. Similarly, Gellerup et al. employed minimum redundancy maximum relevance (mRMR) feature selection [212] that uses an objective function to select a set of features with maximum mutual information between features and labels and minimum mutual information between pairs of features [105]. mRMR is different from the rest of the above filter methods in that it examines all of the fearures simultaneously in order to choose which features to remove. Mitra et al. used a t-test with network-based statistics (NBS) [314] which removes the assumption of indpendence between edge features by taking the network topology of the feature domain into account [184]. Lastly, Munsell et al. employed a bipartite graph between nodes in fMRI and dMRI connectomes and selected clusters of nodes in with similar connectivities in both modalities [189]. Thus, their method simultaneously combines functional and structural connectome features and performs connectivity aware feature selection.

While a wide range of filter type feature selection methods have been used for machine learning on connectome data, there exist many other methods (e.g., the Chi-squared test of independence) which have not been used by the papers in this review. Furthermore, of the methods that have been used, it is not clear which one is best. Nevertheless, based on the best attributes of the methods discussed above, it seems that an ideal filter type feature selector for connectome data should 1) select the most discriminative features, 2) reject redundant features, 3) account for dependencies between features imbued by the network

structure of the data, 4) account for multiple comparisons and 5) not assume any specific distribution of feature values.

Wrapper Type Feature Selectors

One of the main drawbacks of filter type feature selectors is that the selected features are not validated for their combined discriminative power as part of the selection process. In contrast, wrapper type feature selectors simultaneously select and validate the set of selected features by computing some measure of fitness for the entire set, often in the form of a classification error. In this section, we first discuss wrapper based selectors that perform this validation using the accuracy or error of a prediction model that is trained over multiple subsets of the full set of features. Next we discuss wrapper based feature selectors that train a prediction model but use only the weights of that model (possibly summed over multiple rounds) to decide which features to keep. Finally, we discuss a wrapper based feature selector that does not train a prediction model but instead uses an objective function that is designed specifically to retain the best features.

To begin, a variety of papers have employed the popular recursive feature elimination (RFE) method [117] which, using some subset of the training data, recursively trains classifiers on features sets, each with one feature removed [141, 140, 166, 207, 291, 297, 299, 323]. RFE then chooses the subset of features for which the classification error changed the least (up or down). This process repeats until only a specified number of features remain. Typically, the classifier used by RFE is the same type as the one that will be finally trained.

Fei et al. instead used the gSpan algorithm to generate a set of candidate subnetworks and then used discriminative subnetwork mining (DSM) to select only those subnetworks which led to the highest classification accuracy when passed to a graph kernel based classifier (see Section 2.4.4, below) [88].

DSM and RFE are part of a group of methods that is called backwards sequential feature selection because the algorithms begin with a full set of features and progressively remove them. The forward sequential feature selection (FSFS) approach, used by Khazaee et al. begins instead with an empty set of features and iteratively adds features that produce the largest improvement in the fitness function (which in their case was defined by classification accuracy on a subset of the training data) [151].

There is another group of wrapper type feature selectors that don't explicitly use the fitness function of the model but instead used the trained weights of a model (e.g., classifier) to decide the importance of each feature. For instance, recursive feature ranking, used by Mokhtari et al., is similar to RFE in that it recursively removes features by training a model [185]. However, instead of using the classification error to determine the importance of each feature, at each round, the recursive feature ranking examines the weight placed on each feature by the trained classifer, and removes the feature with the smallest weight magnitude. Note, that for linear models, the weight with the smallest magnitude will also

be the weight that changes the cost function the least when set to zero, and so these two feature removal criteria will be equivalent.

Vanderweyen et al. and Wee et al. both used a sparse linear regressor called the least absolute shrinkage and selection operator (LASSO) which uses an ℓ_1 regularization term (described in Section 2.4.4, below) to predict the class labels [284, 300]. After training on a subset of the training data, only those features with non-zero model weights are retained for the final prediction model. The ElasticNet method [324], used by Munsell et al., uses the same sparse linear regressor as LASSO but adds an ℓ_2 regularization term called a smoothness term because it discourages the weight on any one feature from becoming very large [188]. In addition to ElasticNet, Munsell et al. also tried using sparse cannonical correlation analysis (SCCA) [121] and a deep auto-encoder based feature selector [124] but found that ElasticNet outperformed both methods on their dataset [188]. Note that BrainNetCNN, proposed by Kawahara and Brown et al. [149], also uses a deep neural network for feature selection but the feature selector is embedded into the learning model, so we discuss this model below in Section 2.4.4.

Kamiya et al. used a feature selector called the Dantzig selector [54] that also involves an ℓ_1 term over the model weights, but that is subject to a constraint that requires the maximum prediction residual, scaled by feature values for that connectome, to be less than a constant [146]. Those features with non-zero model weights are then passed to the final classifier. The Dantzig selector was designed specifically for cases where the number of features, M, is much larger than the number of scans, N, which is often the case in connectome studies. The manifold regularized multi-task feature selection (M2TFS) [139], used by Jie et al., also trains a sparse linear regressor but is specifically designed for selecting nodal features [138]. M2TFS adds a second regularization term, based on the graph Laplacian, that leverages the network structure and encourages features from neighbouring nodes to be selected together. Additionally, M2TFS generalizes the feature selection model to extract multiple types of features at each node, encouraging features from the same node to be selected together.

Finally, Ball et al. used Boruta feature selection [157], which, instead of model weights, uses the frequency of feature selection of a random forest classifier [15]. First, a null distribution of the frequency selection of each feature is created by training a random forest on data with permuted class labels. Then, a second random forest is trained on the data with the correct class labels and only those features selected by the random forest significantly more often than chance are kept for the final classifier.

As is the case for the filter type feature selectors, different wrapper based feature selectors have their own advantages and disadvantages and so it is not obvious which wrapper type approach is best for features from connectome data. Note, however, that of the wrapper type feature selectors used by studies listed in Table 2.1, only M2TFS and DSM leveraged the network structure of the connectome data. Ideally, a feature selector should utilize the

connectome topology to filter features intelligently. Also, an ideal wrapper type feature selector should employ a model that matches the type of model used in the actual learning algorithm, since different learning models may be sensitive to different types of features and different patterns of feature values. Even with matched model types, all wrapper type feature selectors have the drawback that the feature selection model is divorced from the final learning model. This adds additional training time and also means that the set of selected features may not be optimal for the final model. In Section 2.4.4 we discuss embedded feature selectors which take the form of regularization terms in a learning model's objective function.

Dimensionality Reduction Techniques

Rather than filtering out certain features entirely, dimensionality reduction techniques extract salient information by transforming the input features to a lower dimensional space in which the cardinal directions covary with important factors in the data. For instance, five works from Table 2.1 employed principal component analysis (PCA) which uses eigenvalue decomposition to transform the features onto an orthogonal basis representing the major modes of variation in the feature space [45, 184, 218, 228, 323]. Note that Ziv et al. [323] and Mitra et al. [184] both applied PCA after applying an initial feature selection step to reduce the dimensionality of the features even further. One drawback of PCA is that it assumes a Gaussian model over the data, so data that is distributed over more complex manifolds will not be represented correctly.

In contrast, local linear embedding (LLE) [233], used by Shen et al. [244], makes no assumption about the shape or topology of the feature space manifold other than that it can be approximated as a linear function at each point. LLE provides a mapping between the high dimensional manifold and a flattened, lower dimensional space that captures the dominant modes of variation. Shen et al. used the maximum likelihood estimator [164] to determine the intrinsic dimensionality of the feature space manifold for which to map to. After regularizing each correlation matrix to be semi-positive definite (SPD) using GLASSO, Qui et al. reduced dimensionality using their proposed SPD-LLE which extends LLE by restricting the manifold to the SPD Riemannian space.

Given the matrix of feature row vectors for each connectome, $X \in \mathbb{R}^{N \times M}$, Munsell et al. transformed their features using a linear kernel based on the Gram matrix of X [188]. The output of the linear kernel is a new feature matrix $\hat{X} = X^T X$, in which the ith feature of the jth connectome is the similarity between the original feature vectors of the ith and jth connectomes, computed as $\hat{X}_{ij} = X_i \cdot X_j$. Thus, each connectome is described in terms of its similarity to other connectomes. The linear kernel brings the number of features, M, down (or up) to N, which can be especially adventageous if the number of features is far greater than the number of instances (i.e., $M \gg N$) which, as mentioned above, is common in neuroimaging applications. A variety of other studies also used kernels, specifically in

the context of kernel support vector machines (SVM) and so these approaches are discussed below in Section 2.4.4.

In non-negative matrix factorization (NMF) [161], used for dimensionality reduction by Anderson et al. [6], transformed features, $\hat{X} \in \mathbb{R}^{N \times \hat{M}}$, represent weights of a basis set (dictionary) of learned, high-level, non-negative features $D \in \mathbb{R}^{\hat{M} \times M}$ (where $\hat{M}$ is the number of output features) . The original feature matrix is decomposed as $X = \hat{X}D + \varepsilon$ and ε is a residual matrix that is minimized. Due to the linearity of the model, the non-negativity constraint acts as a sparsity term, encouraging small entries in D and $\hat{X}$ to go to zero. Thus, in the case that X represents edge or node features for each connectome, D represents a basis of consistent subnetworks. Note that each subnetwork may be internally connected or disconnected, depending on which edges or nodes covary consistently.

A general drawback of using dimensionality reduction techniques is that they are not bijective functions, and so information useful for visualizing discriminative features (in particular edges, nodes and subnetworks) can be lost. This issue is especially important for connectome data since basic edge and node features have an intrinsic anatomical location (and often a known or hypothesized anatomical function) which makes visualization of learned models very informative.

2.4.4 Learning Models

In this section, we summarize the wide variety of models that have been used for machine learning on connectome data, including supervised prediction models such as linear prediction models, kernel based models, probabilistic models, ensembles and unsupervised models.

Prediction Models

As mentioned above, the majority of the studies discussed in this review performed supervised outcome prediction. Let $X \in \mathbb{R}^{N \times M}$, be a matrix containing feature row vectors for each connectome and let $\mathbf{y} \in \mathbb{R}^{N \times 1}$ be the associated ground truth labels (assuming a single label per connectome). A prediction model can be represented as a function, $f(X_i; \Theta)$, where $X_i \in \mathbb{R}^{1 \times M}$ is a row vector of X, containing features of the ith connectome and Θ is a set of learnable model parameters.

The learning step of many predictive machine learning models can then be expressed mathematically as an optimization of an objective function,

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}}\, \mathcal{L}(\mathbf{y}, \hat{\mathbf{y}} = \mathbf{f}(X; \Theta)) + R(\Theta), \tag{2.5}$$

subject to some (possibly empty) set of constraints, $\mathcal{C}$, where $\mathcal{L}$ is the loss function (or data term), $\hat{\mathbf{y}} \in \mathbb{R}^{N \times 1}$ is the vector of output labels predicted by $\mathbf{f} = [f(X_1; \Theta), \ldots, f(X_N; \Theta)]^T$, and R is the regularization function (sometimes called the prior as it often encodes prior

information). By optimizing over Eq. 2.5, we find a regularized and potentially constrained set of model parameters, Θ^*, that minimize the loss between the labels predicted by the model, $\hat{\mathbf{y}}$, and the ground truth labels, $\mathbf{y}$.

Linear Prediction Models

The most basic prediction model, linear regression, has been used for connectome data with different regularization terms [46, 208]. For linear regression, the set of learnable weights comprises a single vector of weights $\Theta = \{\mathbf{w}\}$, where, $\mathbf{w} \in \mathbb{R}^{M\times 1}$, $f(X_i; \mathbf{w}) = X_i \cdot \mathbf{w}$ and $\mathcal{L}(\mathbf{y}, \mathbf{f}(X; \mathbf{w})) = ||\mathbf{y} - X\mathbf{w}||_2^2$. Note that f may contain a bias term (i.e., y-intercept) but for simplicity, here and throughout the rest of the paper, we assume that this bias term is included as an extra weight in $\mathbf{w}$. Thus, in the case that a bias term is included, we have $\mathbf{w} \in \mathbb{R}^{(M+1)\times 1}$ and $X \in \mathbb{R}^{N\times(M+1)}$ contains an additional, final column of all 1's.

The LASSO model, used by Vanderweyen et al. [284] and Wee et al. [300] for feature selection is simply linear regression with the ℓ_1 regularization term, $R(\mathbf{w}) = \lambda_{\ell_1}||\mathbf{w}||_1$, where λ_{ℓ_1} is a user specified weight (i.e., hyper-parameter) that determines the strength of regularization. Note that in this case, $R(\mathbf{w})$ can be considered an embedded feature selector, as it helps to determine the set of features used in the final prediction model.

As mentioned above, the objective function for ElasticNet is the same as that for LASSO but with regularization function, $R(\mathbf{w}) = \lambda_{\ell_1}||\mathbf{w}||_1 + \lambda_{\ell_2}||\mathbf{w}||_2^2$, where λ_{ℓ_2} is the weight on the smoothness term. Typically, a unity-sum constraint, $\lambda_{\ell_1} = 1 - \lambda_{\ell_2}$, is also enforced. (Note that if only the ℓ_2 term is present, it is called a Tikhonov regularization, and the entire model is called a ridge regression model [110, 125]).

Similarly, we extended LASSO and constructed a prediction model with two additional terms that encouraged features to be selected 1) from edges on the connectome backbone (i.e., edges with high signal to noise ratio) and 2) from edges connected to one another (i.e., sharing the same nodes) [46] (Chapter 5). We found that this model out-performed both LASSO and ElasticNet on a preterm infant connectome dataset.

Park et al. instead employed partial least-square regression (PLSR) [154] which combines linear regression with PCA [208]. Rather than minimizing prediction error between ground truth and predicted labels, PLSR aims to find a set of $L = \text{rank}(X)$ orthonormal latent variables, $\{\mathbf{t}_\ell \in \mathbb{R}^{N\times 1} : \ell \in [1, L]\}$, that covary with the labels. In particular, for a set of L weight vectors, $\Theta = \{\mathbf{w}_\ell : \ell \in \{1, 2, \ldots, L\}\}$, and sample covariance function, $\text{cov}(\cdot)$, the loss is defined as $\mathcal{L}(\mathbf{y}, \mathbf{f}(X; \Theta)) = -\sum_{\ell=1}^{L} \text{cov}(\mathbf{t}_\ell, \mathbf{y})$, subject to the constraints,

$$C_0 : \mathbf{t}_\ell = X\mathbf{w}_\ell \tag{2.6}$$

$$C_1 : if\ \ell \neq m\ \text{ then } \mathbf{t}_\ell \mathbf{t}_m^T = 0 \tag{2.7}$$

$$\text{else } \mathbf{t}_\ell \mathbf{t}_\ell^T = 1 \tag{2.8}$$

which ensure orthonormality of latent variables. Krishnan et al. presented an iterative algorithm to optimize this objective function that involves use of the singular value decomposition [154]. The advantage of PLSR is that dimensionality reduction and predictive model learning are performed in a single stage.

Linear discriminant analysis (LDA) [97] (sometimes referred to as Fisher linear discriminant analysis) is another classical linear model that has been used with connectome data for supervised prediction of categorical labels [3, 53, 61, 148, 228]. In LDA, the feature space distribution of samples (i.e., of connectomes) in each class is assumed to be Gaussian, and the goal is to find a set of weights, $\mathbf{w}$, that minimizes intra-class variance and maximizes inter-class variance. In two-class LDA, given sample class means, μ_0 and μ_1, and a single sample covariance matrix, assumed (via the homoscedasticity assumption) to be shared by both classes, $\Sigma = \frac{1}{2}(\Sigma_0 + \Sigma_1)$, the LDA loss function is defined as,

$$\mathcal{L}(w) = \frac{2\mathbf{w}^T(\Sigma)\mathbf{w}}{(\mathbf{w} \cdot (\mu_1 - \mu_0))^2}. \tag{2.9}$$

The weights can then be computed analytically as, $\mathbf{w} = \Sigma^{-1}(\mu_1 - \mu_0)$. The binary prediction model takes the form of a linear boundary (i.e., hyperplane) with the prediction function, $f : \mathbb{R}^M \rightarrow \{0, 1\}$, defined as,

$$f(X_i; \mathbf{w}) = \begin{cases} 1, & \text{if } X_i \cdot \mathbf{w} > \frac{1}{2}({\mu_\mathbf{1}}^T \Sigma^{-1} \mu_\mathbf{1} - {\mu_\mathbf{0}}^T \Sigma^{-1} \mu_\mathbf{0}) \\ 0, & \text{otherwise.} \end{cases} \tag{2.10}$$

Maximum uncertainty linear discriminant analysis (MLDA) [272], used by Robinson et al. and regularized LDA (rLDA) [97], used by Alnæs et al. and Kaufmann et al., are both variants of LDA in which the covariance matrix is estimated more conservatively using covariance shrinkage [3, 148, 228].

SVM and Kernels

Of papers examined by this review, SVM was the most widely used prediction model (47 papers used SVM or kernel SVM, discussed below) [45, 74, 100, 116, 151, 152, 166, 186, 189, 195, 218, 217, 219, 230, 238, 273, 284, 298, 297, 299, 306, 322, 321, 323]. Rather than explicitly modelling the feature distributions of each class like LDA, SVM defines a class boundary hyperplane that is informed only by the samples nearest to it, known as support vectors. Formally, given binary ground truth labels, $y_i \in \{-1, 1\}$, SVM aims to find a hyperplane, defined by weights, $\mathbf{w} \in \mathbb{R}^{M \times 1}$, which satisfy the constraints:

$$C_i : y_i(X_i \cdot \mathbf{w}) > 1 - \xi_i, \forall i \in [1, N], \tag{2.11}$$

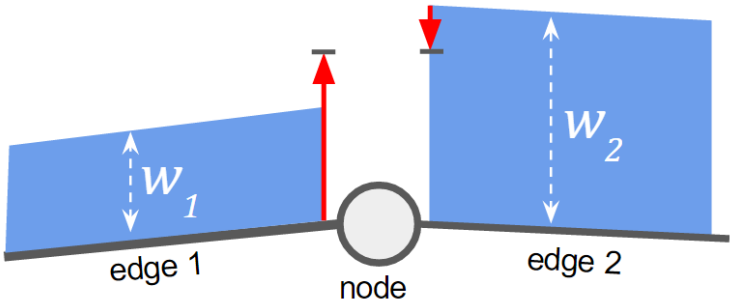

Figure 2.7: A Laplacian regularizer on edge weights encourages pairs of edges that share a node to have a similar weight.

where $\{\xi_i : i \in [1, N]\}$ is a set of slack variables which represent violations of the constraints. The goal of the data term is only to minimize the slack variables, $\mathcal{L}(\mathbf{y}; \xi) = \lambda_\xi \sum_{i=1}^{N} \xi_i$. Because this problem is ill posed, typically an ℓ_2 regularization term is imposed (i.e., $R(\mathbf{w}) = ||\mathbf{w}||_2^2$), though some studies (e.g., Rosa et al. [231]) used an ℓ_1 term instead to promote sparse weights.

Cuingnet et al. and Li et al. both incorporated a Laplacian based regularizer into SVM to promote spatial smoothness in edge features selection weights accross the network [74, 166] (Fig. 2.7). Given the graph Laplacian, $\mathcal{L}$, and regularization strength, λ_β, this regularization term takes the form,

$$R(\mathbf{w}) = ||e^{\frac{1}{2}\lambda_\beta L}\mathbf{w}||^2, \tag{2.12}$$

which resembles a heat flow kernel, diffusing the weights across the network. This diffusion encourages topologically local edges of the network to share strong edge weights and reduces the weight on lone, weakly connected edges.

Prasad et al. used SVM as part of a simulated annealing framework to find the best set of ROIs (i.e., node definitions) [218]. In their case, each ROI represented a non-empty subset of the cortical regions parcellated, in each scan, by an atlas based segmentation. The simulated annealing procedure optimized the ROIs by randomly permuting the parcellated regions in each ROI, quickly at first and then with increasingly smaller pertubations in later iterations. The SVM classification accuracy in each iteration was used to rate the quality of that candidate set of ROIs. Such simulated annealing algorithms are typically computationally intensive, but can provide good quality approximate solutions to NP Hard problems [87].

In order to predict physiological age, Liem et al. used support-vector regression (SVR) [170]. SVR is the regression equivalent to SVM that retains a similar formulation and the notion of support vectors, but replaces SVM's catagorical labels with continuous labels.

Many approaches have used kernel SVM which enables the normally linear SVM to model a (potentially highly) non-linear decision boundary [10, 15, 24, 79, 116, 141, 138, 140, 142, 146, 185, 242, 269, 287, 291, 300]. Kernel SVM modifies the standard SVM

constraint, instead requiring:

$$C_i : y_i(\Phi(X_i) \cdot \mathbf{w}) > 1 - \xi_i, \forall i \in [1, N], \tag{2.13}$$

where $\Phi : \mathbb{R} \rightarrow \chi$ is a function that transforms feature vectors to some new space, χ. The kernel function is then defined as $K_{ij}(X_i, X_j) = \Phi(X_i)^T \Phi(X_i)$ and represents an inner product between X_i and X_j in the new space, χ. Note that if $\Phi(X_i) = X_i$ then K is the linear kernel, and kernel SVM becomes equivalent to standard SVM.

A variety of kernels have been explored for use on connectome data. One of the most common kernels, used by four different studies in Table 2.1, is the radial basis function (RBF) kernel [10, 15, 116, 146]. The RBF kernel is defined as,

$$K_{ij}(X_i, X_j; \sigma) = e^{-\frac{1}{2\sigma^2}||X_i - X_j||^2}, \tag{2.14}$$

where σ is a hyperparameter defining the isotropic standard deviation of the kernel. Note however that the RBF treats each feature equally which may not be ideal when comparing two connectomes, especially if prior information is known about how the topology of the connectomes may vary between classes.

Other approaches have used kernels that explicitly compare different topological features between two graphs including features of random walks, subnetworks and subtrees [185]. For instance, Mokhtari et al., Jie et al., Vega-Pons et al. Fei et al. and Wang et al. all used the Weisfeiler-Lehman (WL) subtree kernel [245], which efficiently counts the number of matching subtree patterns in a pair of networks [88, 141, 185, 287, 291]. Note that the WL kernel can be applied to any two input networks and makes no assumption about correspondence between nodes in the two networks.

A random walk on a graph refers to a sequence of nodes in which the next node is sampled randomly based on the distribution of weights of the edges connected to the current node. Shahnazian et al. used a kernel that measures the similarity between simultaneous random walks, each performed on one of two connectomes [242, 288] (Fig. 2.8). In a comparative study, Shahnazian et al. found that the random walk kernel outperformed the WL subtree kernel [245] and a kernel based on comparing subnetworks [246], in classifying resting from attention states.

Takerkart et al. proposed three kernels that compute similarity in connectome function, ROI locations and ROI adjacency and combined them into a single kernel [269]. In their approach, each connectome was constructed such that edges were defined by spatial adjacency between neighbouring ROIs, nodal positions were defined by ROI centroids and averaged fMRI time series at each ROI defined nodal signals. Given two connectomes, $G(E, V)$ and $G'(E', V')$, the combined kernel is defined as,

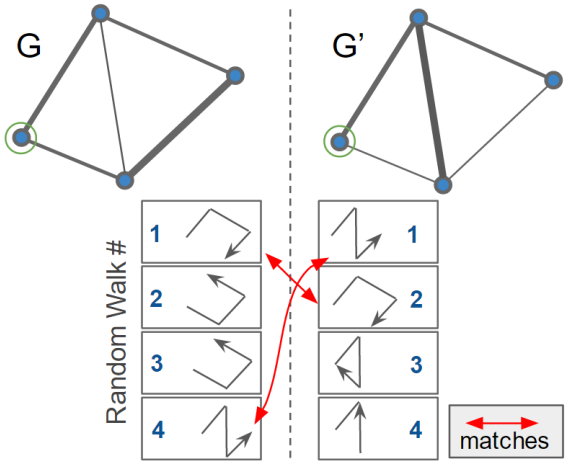

Figure 2.8: Random walk kernel: Given two graphs G and G', the kernel performs random walks on each graph and counts the number of matching walks. In this example, four random walks were peformed per graph and two matches were found.

$$K(G, G') = \sum_{e_{p,q} \in E} \sum_{e_{p',q'} \in E'} k_s(e_{p,q}, e_{p',q'}) \cdot k_g(e_{p,q}, e_{p',q'}) \cdot k_f(e_{p,q}, e_{p',q'}), \tag{2.15}$$

where k_s is 1 if both $e_{p,q}$ and $e_{p',q'}$ exist (i.e., nodes p and q represent spatially adjacent ROIs in G and nodes p' and q' represent spatially adjacent ROIs in G') and 0 otherwise, k_g computes the similarity between positions of nodes p and p' and nodes q and q', and k_f computes similarity between fMRI signals, also between the two pairs of nodes. (Note that there is no special correspondence between p and p' or between q and q' and that Eq. 2.15 sums over every possible pair of edges in the two connectomes.)

Other than the random walk kernel, none of the graph kernels discussed above make an assumption of correspondence between nodes in different connectomes and so are all applicable to cases where connectomes are not constructed using corresponding landmarks or atlas regions. Coversley, for connectomes that are constructed with consistent ROIs across the dataset, these kernel can not leverage the intrinsic correspondence that exists between nodes and edges, and so may be at a disadvantage compared to other measures of connectome similarity.

Probabilistic Models

Logistic regression [127] is a probabilistic linear model that has been used for prediction of outcomes using connectome data but is only applicable when the class labels are binary. Logistic regression uses the same prediction function as the standard linear model, $f(X_i; \mathbf{w}) = X_i \cdot \mathbf{w}$, but uses a different loss function that estimates the negative log likelihood

of the labels, given the features,

$$\mathcal{L}(\mathbf{y}, \mathbf{f}(X; \mathbf{w})) = \frac{1}{N} \sum_{i=1}^{N} log(1 + e^{-y_i(X_i \cdot \mathbf{w})}). \tag{2.16}$$

Of the papers in this review, three used logistic regression and two of those used sparse logistic regression (SLR) [237], which is logisitic regression with the inclusion of an ℓ_1 regularization term [116, 194, 315]. Additionally, Ng et al. also included a regularization term based on the graph Laplacian [194]. Their regularization function, $R(\mathbf{w}) = \lambda||\mathbf{w}||_1 + (1-\lambda)\mathbf{w}^T L\mathbf{w}$, which combines ℓ_1 based sparsity and Laplacian based spatial regularization, is called a GraphNet regularizer [114].

Varoquax et al. classified stroke patients from normal controls using a statistical model of the covariance matrices of the controls only [285]. To do this, they defined a generalized Gaussian distribution over the Riemannian space of covariance matrices. To address the HDSSS problem, they used a fixed variance, greatly reducing the number of model parameters. Instead, Chung et al. used a Gaussian naïve Bayes classifier [316] which models the probability of each class given the input features, which are assumed to be independent and normally distributed [66].

Moyer et al. also employed a naïve Bayes classifier but modelled the feature distribution of each class as distribution of connectomes with assumed community structure using a mixed membership stochastic blockmodel (MMSB) [147, 186]. The MMSB assumes that the membership of each node into K possibly overlapping communities (i.e., subnetworks) is modelled by a multinomial distribution and that the connectivity (i.e., tract count) between any pair of nodes, given the their community memberships, is modelled by a Poisson distribution. An MMSB was learned for each class and then each unseen connectome was assigned the class associated with the MMSB to which that connectome fit with highest probability.

Sweet et al. predicted the locations of epilepsy in the brain by constructing a model of the probability of each ROI in a functional connectome being healthy or epileptic [265]. To predict epileptic regions from observed functional connections, they first introduced latent variables representing the true (but unknown) functional connectivity in normal and abnormal connectomes. Observed functional connectivities of each edge were assumed to be noisy measurements of the true connectivities. They then modelled the probability of an edge connectivity in an abnormal scan, given the connectivity of the same edge in a normal scan and given the health of the two end-point ROIs (i.e., either healthy or epileptic regions). Finally, the probability of given ROI being epileptic in an unseen test connectome was computed as a marginal posterior probability, based on the connectome functional edge connectivities and the learned model parameters.

Generally, probablistic models are riggorously motivated and handle uncertainty in features and labels explicitly which can provide additional information about model predic-

tions. However, this extra information tends to come at the price of (potentially many) additional parameters that must be learned (e.g., $M \times M$ covariance matrices) and/or increased number of samples required per model parameter (e.g., logistic regression requires at least 10 samples per independent parameter versus only 5 for ordinary least squares used for linear regression [211]).

Ensembles and Stacked Models

A variety of papers on the application of machine learning to connectome data used ensemble learning models such as random forests [37], boosting models [95], and deep learning models [160], which combine multiple so-called weak learner models. Six works used random forests classifiers [6, 10, 116, 184, 227, 226]. Random forests are ensembles of decision tree classifiers (often learned in parallel) in which each decision tree (i.e., weak learner) is given a random subset of features and the final prediction is often decided by majority voting among trees [37]. Each tree is trained as a recursive splitting of the dataset at each tree node by the single feature and associated feature threshold value that best splits the data at that node. Functional trees [101], used by Richiardi et al., generalize decision trees by allowing multiple features to define a decision boundary at each node [227, 226]. Also, instead of randomly splitting features, Richiardi et al. split the fMRI signals into four different frequency subbands and trained each tree on connectome features derived from a unique subband.

Rather than training weak learner classifier in parallel (like random forests), boosting algorithms train each classifier sequentially, using the classification accuracies of previously trained classifiers to inform the model parameters of the next classifier. Adaboost [95], a boosting algorithm used by Pariyadath et al. [207], forms a final ensemble classifier by taking a linear combination of the trained classifiers, each weighted by its classification accuracy. Pariyadath et al. used SVM models for each weak learner classifier in the Adaboost ensemble. They showed that this ensemble could predict whether participants of the study were smokers or not by examining functional subnetwork-based features of participants' connectomes [207].

While the k nearest neighbours (kNN) classifier, used by Arbabshirani et al., is not an ensemble method, it does use majority voting to decide the class of each test sample (like random forests) [10]. However, whereas voting is performed by multiple decision tree classifiers in each random forest, in kNN, voting is performed by the k training samples with the most similar feature vectors where each votes with its own class label.

A stacked model is another kind of ensemble in which the outputs of each weak learner are provided as inputs of the next weak learner. Deep learning models are stacked models that posess many layers (e.g., more than 2) and have recently become very popular in certain domains due to their improved performance over other models on a variety of tasks (e.g., AlexNet for natural image classification [155]) [160]. While deep models have been

applied readily to standard grid-like images of the brain [40, 86, 165, 172, 260, 312, 307] and even dMRI and fMRI data (e.g., Plis et al. [213]), their use for connectome data has been much more limited. This is likely because deep models often posses thousands or millions of learnable parameters which require large datasets to train on [160]. As was mentioned in Section 2.3.1, large open MRI connectome datasets have only recently become available and they currently contain hundreds or thousands of scans, but not millions.

Nevertheless, Munsell et al. used a deep stacked auto-encoder [124] to learn a set of relevant edge features to select from each connectome [188]. Breifly, each auto-encoder layer, comprising a complete bipartite graph of inputs and outputs (also known as a fully-connected layer), aims to encode its input into a lower dimensional space with minimal loss of information (i.e., between original input and decoded encoded input). The auto-encoder layers are then stacked such that each layer's output code becomes the input of a smaller (i.e., fewer weights) auto-encoder, where the values of each output are first passed through a non-linear activation function. The result is a model capable of learning highly non-linear relationships in the data. While Munsell et al. found that a linear feature selector outperformed the deep model, their dataset only contain 118 images, and so the deep model likely over-fit to the training set and was not able to generalize well.

One problem with fully-connected layers is that, for M inputs and M' outputs, they require learning $\mathcal{O}(MM')$ parameters, which is often a large number. We proposed a deep CNN called BrainNetCNN with convolutional layers designed specifically for connectome data that leveraged the topology of the input brain networks [149] (Chapter 6). Each of the proposed edge-to-edge convolutional layers required learning only $\mathcal{O}(M)$ parameters. For models with the same number of parameters, we found that the BrainNetCNN model outperformed a model with only fully-connected layers. In order to address the challenges of performing deep learning on a relatively small dataset (168 scans), we also augmented the dataset by generating synthetic connectomes via SMOTE [60].

Generally, ensembles and stacked learning methods enable highly flexible, non-linear models to be trained but require large datasets and greater computational resources to do so. Given that the amount of available connectome data seems to be increasing (Fig. 2.5) and that computational power continues to follow Moore's law, we expect ensembles and deep neural networks in particular to become much more widely studied for applications to brain networks.

Unsupervised Models and Dictionary Learning Models

A small subset of the papers in Table 2.1 used unsupervised learning models on their connectome data, including NMF and clustering algorithms like k-means. The k-means algorithm groups the input data into k clusters by iteratively finding the feature space mean of each cluster and then reassigning each sample to its nearest cluster based on the new mean [92].

Shen et al. used k-means[5] clustering to split the data into two groups after reducing the dimensionality of each connectome via LLE [244]. They found that, in the reduced dimensionality space, the connectome clusters match the ground-truth schizophrenia/control class labels well.

Ashikh et al. instead used the ordering points to identify clustering structure (OPTICS) algorithm [9], which requires, as input, the minimum number of points (in this case connectomes) per cluster in lieu of the final number of custers (as k-means does) [12]. Rather than explicitly determining clusters, OPTICS uses the distance between points to induce a density based ordering on all points in the dataset from which clusters can be inferred. Despite using a Euclidean distance metric between edge features of each connectome, with no dimensionality reduction (aside from filtering via a t-test), Ashikh et al. reported that the OPTICS algorithm was able to group connectomes into three clusters which perfectly matched the early MCI, late MCI and AD class label groups in their dataset.

Li et al. also used k-means clustering but only as a way of algorithmically assigning binary labels to each unlabelled connectome before running Fisher discrimination dictionary learning (FDDL) [168, 303]. In FDDL, the goal is to learn a dictionary, $D \in \mathbb{R}^{M \times K}$, of K atoms (each with M features) and a sparse encoding of the dataset $W \in \mathbb{R}^{K \times N}$, such that,

$$\{D^*, W^*\} = \underset{D,W}{\operatorname{argmin}} \mathcal{L}(X; D, W) + R(W), \tag{2.17}$$

$$\text{and } ||D_m||_2 = 1, \ \forall i \in [1, M]. \tag{2.18}$$

The loss function, $\mathcal{L}(X; D, W)$, encourages the code matrix, W, to encode the original data when projected onto the dictionary, D, and also encourages different subsets of the rows of D (i.e., sub-dictionaries) to encode class specific information. The regularization term, $R(W)$, encourages the codes in W to be sparse and discriminative (via the Fisher discriminant criterion [303]). For full details about the loss and regularization functions, see Yang et al. [303]. Applying this method, Li et al. found two (of the $K = 16$, in their case) learned connectome subnetworks, defined by columns in D, that could differentiate between PTSD and control in 80% of their dataset.

NMF [161], used by Ghanbari et al. (and by Anderson et al. for dimensionality reduction, see Section 2.4.3), is similar to dictionary learning in that the goal is to learn a dictionary, $D \in \mathbb{R}^{M \times K}$, and a code matrix, $W \in \mathbb{R}^{K \times N}$, of X such that the loss $\mathcal{L}(X; D, W) = ||X^T - DW||_F^2$ is minimized (where $|| \cdot ||_F$ is the Frobenius norm) [106]. This is the same as saying that W projected onto D should fit closely to the data X, which was one of the goals of FDDL [303]. However, in NMF, a constraint is added that $D_{ij} \geq 0, \forall i, j$ and also it is assumed that $W = D^T X^T$ (i.e., W is the projection of X onto D) and so the loss becomes $\mathcal{L}(X; D) = ||X^T - DD^T X^T||_F^2$. Like in FDDL, Ghanbari et al. also split the

[5]Shen et al. call their clustering algorithm c-means, which is a fuzzy version of k-means, but the algorithm they describe is actually k-means [244].

dictionary into sub-dictionaries, $D = [\hat{D}, \breve{D}, \tilde{D}]$, with different objectives/penalties for each. However, instead of each sub-dictionary being associated with a class, they encouraged the sub-dictionaries to be class discriminative, age regressive and reconstructive with three separate regularization terms in $R(\hat{D}, \breve{D}, \tilde{D})$.

An advantage of both dictionary learning methods is that, assuming X represents edge features for each connectome (which is true for both Li et al. and Ghanbari et al.) then the atoms of D (or topics as they are referred to in Ghanbari et al.) represent sparse connectome subnetworks which have intrinsic meaning, defined by their parent subdictionary's loss/regularization function, and which can be easily visualized to show which parts of the brain are involved. However, a major disadvantage of both FDDL and NMF is that the number of atoms must be selected *a-priori*, and there may be no clear, principled way to choose this.

Whereas dictionary learning approaches find subnetworks important for groups of connectomes, the stable overlapping replicator dynamics (SORD), proposed by Yoldemir et al. is designed to learn important subnetworks within a single connectome (though they show how the technique can be applied to a group of functional connectomes by concatenating the fMRI signals across time) [311]. A strongly connected subnetwork of nodes can be identified by node weights, $\mathbf{w}$, by optimizing:

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmax}}\, \mathbf{w}^T A \mathbf{w} \tag{2.19}$$

$$\text{such that } ||\mathbf{w}||_1 = 1 \text{ and } \mathbf{w} \geq 1, \tag{2.20}$$

where $A \in \mathbb{R}^{|V| \times |V|}$ is the connectome adjacency matrix. SORD allows multiple, possibly overlapping subnetworks to be identified by iteratively adding a new node and new edges to the connectome (i.e., new row and column in A) after each identification of a subnetwork (Fig. 2.9). The edges of the new node are engineered to destabilizes the previous solution, forcing a new subnetwork to be found. They used a bootstrapping technique to ensure the stability of edges found in each subnetwork. Yoldemir et al. also extended SORD to coupled SORD (CSORD) that can identify subnetworks that are both structurally and functionally well connected [310]. Given adjacency matrices of structural, A^s, and functional, A^f, connectomes of the same brain, and two vectors of nodal weights, $\mathbf{w}^a$, and $\mathbf{w}^b$, the subnetwork can be identified by optimizing:

$$\{\mathbf{w}^{a*}, \mathbf{w}^{b*}\} = \underset{\mathbf{w}^a, \mathbf{w}^b}{\operatorname{argmax}}\, (\mathbf{w}^a)^T A^s \mathbf{w}^b + (\mathbf{w}^b)^T A^f \mathbf{w}^a \tag{2.21}$$

$$\text{such that } ||\mathbf{w}^a||_1 = 1,\ \mathbf{w}^a \geq 1,\ ||\mathbf{w}^b||_1 = 1 \text{ and } \mathbf{w}^b \geq 1, \tag{2.22}$$

for which $\mathbf{w}^a$, and $\mathbf{w}^b$ tend to converge for dense networks, giving a single multi-modal subnetwork (per iteration).

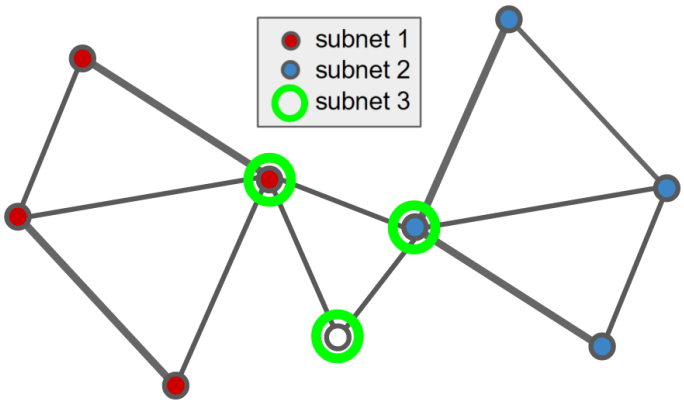

Figure 2.9: Example of three strongly connected yet overlapping subnetworks that can be found by SORD.

Finally, Arslan et al. also proposed a method to group nodes, but instead of replicator dynamics, they employed clustering on an eigenvector representation of the input connectomes. In particular, given a cortical surface mesh with edge weights defined by functional correlation (between spatially local pairs of mesh nodes only), the task was to group mesh nodes into larger ROIs with consistent fMRI signals. To do so, they computed the eigenvalue (spectral) decomposition of the mesh grid's Laplacian matrix and extracted a reduced dimensionality representation of each mesh vertex (by retaining the eigenvectors corresponding to the top eigenvectors). For a single subject parcellation, they applied k-means to the reduced dimensionality mesh vertices, resulting in k ROIs. For a group-wise parcellation, they first created a single mesh graph with correspondence edges between spatially local nodes in different graphs, weighted by the correlations between nodal connectivity patterns. Then, they performed an eigenvalue decomposition on the combined group graph, and again applied k-means to find clusters of nodes across different mesh graphs. The final set of ROIs was computed by majority voting across subjects.

Unsupervised learning methods have provided ways to explore the structure within and between connectomes. Given the complexity of the brain and amount of variation within a single brain across time, or between different subjects' brains across the population, we expect that the use of unsupervised learning for further exploration of the human connectome will continue to expand.

2.4.5 Validation

A wide variety of measures have been used to validate the different machine learning models. By far the most popular measure was standard prediction accuracy (ACC), which was used in 69 of 83 papers. Note however that for imbalanced data (Section 2.4.1), accuracy may over-estimate the performance of the model. Thus, nearly half of the papers in this study also reported sensitivty (SN, also known as recall) and specificity (SP), which convey

how the model performs on positive and negative instances, respectively (assuming 2-class prediction). Other reported validation measures included positive predictive value (PPV, also known as precision), negative predictive value (NPV), F-score (FS) and the area under the receiver-operator characteristic curve (AUC). Some papers have reported variations on these measures including corrected PPV (cPPV) and corrected NPV (cNPV), which take disease (i.e., positive class) prevalence in the overall population into account [4], and balanced accuracy (BAC), which averages accuracy computed over each class separately in order to remove the effect of class imbalance [39]. Wee et al. additionally reported Youden's index (YDI), defined as $SN + SP - 1$, which, similar to balanced accuracy, conveys the sensitivity and specificitiy of the predictor in balanced way [301].

For the minority of studies that performed regression instead of classification, a different set of validation measures were used, including mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMS) and Pearson's correlation between predicted and ground truth outputs (r). Also reported was the area over the regression error characteristic curve (AOC) [32], which provides an estimate of regression error and is analagous to the area under the receiver-operator characteristic curve for classification. Yoldemir et al. didn't perform classification or regression but instead performed unsupervised subnetwork extraction and so, to validate their model, used the Omega index (OmI) [302], which gives a measure of consistency between two sets of graphs [311]. Similarly, Yoldemir et al. and Arslan et al. also measured the consistency of their results using the Dice similarity coefficient (DC), which measures of the degree of overlap between two sets [11, 311].

Finally, in addition to measures which directly report the quality of prediction, Rosa et al. reported the sparsity and stability of feature weights learned by the model. These measures are interesting because they give a sense of how consistent the learned model is across training sets and may suggest how well the model will generalize to unseen data.

The papers in Table 2.1 also used a variety of cross validation schemes to demonstrate the ability of their learning models to generalize to unseen data. Leave-one-out (LOO) cross validation was the most widely used approach, appearing in 41 studies. A variation on this approach, used in particular for cases with imbalanced data sets, is leave-one-subject-per-group-out (LOSGO), which requires at least one instance from each class to be included in every test set, mitigating the bias from class imbalance on the validation measures. Generally, it is possible to perform leave-p-out cross validation, in which a random test set of size p is left out from training in every round. This strategy can be useful for reducing the number of times that the a learning model needs to be trained while retaining the same total number of test instances across rounds. Only 6 machine learning on connectome data studies used leave-p-out with $p > 1$. However, many studies performed K-fold cross validation in which the entire dataset is split randomly into K equally sized, mutually exclusive subsets (called folds) and the model is trained K times on all but one of the folds and tested on the

remaining fold. Often, the random splitting of the data into folds is repeated in multiple rounds and K-fold cross validation is performed once in each round.

As was mentioned above, different cross validation schemes can yield different results, which is at least in part due to the varying sizes of the training sets used to train the model in each scheme. For instance, Iidaka et al. performed 2-fold, 10-fold, 5-fold and leave-one-out cross validation and found wide ranging results (e.g., from 77% to 90% accuracy) across the different schemes [130]. Beyond yielding varying results, different cross validation schemes will impose different computational loads (e.g., number of times the learning model must be trained) and yield different statistical powers (e.g., confidence interval sizes for mean validation measures that are averaged over rounds) based on the number of instances tested and on the number of models trained on unique training subsets. Thus, choice of an appropriate cross-validation scheme depends on the particulars of the data and on the desired subsequent analysis of the results.

2.5 Conclusions and Discussion

Functional and structural connectomes have been proven as informative descriptions of the brain by their use in modelling and classification of a vast array of psychological conditions, neurological conditions, physiological variables and brain states. To do this, researchers have adapted machine learning tools to take advantage of the unique properties of connectome data. Such proposed methods include a variety of local and global connectome features, feature selection and dimensionality reduction techniques that account for the structure and manifold of the connectome data. Additionally, they include graph kernels, topologically aware convolutional layers, objective functions and regularization functions as well as a variety of specialized techniques to extract ROIs and subnetworks from individual connectomes and groups of connectomes. These advances have enabled accurate prediction models as well as techniques for extracting and visualizing salient information from complex brain networks.

Nevertheless, many network specific features and methods for machine learning remain unexplored. For instance, there exist topological measures that, to our knowledge, have never been used in the context of machine learning on connectome data. One such measure is Kirchoff complexity [274], which was proposed as a features of structrual connectomes by Li et al. [167]. Kirchoff complexity is the count of the number of unique spanning trees in a network and provides a measure of its topology. Note that while Li et al. did perform linear regression on these features, they did not split their data into test and training sets and reported results from the context of statistics rather than machine learning (and so this work was not included in Table 2.1). Also, while eigenvalues of the Laplacian (i.e., via spectral decomposition) are fundamental quantities of a network that can be related to almost any other invariant property [67], they remain underutilized, appearing in only one

of the papers listed in Table 2.1. Such network properties may be important in engineering new connectome features that are sensitive to variations in brain networks indicative of injury and disease.

Alternatively, as the size of connectome datasets increase (Fig. 2.5), deep learning will become more feasible and will enable models with the flexibility to capture the wide inter-subject variability found in the brain [64]. It seems likely that with more training data, deep models may be able to improve predictive performance and modelling fidelity on connectome data in the same way that it has in other fields (e.g., natural image classification and natural language processing [160]). One challenge will be to discover how to best leverage the deep learning paradigm for learning on connectome data, given its unique properties. It seems plausible that many of the methods already proposed for learning on connectome data (e.g., graph kernels and Laplacian based regularization) could be applicable in a deep learning context. Another challenge will be to learn how best to augment datasets. Dataset augmentation is a common practice for deep learning on 2D images, multiplying the size of the training set and allowing known symmetries to be implicitly incorperated into the model [58]. Standard augmentation operations, including translating and flipping images (encoding invariances to these transforms), are applicable to images of natural scenes but not to connectome data.

Structured prediction is another branch of machine learning that, for the time being, remains under explored on connectome data. Structured prediction models are akin to regressors or classifiers except that, rather than ouputting scalars or vectors of continuous or categorical variables they output structured objects (e.g., graphs, strings). While none of the papers in Table 2.1 performed structured prediction, the unique topology of connectome data seems to present interesting possibilties for these models. For instance, indentification of particular subnetworks (e.g., finding groups of damaged or disease affected connections) or prediction of thoughts or behaviors from functional connectome data might be amenable to a structured prediction formulation.

A variety of papers listed in Table 2.1 combined structural and functional connectome data and Fig. 2.1 showed that the number of papers using multiple connectome modalities may slowly be increasing. However, most papers that combined modalities used dMRI only to identify landmarks and studied the fMRI connectivity exclusively. Only a select few studies incorporated both structural and functional connectomes into their learning model but those that did found interesting differences and relationships between the two connectome types that highlight the possible advantages of combining modalities [53, 62, 189, 301, 310]. Future work should continue to explore structural and functional connectome data as well as models that combine modalities, especially towards quantifying the strenghts and weaknesses of each modality with respect to modelling different diseases and conditions.

In terms of dMRI alone, only two works extracted features from multiple connectivity measures (e.g., tract count, FA) [45, 299]. However, it remains unclear if any particular

measure of connectivity is most discriminative across a range of learning tasks or if certain measures can provide more information when combined. Thus, an interesting line of research would be to examine how machine learning models weight features of different measures for various learning tasks.

Brain networks that have been studied thus far typically represent connectivity at the largest spatial scale of the brain (i.e., with each ROI representing millions of neurons). As the temporal (fMRI), angular (dMRI) and spatial (fMRI and dMRI) resolutions of scans and scan protocols used to construct connectomes improve, we may choose to construct connectomes that capture these ever finer scales [258, 278]. Thus, yet another challenge will be to design learning and analysis algorithms that can make use of fine scale information across all or part of these brain networks.

As the field matures it will be increasingly important establish the robustness of different machine learning methods on larger connectome datasets with large anatomical variation, especially if these methods are to be used in a clinical setting [289]. As mentioned above, dataset sizes are increasing and a handfull of large, open datasets already exist but only for a limited set of neurological conditions. In order to make use of as much data as possible, standards for connectome construction and/or methods which can incorporate connectomes from different construction methods (e.g., transfer learning [204]), may become important. One step in that direction will be to encourage collaboration and competition in the field, possibly through open challenges with predetermined evaluation approaches and competition leaderboards, which have proved successful in other domains.

We are still, likely, many years away from incorporation of the methods reviewed in this survey as tools into clinical practice. In addition to the lack of large datasets for robust validation and the relatively small number of approaches that have been shown to perform (i.e., predict) with high accuracies (e.g., 95-99%), it is not yet clear exactly how such tools, when they exist, will be used in the clinical workflow. Additionally, machine learning models, which are often 'black boxes', face hurdles appealing to regulatory entities (i.e., FDA) and to clinicians who may wish to understand how a model makes its decisions [56]. Furthermore, not all clinical data is yet structured in a way that makes it ready for computerized analysis (e.g., qualitiative notes). Nevertheless, as methods continue to improve and data becomes more readily available, machine learning techniques for brain network data remain a promising group of technologies for the future of medicine and neuroscience.

Finally, while we have made every effort to cite all papers related to machine learning on connectome data from MRI, some may have been missed. Furthermore, new papers on this topic will continue to be published as the field continues to grow. In order to enable our list of studies to become a living document that remains up-to-date, we have created a website at `http://connectomelearning.cs.sfu.ca/` with a dynamic version of Table 2.1 that the community can contribute to. This website will provide researchers with easy access to

similar works and better exposure for their new research, perhaps helping to encourage competition and collaboration.

Table 2.1: Comparison between different papers that have applied machine learning to connectome data.

Paper	**Image Types**	**Disease / Groups**	**Features**	**Dim. Reduction / Preprocessing**	**Model / Loss**	**Regularization**	**Age Group**	**# Scans**	**ROI Type**	**# ROI**	**Validation**	**Scheme**
Abraham et al., 2016 [1]	fMRI	Autism	ROI signal correlations	-	Linear regression, SVM	L1, L2	Children / Adults (6-64)	871 (ABIDE)	Learned, ICA, Atlas, k-means, Ward	Varying	ACC=68	10 folds, Leave one site out
Alnæs et al., 2015 [3]	fMRI	rest / low / high functional task loads	ROI signal partial + full correlations from GLASSO	-	rLDA	Covariance matrix shrinkage	Adults (29+-10)	101 (37 per class)	Learned (dual-regression)	46	ACC=100 / 70 / 70	LOSGO
Anderson et al., 2011 [7]	fMRI	Autism	ROI signal correlations	t-test	Age-based per-class linear model fit test	-	Adolecents (8-42)	80 (40+, 40-)	Voxel selection (Min. 5 mm radius)	7266	ACC=79, SN=83, SP=75	LOO
Anderson et al., 2014 [6]	fMRI, Cortical Thickness	ADHD	Local network measures + cortical measurements + metadata	NMF	Decision tree	-	Adolescents (8-19)	730 (276+, 472-)	Atlas	90	ACC=67, SN=76, SP=61	LOO
Arbabshirani et al., 2013 [10]	fMRI	Schizophrenia	ROI signal correlations	Remove features effected by medication, various kernels	Many (KNN, decision trees, RBF neural network, kernel SVM: all tied)	-	Adults (25-50)	56 (28+, 28-)	Learned (ICA)	9	ACC=96, SN=100, SP=92, PPV=92, NPV=100	LOO

Paper	Image Types	Disease / Groups	Features	Dim. Reduction / Preprocessing	Model / Loss	Regularization	Age Group	# Scans	ROI Type	# ROI	Validation	Scheme
Arslan et al., 2015 [11]	fMRI	Healthy Adults	Locally connected cortical ROI signal correlations	-	Multi-graph spectral decomposition	-	Adults (22-35)	40	Learned	50, 100, 150, 200	DC=0.72	leave 20 (50%) out, 20 rounds
Ashikh et al., 2015 [12]	fMRI	NC / eMCI / MCI / AD,	Temporal variance of ROI signal correlations	t-test	OPTICS	-	Seniors (ADNI)	123 (35, 34, 34, 29)	Not specified	200	ACC=100	-
Ball et al., 2016 [15]	fMRI	Preterm / term, age	ROI signal covariances	Boruta feature selection	SVM + RBF, Linear regression	-	Preterm (36-49 wks)	131 (105+,26-)	Learned (ICA)	71	BAC=80, AUC=92, r^2=0.57, MSE=8.9	5 folds
Bassett et al., 2012 [24]	fMRI	Schizophrenia	Largest connected component size curve over thresholds	-	SVM	L2	Adults (30-52)	58 (29+, 29-)	AAL	90	ACC=75, SN=85, SP=64	1000 rounds, leave 28 (14+, 14-) out
Brown and Kawahara et al., 2016 [149]	dMRI	Motor, Cognitive Function, Age	Edge tract counts	-	Brain-NetCNN	L2	Preterm	168	Atlas	90	r=0.31, 0.19, 0.86, MAE=10.6, 10.5, 2.3	3 folds
Brown et al., 2015 [45]	dMRI	Low Motor Function	Network Measures	PCA (After net. measures)	SVM	L2	Preterm	168 (22+, 146-)	Atlas	90	ACC=72, SN=77, SP=69	1000 rounds, LOSGO
Brown et al., 2016 [46]	dMRI	Low Motor, Cognitive Function	Edge tract counts	-	Non-negative linear regression	L1, backbone prior, connectivity prior	Preterm	168 (29+, 139-), 168 (13+, 155-)	Atlas	90	ACC=71, 60, AOC=14.3, 17.4, r=0.44, 0.34	1000 rounds, LOSGO

Paper	Image Types	Disease / Groups	Features	Dim. Reduction / Preprocessing	Model / Loss	Regularization	Age Group	# Scans	ROI Type	# ROI	Validation	Scheme
Caeyenberghs et al., 2013 [53]	dMRI + fMRI	TBI	structural and functional nodal degrees	-	Discriminant function analysis	-	Adults (25+/-6)	33 (17+, 16-)	Motor Regions	22	ACC=61, SN=43, SP=77	Not stated
Chen et al., 2011 [61]	fMRI	AD	ROI signal correlations	Wilcoxon rank-sum test	LDA	-	Seniors (69-82)	55 (20NC, 15MCI, 20AD)	AAL	116	AUC=0.87, SN=85, SP=80	LOO
Chen et al., 2015 [62]	dMRI + fMRI	TBI	Edge FA, ROI signal correlations	-	Multi-view Clustering	-	Not specified	40 (16+, 24-)	DICC-COL	358	-	-
Chen et al., 2014 [63]	dMRI	Healthy Adults	Edge tract counts (binary)	-	Training group-wise consistent connections	-	Not specified	120	DICC-COL, learned	358, 272	-	-
Cheng et al., 2012 [65]	fMRI	ADHD	ROI signal partial + full correlations	t-test	SVM	L2	Adolescents (8-19)	730 (276+, 472-)	Atlas	90	ACC=76, SN=63, SP=85	LOO
Chung et al., 2016 [66]	dMRI	Low Motor Function	Heat flow features	Flow averaged between 6 partitions	Gaussian naive Bayes	-	Infants (24 months)	290 (55+, 235-)	AAL	116	ACC=82, SN=75, SP=83, FS=79	10 folds, 5 reps
Cradock et al., 2009 [72]	fMRI	MDD	ROI signal correlations + r-to-z	Reliability filter	SVM	L2	Adults (22-54)	40 (20+, 20-)	AFNI + Relevance to MDD	15	ACC=95, SN=92, SP=96, AUC=96	LOO
Cuingnet et al., 2011 [74]	dMRI	Stroke	Nodal features	-	SVM	Exp. of Laplacian (over nodes)	Adults (24 - 81)	72	Per-voxel	96 x 64 x 24	ACC=76	Not stated
Dodero et al., 2015 [79]	fMRI, dMRI, fMRI	ASD, ASD, Schizophrenia	Graph Laplacians	-	Kernel SVM	L2	Adolescents, Adults (9-18, 9-18, 33+/-9)	79 (42+, 37-), 94 (51+, 43-), 27 (12+, 15-)	Atlas	264, 264, 74	ACC=61, 74, 68	LOO

Paper	Image Types	Disease / Groups	Features	Dim. Reduction / Preprocessing	Model / Loss	Regularization	Age Group	# Scans	ROI Type	# ROI	Validation	Scheme
Dosenbach et al., 2011 [81]	fMRI	Age groups	ROI signal correlations	t-test	SVM	L2	Children / Adults (7-30)	122 (61+, 61-)	Learned	210	ACC=91, SN=90, SP=92	LOO
Fei et al., 2014 [88]	fMRI	MCI	gSpan frequent subnetworks from different thresholds	DSM + WL kernel	Graph-kernel-based	-	Seniors (74 +/- 8)	37 (12+, 12-)	AAL	116	ACC=97, AUC=96	LOO
Finn et al., 2015 [90]	fMRI	Healthy Adults	ROI signal correlations	None, various anatomical subnetworks	Max correlation between feature vectors	-	Adults	504 (126 restA, 126 restB, 126 taskA, 126 taskB)	Atlas	268	ACC=99	-
Galvis et al., 2016 [100]	dMRI	Parkinson's Disease	ROI signal correlations (no EPI correct, 3DMI, PDEC)	ICC + t-test	SVM	L2	Seniors (51-71)	189 (131+, 58-)	FreeSurfer	68	BAC=60	10 folds, 10 reps
Gao et al., 2015 [102]	dMRI	NC / SMC / MCI / AD	Edge tract counts normalized by ROI volumes	-	Multi-graph normalized-cut	Graph Laplacian (over edges)	Seniors (ADNI)	154 (30 NC, 34 SMC, 62 MCI, 28 AD)	FreeSurfer	129	-	-
Gellerup et al., 2016 [105]	fMRI	Parkinson's Disease	ROI signal correlations + network measures across 5 frequency bands	mRMR	Proximal SVM ensemble	L2	Seniors (60+/-10)	45 (24+, 21-)	Atlas	264	ACC=84, SN=73, SP=93	5 folds
Ghanbari et al., 2014 [106]	dMRI	ASD, Age	Edge probability	Built-into NMF	NMF + graph embeddings	-	Adolescents (8-18)	83 (24+, 59-)	Atlas (Desikan)	79	-	-

Paper	Image Types	Disease / Groups	Features	Dim. Reduction / Preprocessing	Model / Loss	Regularization	Age Group	# Scans	ROI Type	# ROI	Validation	Scheme
Guo et al., 2012 [116]	fMRI	MDD	Nodal degree, efficiency, betweeness centrality from partial correlation at multiple thresholds	-	Many (SVM, RBF-SVM, LDA, random forest, logistic regression)	L2	Adults (17 - 54)	76 (38+,28-)	AAL	90	ACC=79	100 rounds, leave 30% out
Iidaka, 2015 [130]	fMRI	ASD	ROI signal correlations	Min. correlation threshold	Probabilistic Neural Networks	-	Children (6-19)	640 (312+, 328-)	AAL	90	ACC=90, SN=92, SP=88, PPV=88, NPV=92, cPPV=8, cNPV=100	2, 10 and 50 folds, Leave one out
Jie et al., 2014 [141]	fMRI	MCI	Local CC from multiple thresholds	t-test, RFE	Multi-kernel SVM	L2	Seniors (74+/-8)	37 (12+, 25-)	AAL	90	ACC=92, BAC=94, SN=100, SP=88, AUC=94	LOO
Jie et al., 2013 [140]	fMRI	MCI	Local CC from thresholded ROI signal correlations	t-test + RFE + Linear and WL Kernel	Multi-kernel SVM	L2	Seniors (65-83)	37 (12+, 25-)	AAL	116	ACC=92	LOO
Jie et al., 2014 [138]	fMRI	MCI	CC on Hypergraph	M2TFS + linear kernel	Multi-kernel SVM	L2	Seniors (65-83)	37 (12+, 25-)	AAL	116	ACC=95, SN=92, SP=96, AUC=96	LOO
Jin et al., 2015 [142]	dMRI	ASD	Edge FA, MD and TC, Multiple Scales	t-test, LASSO logistic regression	Multi-kernel SVM	L2	Infants	80 (40+, 40-)	Atlas	90, 203, 403	ACC=76, SN=72, SP=79, AUC=0.8	5 folds, 10 reps
Kamiya et al., 2016 [146]	dMRI	TLE	Local network measures (4x83=332)	Dantzig selector + RBF Kernel	Kernel SVM	L2	Adults (21-45)	58 (44+, 14-)	FreeSurfer	83	ACC=90, AUC=97	LOO

Paper	Image Types	Disease / Groups	Features	Dim. Reduction / Preprocessing	Model / Loss	Regularization	Age Group	# Scans	ROI Type	# ROI	Validation	Scheme
Kaufmann et al., 2016 [148]	fMRI	Sleep Deprevation	ROI signal partial correlations from GLASSO	-	rLDA	Covariance matrix shrinkage	Adults (20-24)	60 (41+, 19-)	Learned (MELODIC ICA)	27	ACC=73 / 54 / 85	LOO
Khazaee et al., 2015 [152]	fMRI	AD	Local + global network measures (454)	Fischer score	SVM	L2	Seniors (ADNI)	40 (20+, 20-)	AAL	90	ACC=100	LOO
Khazaee et al., 2015 [151]	fMRI	NC / MCI / AD	Local + global network (2909) on global cost efficiency max. network	Fischer score + FSFS	SVM	L2	Seniors (ADNI 64-83)	168 (41NC, 89MCI, 34AD)	Atlas	264	ACC=88, per-class SN, SP, PPV	Cross-validation with hold-out
Li et al., 2012 [166]	dMRI	ASD	Edge connectivity	RFE	SVM	Graph Laplacian (over edges)	Children (7-14)	20 (10+, 10-)	Atlas	68	ACC=100	LOO
Li et al., 2014 [168]	fMRI + dMRI for landmarks	PTSD	Temporal sliding window nodal degree	Manual selection of quasi-static states	K-means + FDDL, finite state machines	-	Adults	95 (45+, 51-)	DICCCOL	358	Avg. diff. of subnet histograms	10 folds
Liem et al., 2016 [170]	fMRI	Age	ROI signal correlations	-	SVR	L2	Adult (19-82)	2354	Atlas (BASC)	197, 444	MAE=4.29, R2=0.87	5 folds
Mitra et al., 2016 [184]	dMRI	TBI	Edge FA	NBS Edge t-test + PCA	Random Forest	-	Adults	215	AAL	116	ACC=68, SN=80, SP=46, PPV=68, NPV=69	10 folds
Mokhtari et al., 2012 [185]	fMRI	Rest/Attention	ROI signal correlations	Recursive feature ranking + WL Kernel	Kernel SVM	L2	Adults (20-30)	38 (19+, 19-)	Atlas	24	ACC=100	LOO

Paper	Image Types	Disease / Groups	Features	Dim. Reduction / Preprocessing	Model / Loss	Regularization	Age Group	# Scans	ROI Type	# ROI	Validation	Scheme
Moyer et al., 2015 [186]	dMRI	AD, Bipolar Disorder	Edge tract counts	-	MMSB + (naive Bayes, SVM)	-	ADNI-2, ?	96 (46+, 50-), 92	FreeSurfer	68	ACC=82, 61, REC=80, 53, PRS=82, 69	10 folds
Munsell et al., 2015 [188]	dMRI	TLE, surgical outcome	Edge tract counts	ElasticNet + linear kernel, or SCCA or deep model	SVM	L2	Adults (18 - 70)	118 (70+, 48-)	Atlas (Lusanne)	82	ACC=80, SN=74, SP=84, PPV=90, NPV=70	10 folds
Munsell et al., 2016 [189]	dMRI + fMRI	Autism	Nodal betweeness of ROI signal correlations + Edge tract counts	Clusters connecting modalities on bipartite graph between fMRI and dMRI nodes	SVM	L2	Adolescents (13-18)	70 (35+, 35-)	Atlas	264	ACC=88, SN=89, SP=86, PPV=87, NPV=89	10 folds
Ng et al., 2016 [195]	fMRI	Before/After Memory Task	Covariance matrices	Matrix whitening transport	SVM	L2	Not specified	102 (51+, 51-)	Atlas	78	ACC=98	10000 rounds, leave 34 out
Ng et al., 2012 [194]	dMRI + fMRI	Viewing a face vs rest	Per - voxel time series.	-	SLR	L1 + Graph Laplacian	Not specified	36 (18+, 18-)	Ward clustering	500	ACC=86	9 folds
Pariyadath et al., 2014 [207]	fMRI	Smoking Status	Correlations within and between subnetwork regions	RFE	AdaBoost + SVM	L2	Adults (28-50)	42 (21+, 21-)	ICA + Clustering	56	ACC=79, PPV=83	LOO

Paper	Image Types	Disease / Groups	Features	Dim. Reduction / Preprocessing	Model / Loss	Regularization	Age Group	# Scans	ROI Type	# ROI	Validation	Scheme
Park et al., 2015 [208]	dMRI, fMRI	BMI	Edge fiber density + fMRI mean nodal degrees	t-test + permutation testing, region prior, functional-structural correlation	PLSR	-	Adults (29)	120 (60+, 60-)	AAL	116	MAE=15%, RMS=5.3	LOO
Prasad et al., 2014 [218]	dMRI	AD	Normalized tract counts	Region grouping, PCA	Simulated annealing + SVM	-	Adults (ADNI-2)	87 (37+, 50-)	Freesurfer	68	ACC=85, SN=88, SP=81	LOO
Prasad et al., 2015 [217]	dMRI	NC / eMCI / MCI / AD,	Normalized edge TC, max flow between ROIs, network measures	-	SVM	L2	Seniors (64-85)	200 (50, 74, 38, 38)	Freesurfer	68	ACC=78, SN=84, SP=69	10 folds
Pruett et al., 2015 [219]	fMRI	Age (6 v.12m) and clincal risk (low, high)	ROI signal correlations	t-test + linear kernel	SVM	L2	Infants (6-12 months)	128 (32 per group)	Atlas	230	ACC=81, 75, SN=78, 81, SP=84, 69	LOO
Qiu et al., 2015 [221]	fMRI	Age	ROI signal partial correlations from GLASSO	LLE-SPD	Linear regression	L2	Adults (22-79)	178	Atlas	80	r=0.59, RMS=12.9	LOO
Richiardi et al., 2011 [226]	fMRI	Resting vs watching movie	Multi-band time series	t-test + FDR	Polythetic trees	Tree pruning + ensembles	Adults (18-36)	15	Altas	90	ACC=97	LOO
Richiardi et al., 2012 [227]	fMRI	MS	ROI signal correlations	-	Ensemble of functional trees	-	Adults (29-45)	36 (22+, 14-)	AAL	90	ACC=83, SN=82, SP=86	LOO
Robinson et al., 2010 [228]	dMRI	Age (young adults / seniors)	Edge FA	PCA	MLDA	-	Adults (20-30, 59-90)	96	Atlas	83	ACC=87, SN=90, SP=88, Bayes err.=0.87	10 folds

Paper	Image Types	Disease / Groups	Features	Dim. Reduction / Preprocessing	Model / Loss	Regularization	Age Group	# Scans	ROI Type	# ROI	Validation	Scheme
Rosa et al., 2013 [231]	fMRI	MDD	ROI signal sparse inverse covariance	-	SVM	L1	Adults (29-58)	38 (19+, 19-)	Atlas	137	ACC=82, SN=74, SP=89	LOSGO
Rosa et al., 2015 [230]	fMRI	MDD, Depression Spectrum	ROI signal sparse inverse covariance	-	SVM	L1	Adults (29-58), (27-49)	38 (19+, 19-), 60 (30+, 30-)	Atlas	137	ACC=85, SN=83, SP=87, sparsity=0.6%, stability=57	LOSGO
Sacchet et al., 2015 [238]	dMRI	MDD	Global Network Measures (9)	-	SVM	L2	Adults (18-55)	32 (14+, 18-)	Atlas	68	ACC=72, SN=71, SP=72	LOO
Shahnazian et al., 2012 [242]	fMRI	Rest / Attention	Bi-variate Granger causality network	Random walk kernel (tried other kernels)	Kernel SVM	L2	Adults (20-30)	38 (19+, 19-)	Atlas	24	ACC=100	LOO
Shen et al., 2010 [244]	fMRI	Schizophrenia	ROI signal correlations	KTRCC + LLE	C-Means	-	Adults (19-30)	52 (32+, 20-)	AAL	116	ACC=92, AUC=96	LOO
Smyser et al., 2016 [257]	fMRI	Age	ROI signal correlations + r-to-z	t-test + linear kernel	SVM	L2	Preterm Infants (36 - 41 weeks)	100 (50+, 50-)	Spheres in Talairach atlas space	214	ACC=84, SN=90, SP=78	LOO
Sweet et al., 2013 [265]	fMRI, EEG	Epileptic Regions	ROI signal correlations	-	Baysean model for abnormal regions given edges	-	Not specified	44 (6+, 38-)	Surface subdivision (50-100mm patches)	1153	-	-
Takerkart et al., 2012 [268]	fMRI	Differing Auditory Stimuli	ROI time series, ROI barycenters	Functional, geometrical, structural kernels	Kernel SVM	L2	Not specified	45 (9 per class)	Clustering (Ward's)	5-30	ACC=45	LOO

Paper	Image Types	Disease / Groups	Features	Dim. Reduction / Preprocessing	Model / Loss	Regularization	Age Group	# Scans	ROI Type	# ROI	Validation	Scheme
Tunc et al., 2016 [273]	dMRI	Sex	Mean anatomical subnetwork inter-connectivity, cognitive + motor scores	-	SVM	L2	Adolescents (15+/-3.5)	900	Atlas (Desikan)	95	ACC=79, 64	10 folds
Vander-weyen et al., 2015 [284]	fMRI	TBI and AD	ROI signal partial correlations	LASSO	SVM	L2	Seniors (70)	69 (40NC, 15AD, 14TBI)	Atlas	264	ACC=82, SN=40, SP=98, PPV=86, NPV=81	LOO
Varoquaux et al., 2010 [285]	fMRI	Stroke	ROI signal correlations	-	Gaussian model over manifold of (control) covariance matrices	-	Not specified	30 (10+, 20-)	Seeded Regions	33	Class log likelihood	LOO
Vegas-Pons et al., 2014 [287]	fMRI	Differing Auditory Stimuli	Thresholded ROI signal correlations	WL kernel	Kernel SVM	-	Not specified	38	Clustering	?	ACC=74	LOO
Wang et al., 2014 [291]	fMRI	MCI	Local CC from thresholded ROI signal correlations	t-test + RFE + gSpan + Linear, WL Kernels	Multi-kernel SVM	L2	Seniors (65-83)	37 (12+, 25-)	AAL	116	ACC=97, AUC=92	LOO
Watanabe et al., 2016 [295]	dMRI	TBI	Edge probability	1000 features with top rank-sum correlations with clinical scores	NMF	Class discriminative subnetworks	Adults	111 (79+, 32-)	Atlas (Desikan)	86	ACC=82, SN=82, SP=81, BAC=82	LOSGO
Wee et al., 2016 [300]	fMRI	ASD	ROI signal correlations from temporal clusters	LASSO + various kernels	Kernel SVM	L2	Children, Young Adults (4 - 22)	92 (45+, 47-)	AAL	116	ACC=71, SN=80, SP=61, PPV=79, NPV=65	10 folds

Paper	Image Types	Disease / Groups	Features	Dim. Reduction / Preprocessing	Model / Loss	Regularization	Age Group	# Scans	ROI Type	# ROI	Validation	Scheme
Wee et al., 2010 [299]	dMRI	MCI	Local CC (for 6 connectome types)	Ranked correlation + RFE	SVM	L2	Seniors (74+/-8.6)	27 (10+,17-)	AAL	90	ACC=89, AUC=93	LOO
Wee et al., 2013 [298]	fMRI	MCI	Temporal sliding window region activations	Fused multiple graphical LASSO + t-test	SVM	L2	ADNI-2 (68-80)	59 (29+, 30-)	AAL	116	ACC=90, BAC=79, SN=76, SP=83, AUC=79	LOO
Wee et al., 2013 [297]	fMRI	AD, MCI	Granger causality networks	t-test on MAR model order distribution + RFE	SVM	L2	Seniors (65-83)	37 (12+, 25-)	AAL	116	ACC=92, AUC=90	LOO
Wee et al., 2012 [301]	dMRI + fMRI	MCI	Structural and functional nodal CC	t-test, Linear + Polynomial + RBF kernels	Multi-kernel SVM	L2	Seniors (64-83)	27 (10+, 17-)	AAL	90	ACC=96, SN=100, SP=94, AUC=95, YDI=94, BAC=95, FS=97	LOO
Yang et al., 2016 [306]	fMRI	MCI	CC on fused randomly thresholded networks	LASSO	SVM	L2	Seniors (68-80)	59 (29+, 30-)	AAL	116	ACC=83, SN=86, SP=80, AUC=81	LOO
Yoldemir et al., 2015 [310]	dMRI + fMRI	7 Functional Tasks	fMRI time series, weighted nodal degree	-	CSORD + SVM	Non-negative weights, fixed weight magnitude	Adults	40	Clustering	200	ACC=79, ICC=0.43, DC=0.63	100 rounds, leave 5 (12.5%) out
Yoldemir et al., 2013 [311]	fMRI	Healthy Adults	fMRI time series	-	SORD	Non-negative weights, fixed weight magnitude	Adults	40	Clustering	200	ICC of net. measures, OmI=0.72	Test-retest

Paper	Image Types	Disease / Groups	Features	Dim. Reduction / Preprocessing	Model / Loss	Regularization	Age Group	# Scans	ROI Type	# ROI	Validation	Scheme
Zhan et al., 2015 [315]	dMRI + MRI	NC / MCI / AD	Edge tract counts	High order SVD	SLR	L1	ADNI-2	202	Atlas	113	ACC=71, SN=68, SP=72, AUC=76	20 rounds, leave 15% out
Zhu et al., 2014 [322]	fMRI, dMRI for landmarks	MCI, Schizophrenia	ROI signal correlations	Edge t-test + CFS	SVM	L2	Adults	28 (10+,18-), 20 (10+,10-)	DICCCOL	358	ACC=96, 100	LOO
Zhu et al., 2014 [321]	fMRI + dMRI for landmarks	MCI	ROI signal correlations	CFS	SVM	L2	Seniors (55-84), (66-84)	28 (10+, 18-), 24 (10+, 10-)	DICCCOL	358	ACC=100, 96	LOO
Ziv et al., 2013 [323]	dMRI	Neonatal Encephalopathy	Counts of binary subgraphs	RFE, PCA	SVM	L2	Neonates	24	Clustering (Recursive partitioning)	100	ACC=79	Cross-validation

Chapter 3

Structural Network Analysis of Brain Development in Young Preterm Neonates

3.1 Introduction

The early configuration and development of the brain's structural network is not yet well understood. *In vivo* analysis of white matter connections typically requires a diffusion magnetic resonance (dMR) image of the brain which, for *in utero* subjects, presents significant challenges [136]. Preterm neonatal subjects provide an opportunity to study the early connectome without the difficulties associated with *in utero* imaging. Understanding the connectomes of these infants is doubly important due to the risk factors associated with preterm birth, including white matter injury and abnormal neurodevelopment [85]. Here, we examine a normative cohort of preterm neonatal infants scanned between 27 and 45 weeks post-menstrual age (PMA) and identify consistent topological and developmental trends in their structural brain networks. Our goal is to develop a better understanding of early brain configuration and growth which will enable future studies to better characterize abnormal development and injury.

Previous works have examined white matter development in young infants. Many early studies focused on voxel-wise measures of fractional anisotropy (FA) and mean diffusivity (MD) [33, 128, 191]. These works discussed the effects of myelination and reduction in brain water over time on increasing FA and decreasing MD [85, 103].

Many other studies have looked at functional network development in young infants [93, 94, 104, 293]. Fransson et al., in particular, examined the resting-state functional network architecture of very young preterm infants (25 weeks mean gestational age) and found that only half of the number of resting-state sub-networks found in healthy adults were present at the preterm stage [94]. Recently, van den Heuvel et al. found that functional

networks in preterm infants agreed well with the underlying anatomical structure [280]. In general, the relationship between functional networks and structural networks is complex and still not fully understood and there is still much work being done trying to explain causal relationships between the two [31, 264].

Some other recent works have focused on the examination of the structural connectome of young infants by performing tractography between numerous anatomical regions in the brain [17, 205, 266, 275, 276, 309, 280]. Takahashi et al. examined results of full-brain tractography qualitatively and described trends across postmortem infants between 17 and 40 weeks [266]. In order to quantify and organize tractography results, many studies abstract the connections in the brain as a network, where nodes typically represent anatomical regions and edges represent some measure of connectivity between those regions. Ball et al. examined connections in the thalamocortical network of preterm infants and showed that early birth correlated with reduced connectivity [17]. Pandit et al. studied the change in connection strengths across scan age and birth age on a cohort of preterm infants scanned as early as 47 weeks post-conception [205]. They reported that the frontal lobe showed a higher rate of development than other regions across their age group. They further noted that babies born prematurely showed lower overall cortical and sub-cortical connectivity than infants born at term.

Other preterm infant studies have looked into summarizing structural connectomes using network measures. Yap et al. examined the development of connectomes in young children, across a range of ages between 2 weeks and 2 years, using measures of network integration and segregation [309]. Tymofiyeva et al. used an atlas-free approach to analyze connectome development in preterm infants, children and adults, also employing network measures to capture topological changes [275, 276]. Very recently, Ball et al. studied a specific network measure known as *rich-club organization* in a cohort of preterm infants [16]. They found that this rich-club structure, known to be present in adult brain networks, emerges as early as 30 weeks PMA.

Such network measures allow high-level summaries of brain network topology which have been shown to be useful, reliable bio-markers in discriminating normal and abnormal brain networks [174, 202, 248]. Rubinov and Sporns recently presented a comprehensive summary of such measures in relation to their use on structural and functional brain networks [234].

To date, network analysis of the entire preterm infant connectome, particularly over anatomically defined regions, has not been done for gestational ages earlier than term equivalent age. This gap is likely because of the difficulties in acquiring a large dataset of subjects at such a young age and because, until recently, brain atlases of young infants were not available. It is possible to perform a similar analysis without an atlas, as demonstrated by Tymofiyeva et al., however, this strategy makes it difficult to identify the anatomical significance of specific connections and sub-networks.

In this work, we compute structural brain networks for a cohort of young preterm neonates and analyze both local and global longitudinal trends. In performing this analysis, we observe that the brain networks of preterm infants show high efficiency and clustering measures across a range of network scales, a result seen in analogous studies of term infants at slightly older ages. We also note that the development of individual region-pair connections is often significantly correlated with age. In particular, we find that connections in the frontal and occipital lobes show high rates of development during this period. Finally, using established brain network measures [234], we see that the preterm infant connectome remains highly efficient and becomes more clustered across this age range, leading to a significant increase in small-worldness. As far as we are aware, this is the first connectome analysis of subjects as young as 27 weeks PMA and the first work to look at whole-brain network integration and segregation in a large, normative cohort of preterm infants.

3.2 Materials and Method

3.2.1 Study Population

To establish normative development of preterm structural brain connectivity, we selected "normal" infants from a prospective cohort described in Chau et al. [59]. This cohort consists of premature newborns born between 24 to 32 weeks post-menstrual age at the Children's & Women's Health Centre of British Columbia. Exclusion criteria included 1) congenital malformation or syndrome; 2) antenatal infection; or 3) large parenchymal hemorrhagic infarction ($>$ 2 cm) on head ultrasound scanning. The use of this data for this study (and the studies detailed in subsequent chapters of this thesis) was approved by the University of British Columbia Clinical Research Ethics Board. The newborns enrolled in this cohort were evaluated with MRI scans in the neonatal period (outlined below) and had neurodevelopmental assessments at 18 months of age (corrected for prematurity) with the Bayley Scales of Infant and Toddler Development, Third Edition (BSID-III) [26] and the Peabody Developmental Motor Scales, Second Edition (PDMS-II) [91]. The 3 composite scores (cognitive, language and motor scores) of the BSID-III have a mean of 100 and a standard deviation of 15. The PDMS-II provides a more sensitive assessment of motor function yielding gross, fine and total motor scores with a mean of 100 and standard deviation of 15. To ensure a normative sample of preterm neonates, we included those infants without acquired brain injury on MRI (no white matter injury, no intraventricular hemorrhage) and with scores on all six composite measures of neurodevelopment within 1 standard deviation of the normal mean ($>$ 85). After removing subjects with low cognitive test scores, detectable brain injury and low image quality (described below), the final number of subjects used in this study was 47 (28 males, 19 females).

Table 3.1: Ages and counts for subjects and subject scans. Post-menstrual ages at birth are shown for the subjects while ages at time of scan are displayed for the scans.

	Counts			Post-Menst. Age (wks)			
	Total	M	F	Mean	SD	Min	Max
Subjects	47	28	19	28.19	2.12	24	32
Scans	70	40	30	35.8	5.29	27	45

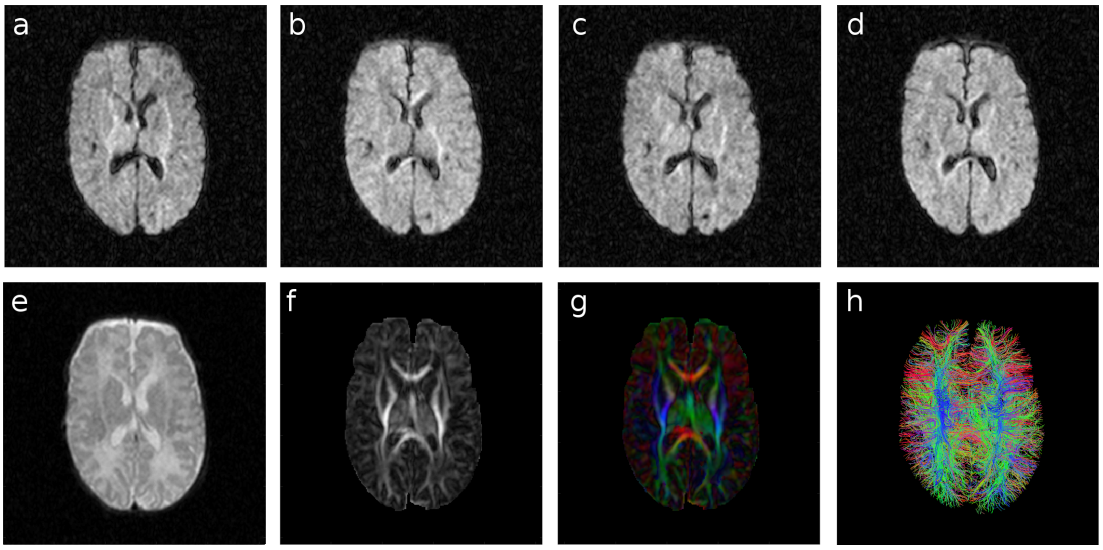

Figure 3.1: Axial slices of example scans of one representative preterm infant. a-b) Diffusion weighted images (DWI) for two gradient directions at $b = 600$, c-d) DWIs for two gradient directions at $b = 700$, d) a b0 image (i.e., DWI with $b = 0$), e) FA map, f) map of primary modes of diffusion tensors, represented by colours (using the standard colour mapping), g) a sample of streamlines from tractography. Note that the SNR of the b0 image is 34.9 dB where as the average SNR of the $b = 600$ and $b = 700$ images is 25.7 dB.

3.2.2 Magnetic Resonance Imaging

Each of the 47 preterm neonates were scanned within the first weeks of life once they were clinically stable. Twenty-three of these 47 infants were scanned again at term-equivalent age, with 2 to 15 (9.49±3.45) weeks between scans. The resulting 70 structural and diffusion MRI scans cover the age range of 27 to 45 (35.8±5.29) weeks PMA (Table 3.1).

Our MRI studies were carried out on a Siemens (Berlin, Germany) 1.5T Avanto using VB 13A software and included the following sequences: 3D coronal volumetric T_1-weighted images (repetition time [TR], 36 ms; echo time [TE], 9.2 ms; field of view [FOV], 200 mm; in-plane resolution, 0.625 mm; slice thickness, 1 mm, no gap) and a 3D axial volumetric diffusion tensor image set (TR 4900 ms; TE 104 ms; FOV 160 mm; slice thickness, 3 mm; no gap) with 3 averages of 12 non-colinear gradient directions over 2 diffusion weightings of 600 and 700 s/mm^2 (b-value), resulting in an in-plane resolution of 1.3 mm. Each diffusion

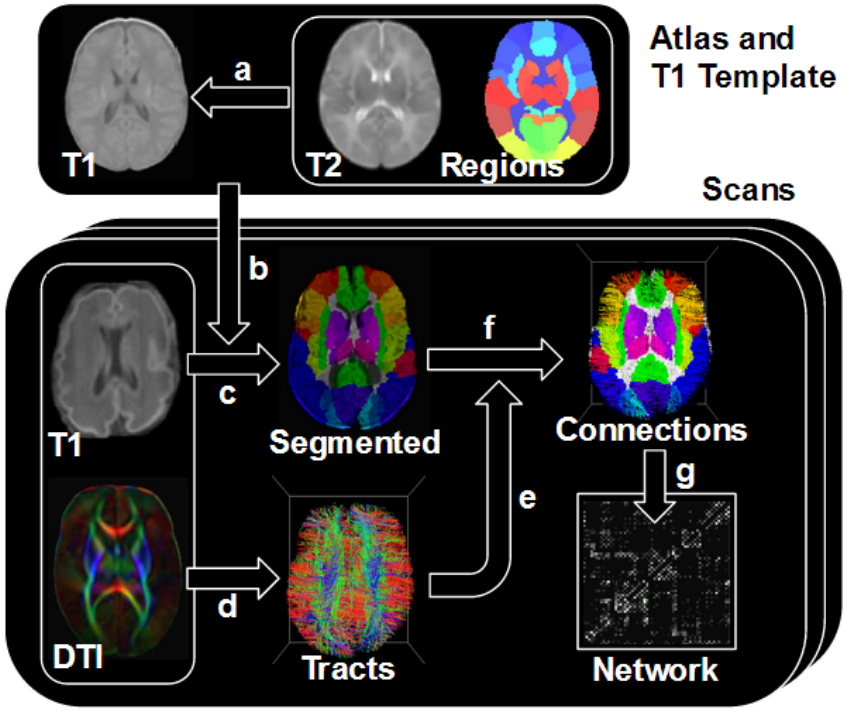

Figure 3.2: High level schematic representation of connectome pipeline. Arrows represent a) T2-T1 registration, b) T1-T1 registration, c) atlas-based segmentation, d) tractography, e) T1-tract alignment, f) assignment of tracts to region pairs, and e) tract counting or mean FA calculation

weighted image set was preprocessed using the FSL Diffusion Toolbox (FDT) pipeline[1] and tensors were fit using RESTORE [57]. Fig. 3.1 shows examples of different diffusion weighted images and computed tensors for one infant from our cohort.

An experienced neuroradiologist reviewed the resulting MR images for presence of white matter injury (WMI), intraventricular hemorrhages (IVH), and poor image quality due to subject motion. The presence of WMI was identified using a system found to be predictive of adverse neurodevelopmental outcome at 12 to 18 months of age [182]. We noted IVH using the protocol of Papile et al. [206]. The 70 scans used in this study were selected so as to be of sufficient quality and be free of these pathologies.

3.2.3 Atlas Based Segmentation

We used an atlas-based method to segment each scan. A brain region atlas of T2 MR images from the IDEA group at University of North Carolina (UNC) School of Medicine, Chapel Hill was aligned to each skull-stripped T1 image in the dataset [247]. The T2 atlas was used due to the lack of availability of an anatomically labelled T1 brain region atlas of young infants at the time of this study. The T2 UNC atlas was constructed from 95

[1] http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FDT

subjects, born between 33 and 42 weeks, with scans taken between 38 and 46 weeks PMA (41.5±1.7 weeks). The atlas contains 90 (45 per hemisphere) anatomically defined cortical (66) and sub-cortical (24) gray matter regions (Table 3.4), each with an average volume of 4.7±3.1 cm^3.

We first register the UNC neonatal T2 atlas to a T1 template of infant brains, followed by aligning the T1 template to each subject's T1 scans. We perform the registration this way as inter-modality registration generally presents more challenges than intra-modal registration and has a higher risk of image registration error. Aligning the T2 atlas to a single T1 template allows the visual examination of a single result to ensure that the inter-modality registration was accurate. The T1 template is a 3D time-slice selected from the 4D Imperial College London (ICL) infant atlas, an atlas constructed from 204 premature neonates between 28 and 44 weeks [240]. We registered to the 3D time-slice corresponding to 38 weeks PMA, the youngest age of the UNC dataset.

Each registration was performed first by an affine transformation using FMRIB's Linear Image Registration Tool (FLIRT) [133], then a deformable (non-linear) transformation using the MATLAB Image Registration Toolbox (MIRT)[2]. For the inter-model registration, both registration steps used a mutual information similarity metric since it is known to be a good choice for such registration problems [215]. For the intra-modal registrations, a cross-correlation similarity metric was sufficient. Anatomical regions in the UNC atlas were propagated to the 38-week T1 template and then to each T1 image in the dataset by applying the same transforms obtained from the registration steps. (Note that we improve and simplify this registration process in Chapter 4).

Prior to registration, each infant's T1 image was neck-stripped using a manually selected slice-plane, then skull stripped first using BET [254] and then using an age-matched, rigidly aligned brain mask from the ICL 4D T1 atlas. The combination of this skull stripping and registration gives us an anatomical segmentation of each T1 image in it's native frame of reference which, upon careful visual inspection, was found to be highly accurate.

3.2.4 Connectome Mapping

We perform whole-brain tractography on each infant's DTI using TrackVis [292]. We seed streamlines at all voxels with FA greater than 0.1 since this is the value of the noise floor [145]. By choosing this FA threshold, we obtain all possible tracks that can be extracted from the DTI. However, this strategy leads to a developmental bias as the number of streamline seeds depend on the brain volume and overall FA levels in the DTI. We will address this bias later in Section 3.2.6. We spline filter all tracks (which models each track as b-spline to reduce memory usage) and then align the tracks to the infant's T1 image space.

[2] https://sites.google.com/site/myronenko/research/mirt

Table 3.2: Network measures used in this study.

Category	Network Measure Name	Acronym
Connectivity	Mean Weighted Nodal Degree	MWD
Integration	Characteristic Path Length	CPL
	Global Efficiency	GE
	Local Efficiency	LE
Segregation	Clustering Coefficient	CC
	Modularity	ML
Small-World	Small-Worldness	SW

The number of streamlines with end-points in each pair of anatomical brain regions were counted to create an n by n connectivity matrix, C^k, ($n = 90$) for each scan, k, where c^k_{ij} is the number of tracts in scan k between regions i and j. Note that each streamline is counted only once as we count a streamline based its end points and not whether a streamline goes through a region. Also, streamlines connecting a region to itself are ignored, causing the diagonal entries of the matrix to be 0. While the streamlines provided by TrackVis are presented with an implied direction, the direction of diffusion measured by the underlying DTI is ambiguous. Thus, we discard the directional information in each network by summing each connectivity matrix with it's transpose, imposing symmetry. No dialation of brain regions was performed since a large number of tracks (e.g., $3.07 \times 10^5 \pm 1.10 \times 10^5$ per scan) were found to be connecting pairs of regions without additional processing.

Using this symmetric connectivity matrix, we define our structural connectome as a network $G^k = (V^k, E^k)$ (for each scan $k = 1 \dots 70$). In this network, the brain regions are represented as a set of nodes $V^k = \{v^k_1, \dots, v^k_n\}$. The connections between brain regions are represented by edges $E^k = V^k \times V^k = \{(v^k_1, v^k_2), \dots\}$ where $e^k_{ij} = (v^k_i, v^k_j)$ represents the connection between regions i and j. The connectome network is then weighted by assigning a weight $x(e^k_{ij})$ to each edge e^k_{ij} corresponding to the amount of structural connectivity between the two regions. This connectome weighting usually takes one of two forms: we can assign an edge weight based on the raw number of streamlines between regions (i.e., $x(e^k_{ij}) = c^k_{ij}$), or we can examine tract "integrity" by using the mean FA computed over the c^k_{ij} streamlines as an edge weight [17]. (To measure the FA of an individual streamline, FA is sampled from the image uniformly along the track and averaged.) We will examine both weighting schemes and refer to their networks as tract-count and mean-FA connectomes respectively.

Figure 3.2 summarizes, at a high level, the pipeline used to generate each brain network.

3.2.5 Network Analysis

The structural connectomes defined in the previous section are networks and it has become popular to examine these networks using various network measures [16, 174, 202, 234, 248, 309]. The collection of network measures used here are given in Table 3.2. At a high

level, these measures capture four intrinsic properties of networks: connectivity, segregation, integration, and small-worldness. Below, we describe the relationship between these measures and the network properties they capture. For a more extensive discussion of network measures, we refer the reader to Rubinov et al. [234] or to our discussion of them in Chapter 2.

To begin, the network property of connectivity refers to the number of connections between nodes. A highly-connected network has more edges, or more highly-weighted edges, than a marginally-connected network. In the context of our connectomes, connectivity directly relates to the number of streamlines between brain regions, or the average FA value of tracts between two regions. We can capture this network connectivity using the *mean weighted nodal degree* (MWD) measure, which reports the average number of connections for a network node. Computing MWD involves computing each node's degree: the sum of edge weights for edges connected to a node. The MWD is then average of these node degree values across the network.

Complementing network connectivity is the property of network integration. While connectivity refers to the amount of connectivity in a network, integration refers to how densely connected all nodes are to one another. For example, a fully connected network has high network integration as each node is connected to every other node. On the other hand, a network with fully disconnected parts has low network integration. In the context of our connectomes, network integration relates to how interconnected all brain regions are to one another. The most common network integration measures are *characteristic path length* (CPL), *global efficiency* (GE), and *local efficiency* (LE). A network's CPL relates to the length of the shortest paths through the network, where path length is the inverse sum of the edge weights for edges along a path. The CPL measure is computed as the average of these shortest path lengths between all pairs of nodes in the network [296]. Similarly, GE measures network integration using shortest paths and, in fact, is simply the average inverse of the shortest path lengths [234]. Both of these measures examine shortest paths through the entire network. However, we can also look at shortest paths locally by measuring the length of the shortest path between the neighbours of a given node. These local shortest path lengths are used to compute LE. Like GE, LE is the average inverse of the shortest path lengths, but LE uses these local shortest paths instead of the overall shortest paths [159].

Conversely to network integration, network segregation refers to how much a network is organized into a collection of sub-networks. For example, a fully connected network, where each node is equally connected to all other nodes, has low network segregation; whereas a network with fully disconnected parts has high network segregation. In the context of our connectomes, network segregation relates to what degree a brain's anatomical regions are arranged into small, distinct clusters. There are two popular measures that capture network segregation: *modularity* (ML), and *clustering coefficient* (CC). The ML measure captures network segregation by measuring the number of connections (i.e. the sum of

the edge weights) within a sub-network and comparing that to the number of connections exiting the same sub-network. ML is then taken as the maximum of these ratios across all possible sub-networks [192]. As it is infeasible to create all possible sub-networks, ML is typically approximated using an algorithm like that proposed by Newman [193]. Similarly, the CC of a network captures its segregation, but unlike ML, CC can be computed exactly. A network's CC captures the fraction of a node's neighbours that are also connected to each other. These fractions at each node are then averaged over the network to give the overall network CC. Each node's CC can also be weighted by the product of the relevant edge weights to obtain a weighted version of this measure [296, 200]. This weighted CC is what is calculated in our study.

Finally, the small world property of a network builds off of these simpler network properties to capture something more complex. A small work network is one that has evidence of both segregation and integration, where nodes are grouped into sub-networks but those sub-networks are also connected in an organized way [20]. A small-world network is known for its efficiency; it keeps high connectivity between nodes with a minimal number of connections. This small world property can be captured using the *small-worldness* (SW) measure, which is simply a normalized ratio between CC and CPL (where the normalization is discussed in Section 3.2.6). Many complex networks, including functional and structural brain networks of animals, have been shown to have this small-world property. It has also been shown that SW is reduced in brain networks of patients with neurodegenerative diseases, suggesting that SW might be an integral property of a well functioning brain [20]. It is not yet known when SW emerges in human brain networks and, given the early ages of the subjects in this study, we have the opportunity here to study how SW develops.

3.2.6 Normalization

By seeding tractography at every voxel with FA>0.1, the number of tracts generated differs between scans. In particular, brain volume and white matter maturation may affect the number of tractography seeds and thus the edge weights in our connectomes. Since the focus of this work is more on the structural topology of white matter than on its volume or degree of maturation, we normalize each connectome in a way that minimizes the effect of these variables. Two different normalization strategies are adopted here and used where most appropriate.

When analyzing changes in individual connections (Section 3.3.2) with respect to age, we normalize each connectome by dividing by the total number of tracts. Edge weights in each normalized connectome then represent relative connectivity and are independent of seeds used when performing tractography.

We use a different strategy when analyzing results from network measures (Sections 3.3.3 and 3.3.4). While normalizing the edge weights in the connectome removes the tract count bias, an additional bias factor exists due to the fact that a random, 'null-hypothesis' net-

work may have a non-zero value for a given network measure. We address this additional bias in the network measures by comparing each measure's value to one computed from a similar network where edges have been placed randomly. These randomly-built networks are generated with the same number of nodes as our connectome and the same distribution of node degrees, thereby removing any bias due to connectome magnitude as well as the fraction of a measure's value that arises by chance [234]. In this work, we compute ten random networks for each connectome and the network measures of these random networks are averaged. The normalized network measures are then obtained by dividing the original, unnormalized, measures by the average obtained from the randomly-built networks.

Note that we do not normalize the mean-FA connectomes because the FA values for a region-pair connection are averaged across all its c_{ij}^k streamlines, making it much more invariant to the number of tractography seeds used. Also, we do not wish to normalize for increasing FA since examining change in FA and comparing it to change in tract count is the exact reason these connectomes were computed.

3.3 Results

With a brain network computed for each subject scan, we performed four experiments designed to expose important aspects of early connectome development in preterm infants. We first investigate group-wise properties of the preterm connectomes using high-level network measures and compare the results to network measure values from a group of infants born at term (Section 3.3.1). We then examine longitudinal trends in for each region-pair connection using a linear mixed-effects model (Section 3.3.2). Finally, we look at network measures in the preterm infant brain networks across time, first as trends within the entire set of scans (Section 3.3.3), and then as changes between scans in each subject (Section 3.3.4).

3.3.1 Groupwise Connectome Analysis

Our first goal is to show that our results are reasonably consistent with results from a similar study. In particular, we repeat an experiment by Yap et al. originally done on a group of young infants born at term [309]. By doing so, we are able to both validate our connectome pipeline against another from a different group and compare preterm infants to those born normally at term with similar scan ages. Such a validation is important as it has been suggested that comparing different connectome studies is difficult due to the large number of methodological choices and parameters [25].

In attempt to make the comparison fair, a subset of scans from our dataset is selected to match the distribution of ages from the term group in [309] as closely as possible. By selecting scans acquired after 39 weeks PMA, we obtain a group of scans with mean age 41.7±1.9 weeks; very close to the mean age of the term group at 41.6±1.7 weeks. We also

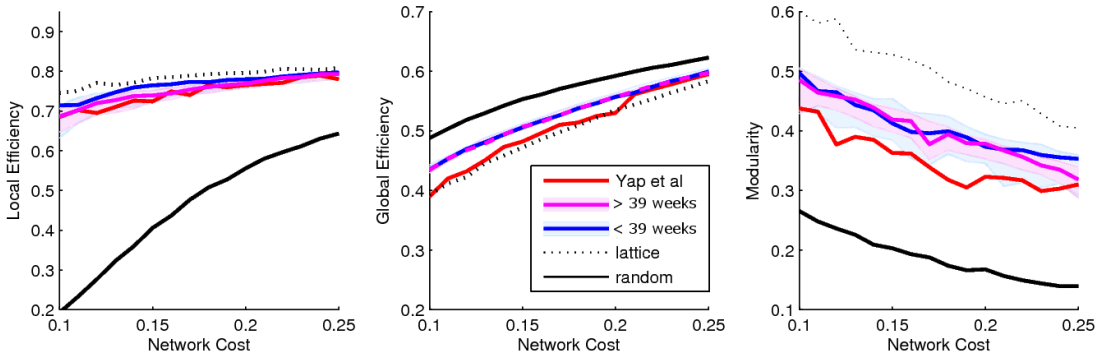

Figure 3.3: Backbone network cost v.s. local efficiency, global efficiency and modularity. Network cost is varied by applying different thresholds to the group connectivity SNR matrix (see text). Network measures are plotted for three groups of infant scans. Results from Yap et al. of a group scanned 2 weeks after normal term birth (red) are plotted next to group of preterm infants scanned around term-equivalent age (magenta). A third group of preterms scanned before 39 weeks PMA (blue) is also shown. Bootstrap-estimated confidence intervals for our preterm groups appear in light magenta and light blue. Note that the preterm and term groups show a similar trend in connectome structure with some slight differences in global efficiency and modularity. We elaborate on these differences in Section 3.4.

compare these groups to a third group, comprised of all remaining preterm scans, taken before 39 weeks (32.6±3.1 weeks).

The connectome results from Yap et al. rely on the concept of a backbone network computed from a group of subjects [111, 309]. This backbone network captures connections that are consistent across the group through the use of a signal-to-noise (SNR) connection matrix. For each group, $\mathcal{G}$, an element in the SNR matrix is defined as,

$$SNR_{ij}^{\mathcal{G}}(c_{ij}^k) = \frac{MEAN_{k\in\mathcal{G}}(c_{ij}^k)}{SD_{k\in\mathcal{G}}(c_{ij}^k)}. \tag{3.1}$$

where SD is the standard deviation across subjects. Note that to stay consistent with Yap et al., we only use the tract-count connectomes when computing this matrix. The backbone network is then built from the SNR connection matrix by thresholding over a range of values to produce binary matrices of varying network costs (i.e., the fraction of possible edges present after thresholding). Each binary matrix can be interpreted as a network in the fashion described earlier and we can compute the GE, LE and ML for each of them. These three measures of the thresholded backbone network are plotted versus network cost in Figure 3.3. Also plotted are the GE, LE, and ML for the special cases of random networks (where edges are distributed randomly) and lattices (where nodes are connected to form a network with a grid-like structure). We also extended the analysis of Yap et al. by using statistical bootstrapping to estimate confidence intervals for all three network measures

(using 50 bootstrap iterations of 39 samples with replacement). These confidence intervals give us further insight into the stability of the backbone network structure.

As we would expect, both LE and GE rise with greater network costs (i.e. more edges) in all groups. Also expected is that the three measures for our connectomes fall in the range between the extreme cases of random networks and lattices. These outcomes are consistent with those reported by Yap et al. for infants born at term and unsurprising since we expect the brain to have a strong local structure combined with cross-network connections. However, we do see notable differences between our preterm connectomes and those term infant connectomes analyzed by Yap et al.. First, the GE of our two preterm groups is higher for most network costs than seen in the group born at term. Second, the ML of our preterm infants is significantly higher than those reported for term infants. Possible interpretations of these differences are discussed later in Section 3.4.

3.3.2 Per-Connection Analysis

In order to understand inter-region white matter development over the range of scan ages in our group, we examined each connection in our tract-count and mean-FA connectomes across time. Since certain subjects were scanned twice, we required a statistical model that assumes possible intra-subject biases. Here we employed a linear mixed-effects model, a generalization of a linear regression model that assumes possible correlation between scans from the same subject [199]. A model of this kind is fit to the tract counts and mean FA values of each region-pair connection across PMA.

Our particular interest is to discover which region-pairs connections are consistently exhibiting longitudinal trends across the cohort. To test for this, we compute a confidence interval (CI) for our linear model's slope and check to see if that CI contains 0. If the CI falls strictly above zero (or strictly below zero) then we can say it is significantly likely that the values are increasing (or decreasing) across time. To correct for multiple statistical computations across the $4050 = 90 \times 90/2$ region-pairs, we adopt the approach of Benjamini et al. to select significant CIs using a false discovery rate (FDR)-based criterion [29].

After correction for multiple hypothesis testing, 664 (of a possible 4050) region pairs had tract counts with slopes significantly likely ($p < 0.01$, FDR corrected [29]) of being non-zero. The locations of these region pairs and the magnitudes of their slopes are displayed spatially in Figure 3.4. 571 of these slopes were positive implying that a majority of these connections are gaining tracts over time.

After normalization of total number of tracts in each connectome, the same analysis was run. As the edge weights in these normalized connectomes convey the relative strength of connections between regions, we expect a greater balance between those increasing in strength over time and those decreasing in strength. Of region-pairs with slopes found to be significantly likely of being non-zero, only 211 were positive where as 250 were negative. As expected, many of the region-pair connections with steeply increasing tract-counts also have

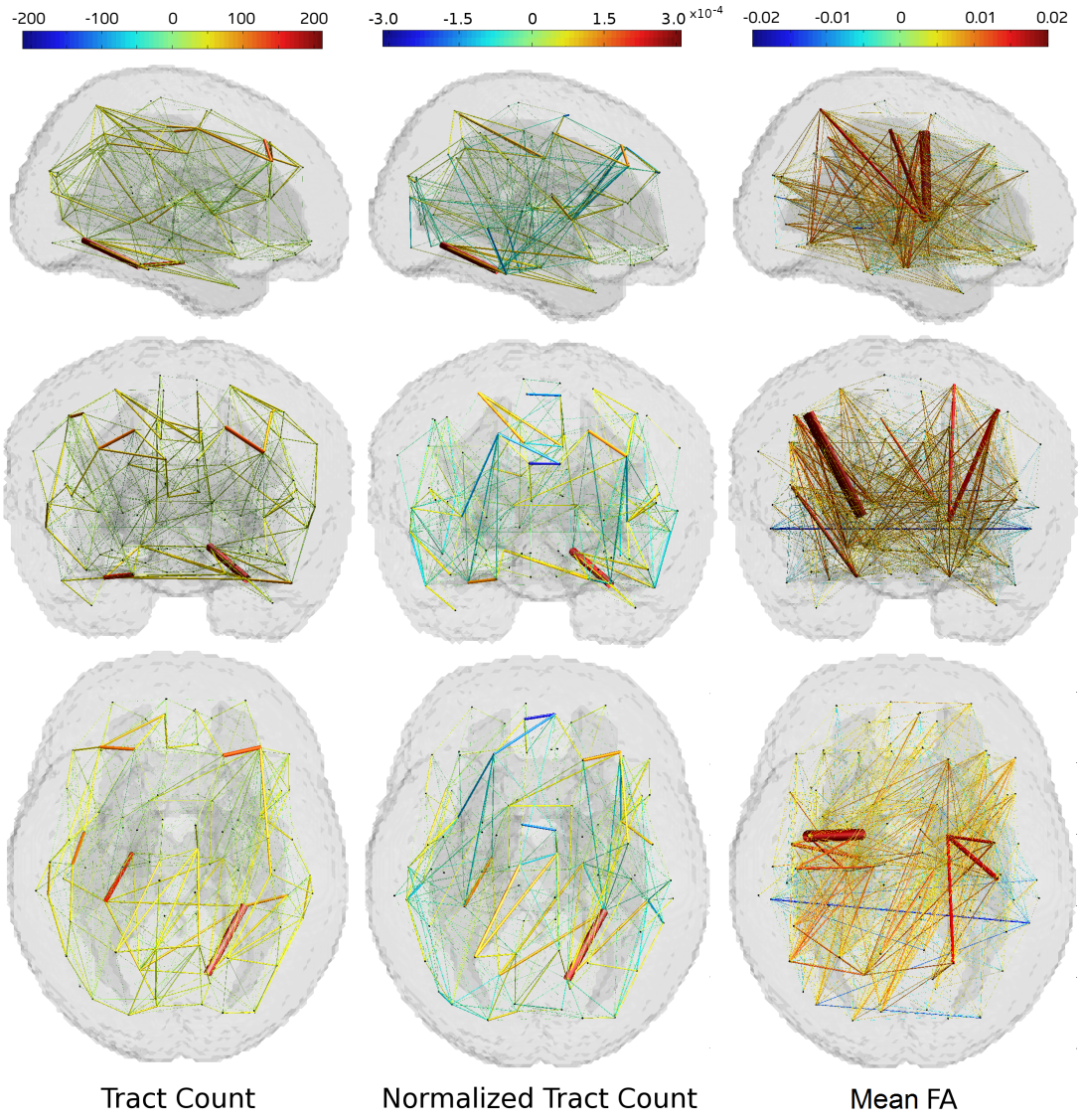

Figure 3.4: Edge weight slopes for linear mixed-effects models fit longitudinally to tract-count, normalized tract-count, and mean-FA connectomes. Connections are mapped spatially onto UNC atlas. Connections are only displayed if the 99% CI of their longitudinal slope did not include zero (after correction for multiple comparisons). Edge colour maps to the linear model's slope value and thicker lines map to greater slope magnitudes.

steep positive slopes in the normalized tract count connectome (Figure 3.4). Also expected is that connections with negative slopes in the normalized connectomes are frequently between regions where there was no significant positive trend in the unnormalized connectomes.

Of the 211 connections with positive slopes in the normalized connectomes, 40% are within the left hemisphere, 38% are within the right hemisphere and 22% are inter-hemispheric. Connections with negative slopes are distributed between hemispheres in a similar fashion. We also find that 34 region pairs in the frontal lobe (atlas regions 1 through 30, Table

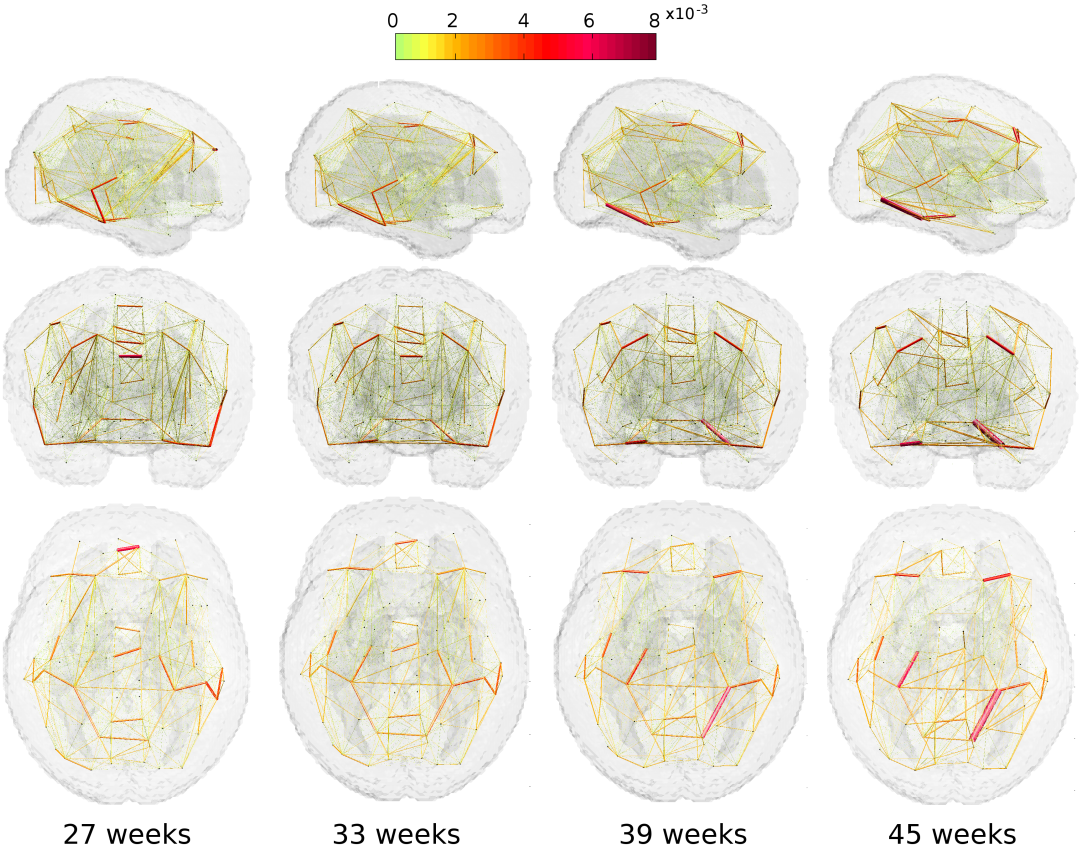

Figure 3.5: Predicted normalized tract-count connectomes exposing relative changes in network topology from 27 to 45 weeks PMA. Edge weights are predicted at four ages using linear mixed-effects models fit to each connection across scans. Connections with at least 0.1% of the brain's fiber tracts are rendered spatially over UNC atlas from sagittal, coronal and axial views (top to bottom). Each connection's thickness and colour represents the predicted fraction of tracts in the brain that connect each region pair.

3.4), had normalized tract counts significantly increasing with age. Meanwhile, the occipital lobe (atlas regions 37 through 66), there are 28 such connections. The development of connections in these lobes was noticeably higher than the 23.0±5.73 developing connections obtained for a random selection of 30 regions (over 1000 random samplings with replacement).

We also examined the split between developing cortico-cortical connections and those connections with deep gray matter regions. In the normalized connectomes, $\sim 59\%$ of edges with significantly increasing tract counts were between cortical regions versus only $\sim 5\%$ of edges between deep gray matter regions (with the remainder of edges between the cortex and the deep gray matter). This comparison may be unfair, since 76 of 90 atlas regions are within the cerebral cortex. However, even when weighted by the model slopes, (i.e., the relative tract-count increases at each edge,) these cortico-cortical connections represent $\sim 72\%$ of the total rate of tract count increases in the infant brains.

The differences in brain volumes between different subjects in each scan makes it difficult to compare tract lengths between scans in a meaningful way. However, using the centroids of the UNC atlas regions we can determine the relative distances between region pairs in a common space. The UNC atlas is 115 mm long along the coronal axis with centroids an average of 48.6±16.2 mm apart. The two most distant regions are 98 mm apart. Using these Euclidean distances as an indicator of relative tract lengths between regions we can explore the development of spatially local versus spatially distant connections. Of edges with significantly increasing weights in the normalized tract-count connectomes, $\sim$ 50% were between region pairs less than 30 mm apart. However, some long-range connections were also found to be developing, with 12 significantly increasing region-pairs greater than 60 mm apart and 2 greater than 70 mm apart.

Finally, the same statistical analysis was carried out on the mean-FA connectomes. For these connectomes, 840 region-pairs showed significantly increasing FA whereas 169 showed significantly decreasing FA. While there are more region-pair connections with significantly increasing FA than those with increasing tract count, 75% of region-pairs with significantly increasing tract count also show significantly increasing FA. Furthermore, 82% of region-pair connections with relatively increasing tract counts (observed in the normalized connectomes) also show significantly increasing FA. (At random, we only expect only 21±1.6% and 21±2.8% of connections to overlap in these two cases, respectively.) Despite appearing to have very different configurations in Figure 3.4, the longitudinal trends in the tract count and FA connectomes are certainly related. We discuss possible reasons for discrepancies in Section 3.4.

To visualize the changes in brain network topology across time, we used our linear mixed-effect models for each connection to estimate tract counts at four ages spanning our cohort's age range. The resulting predicted brain networks are shown in Figure 3.5. These four evaluated brain networks show connections gaining and losing relative connection strength over time. Certain connections in particular show high rates of development, particularly around the lingual, fusiform, and parahippocampal gyri (see Figure 3.8 for locations of these regions). Others, like the connection between the medial parts of the left and right superior frontal gyri show a rapid decrease in relative strength, implying that the number of tracts between these regions are not increasing as quickly as between a typical region-pair.

3.3.3 Network Measures Versus Age

In order to summarize how the organization of the structural connectome develops over time, we plot measures of structural connectivity, segregation, integration and small-worldness versus age in Figure 3.6. We used the Brain Connectivity Toolbox to compute all reported network measures [234].

For the unnormalized network measures, MWD was found to have a significant positive correlation with age, while CC and CPL are significantly correlated with age, positively

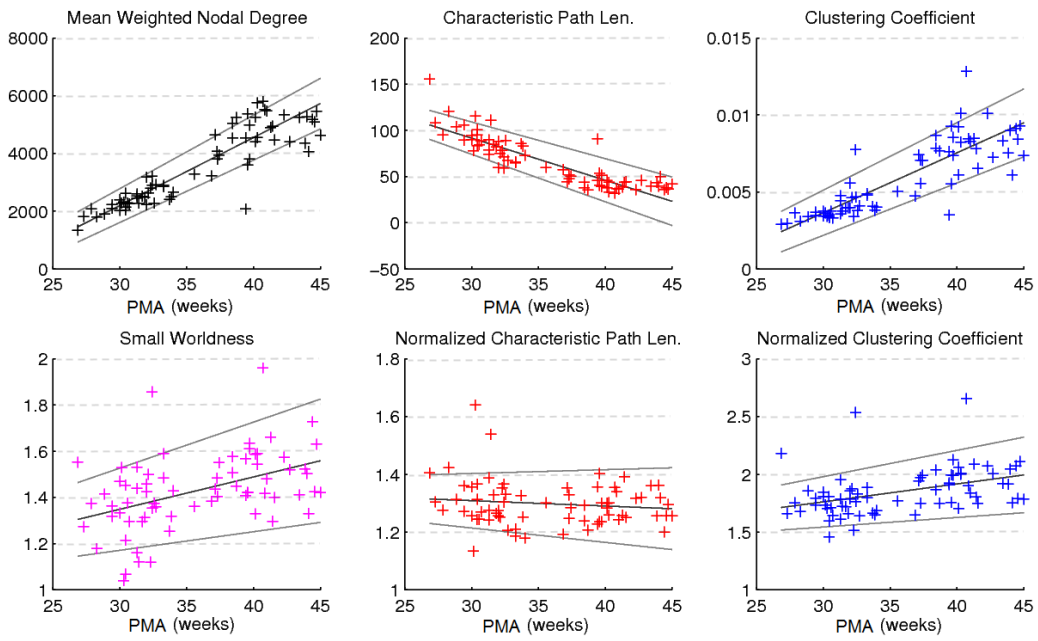

Figure 3.6: Mean degree and other network measures plotted against age at time of scan. Linear mixed-effects model fit to each measure (black) and are displayed with the 95% confidence interval for the parameters of each linear model (gray). Note the significant increases in normalized clustering coefficient and small-worldness across this age range.

and negatively, respectively. These results confirm the hypothesis that most region-pair connections are gaining tracts over time, likely due to brain volume increases and white matter maturation.

In the case of the normalized network measures, we found that normalized CPL showed no significant correlation with age and stays roughly constant. This implies that as the brain develops, the topological distance between any two regions is remaining constant on average. However, normalized CC is significantly positively correlated with age, implying that regions are becoming segregated into distinct clusters. The combination of these results suggest that over this age range, the preterm infant brain is organizing itself into clustered sub-networks while maintaining its larger scale, cross-network connections. The SW measure, having been computed by dividing normalized CC by normalized CPL, (which, we note, also makes it a normalized measure,) shows a similar significant positive correlation. This implies that the structure of the infant connectome is becoming more small-world, independent of total number of tracts.

We summarize SW slope CIs across tract count, normalized tract count and mean-FA connectomes in Table 3.3. Note that SW slope CIs for normalized and unnormalized tract count connectomes are very similar. This validates our assumption that the SW measure is already independent of overall edge-weight magnitude. Also note that the mean-FA

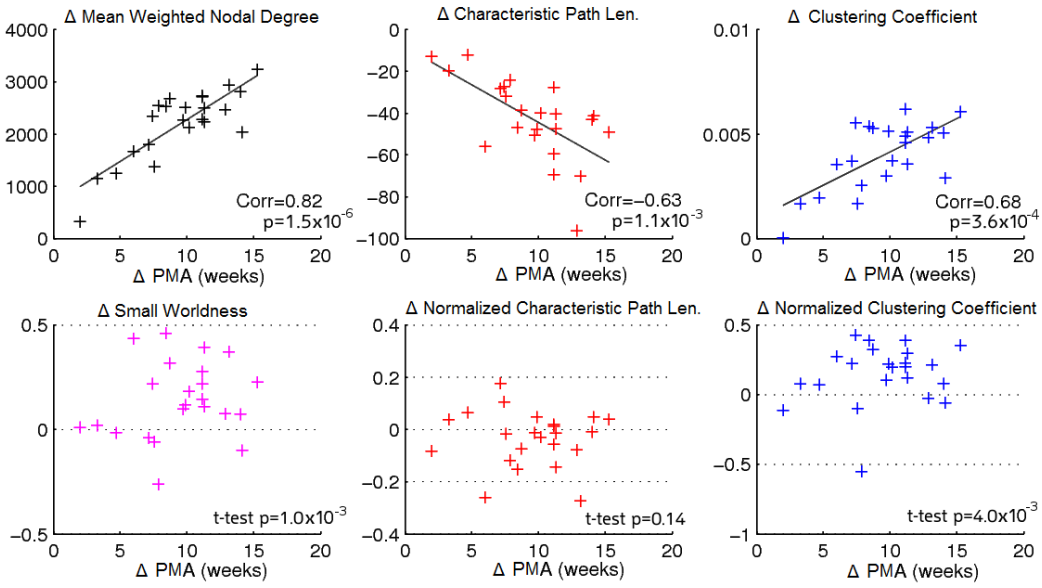

Figure 3.7: Change in network measures between an infant's first and second scan versus number of weeks between scans. Pearson's correlation reported for changes in unnormalized measures (top row). For each normalized measure (bottom row), a t-test is run to find if the mean connectome change between scans is significantly different from zero. p-values are reported for each measure. Note that normalized clustering coefficient and small-worldness show a significant increase between scans.

connectomes lack a clear trend of increasing SW with age. We discuss potential reasons for the appearance of this result in Section 3.4.

3.3.4 Intra-Subject Network Changes

With 23 subjects scanned twice, we can examine intra-subject longitudinal trends of these structural connectomes. In Figure 3.7, we show the differences in the network measure values versus the time between scans.

For the unnormalized measures, all subjects displayed increasing MWD and CC while CPL decreased between scans. This result is what we would expect based on the inter-subject trends in the previous experiment. Further, changes in MWD and CC are significantly ($p < 0.01$) positively correlated with age, while changes in CPL are significantly negatively correlated with age. Thus, longer durations of time linearly relate to larger changes in the measures.

For the normalized measures, CPL values are scattered above and below zero. This result suggests that, compared to random networks, about as many subjects became relatively less efficient as those that became relatively more efficient. By applying a t-test, we found that the mean normalized CPL value was not significantly different from zero ($p = 0.1333$). Normalized CC, however, increased between scans in most subjects which,

Table 3.3: Summary of small-worldness (SW) trends in tract count, normalized tract-count and mean-FA connectomes. The middle three columns report slope estimates and 95% CI bounds for linear mixed-effects model fit across scans. The right-most column reports t-test p values for the hypothesis that SW does not change over time for scans from the same subjects. Note that SW increases significantly for the tract-count connectomes but not the mean-FA connectome.

	Mixed Model Slope CI			Subject Δ
Connectome	Lower	Est.	Upper	t-test p val.
Tract-Count	0.0081	0.0140	0.0199	0.0017
Norm. Count	0.0070	0.0133	0.0196	0.0032
Mean-FA	-0.0044	-0.0007	0.0030	0.6498

in turn, caused SW to increase. Using t-tests, the mean values of both measures were found to be significantly greater than zero ($p = 0.0056$ and $p = 0.0017$ respectively). Both this experiment and the previous one strongly suggest that SW increases with PMA in the tract-count connectomes of preterm infants (Table 3.3).

3.4 Discussion

In the first experiment of this study, we found that our dataset of preterm infants, represented as two SNR group-connectomes of younger and older scans, showed similar responses to thresholding as a group of similarly aged infants born at term and studied by Yap et al. [309] (Section 3.3.1). The comparisons between the infant groups and random and lattice networks resulted in the predicted outcomes and all three groups of infant connectomes exhibited a balance of network integration and segregation. Despite these similarities, Yap et al. reported, at certain network costs, lower ML and GE in the term group than we find in the preterm group. These results seem to disagree with our later analysis on how network integration and segregation develop over time. However, it is likely the case that differences between connectome pipelines can explain this discrepancy.

One important difference between the two connectome pipelines is the way in which tract-count connectomes are constructed. In Yap et al.'s work, the weight of each tract-count connectome edge represented the number of tracts *passing through* a pair of atlas regions. In contrast, we only count a tract based on what regions its end-points are in. So, for the same scan, Yap et al.'s method will report strictly larger edge weights, especially for regions that are intermediaries on long-range fiber tracts. This bias towards intermediate regions means that small sections of these long-range connections will dominate the topology of the backbone network more so than their full, long-range connection.

This counting bias explains why, at low network costs, GE was higher in the age-matched preterm group compared to the group born at term. In the connectomes computed by Yap et al. , long-range connections, which increase GE, will be counted less than their

corresponding sub-connections and thus will be less likely to be included in the backbone network. Further, we see that for higher network costs, these long-range connections are preserved in the backbone network and the difference in GE vanishes.

We also found that at most network costs, the ML computed by Yap et al. was lower than what we computed in the age-matched preterm group. We expect this is also due to the counting bias mentioned above. The edges that receive larger weights by Yap et al.'s method are likely to have low weights in our connectomes, since these intermediate connections are unlikely to have a large number of similar tracts that terminate in the two corresponding regions. These intermediate connections are also likely to connect regions along long-range fiber bundles, thereby reducing network segregation and ML.

Given that differences in connectome construction can account for discrepancies in network measures between group backbone networks in the two studies, the degree of network segregation and integration in infants born preterm versus those born at term may, in fact, be similar. It has been well-established that preterm birth is associated with reduced FA and increased mean diffusivity, suggesting reduced structural connectivity [8, 17, 128]. Our results may complement those findings by further showing that this reduced connectivity does not affect the overall organization of the connectome, or that the differences in organization are subtle enough that they cannot be discerned from measures of the backbone network. The second of these two explanations seems more likely as we did not see a difference in ML between our younger and older preterm groups despite seeing a trend of increasing segregation with time in Section 3.3.3.

An examination of individual region-pair connections (Section 3.3.2) revealed higher than average numbers of developing connections in the occipital and frontal lobes. Early development of neurons and connecting axonal fibers in the occipital lobe, prior to those in the frontal lobe, has been reported in the micro-scale brain anatomy literature [129, 271]. At the same time, Pandit et al. observed more connections developing with age between frontal lobe structures within 47 to 134 weeks post-conception [205]. It is possible that within the age range of this study, we are seeing both effects: the greater development in the occipital lobe could be occurring earlier in our cohort's age range, with greater development in the frontal lobe occurring later on. This hypothesis would confirm results reported by Takahashi et al., who found in postmortem infants, between 17 and 40 weeks post-conception, that emergence of connectivity moved generally from posterodorsal to anteroventral [266]. Unfortunately, the linear modeling we performed did not allow us to see these distinct periods of development separately. Our future work will look at exploring this hypothesis even further, either by fitting a higher order model or by examining smaller age ranges individually.

It is likely that the relative increases in tract count seen between certain region-pairs is primarily due oligodendrocyte maturation [83]. Myelin sheath maturation begins as early as 30 weeks but does not occur simultaneously throughout the brain, instead propagating

up the corticospinal tracts and outwards, affecting major fiber bundles first [180]. We therefore expect then, in our cohort, that the majority of increased connectivity is due to oligodendrocyte maturation and not myelination.

We examined both tract-count connectomes and mean-FA connectomes and found that there was significant overlap between developing edges between the two (Section 3.3.2). This results is expected since higher FA suggests the presence of more cell structure connecting two regions [28]. However, high mean FA does not always imply higher oligodendrocyte maturation or that more neuronal fibers are connecting two regions. Regions of crossing fibers may have high fiber density but will present a low FA value in a standard DTI. That said, we expect that the effects of crossing fibers are less likely to present in this study given the early ages of the infants.

A notable difference between tract-count and mean-FA connectomes is that the mean-FA edge weights are not as sensitive to relative bundle thickness. While certain fiber tracts may be thicker than others, they don't have to exhibit higher mean FA. For this reason, and those mentioned above, we conclude that there is value in examining both types of connectomes, especially when examining early infant development when a variety of complex processes may be influencing FA values.

The rapid development of cortico-cortical tracts found in our analysis of the normalized tract-count connectomes matches what is known in the literature. Long-range fibers between cortical regions develop across the corpus callosum between 33-35 weeks PCA and short-range cortico-cortical fibers begin developing after that [153]. Again, due to our linear model, here we only observe the combined effect of these two stages of development.

Examining the distance between regions of developing connections, we found that the majority of region-pairs exhibiting significantly increasing tract-counts were between spatially local regions ($<$30 mm). This bias towards shorter tracts could be because longer connections are harder to track [176, 224]. However, we did find some long-range connections between regions greater than 70 mm apart. This finding of a small number of long-range connections developing in the preterm infant brain is supported by the results of Takahashi et al., who also found such long range connections developing between 17 and 40 weeks gestation [266].

In Section 3.3.3, it was shown that for tract-count connectomes, SW consistently increases with age across our age range. This result is consistent with results from Tymofiyeva et al. who found higher small-worldness in infants scanned shortly after normal term birth than in a group of preterm infants scanned at an average of $\sim$ 35 weeks after conception [275]. In their work, as here, rise in small-worldness was due to increased normalized CC values and stable normalized CPL. van den Heuvel et al. also examined small-worldness in structural connectomes of preterm infants, scanned at 30 and 40 weeks PMA. Despite evidence of WMI and other pathologies in their cohort, our small-worldness results strongly agree with their findings of mean SW values of $\sim$ 1.3 at 30 weeks and $\sim$ 1.5 at 40 weeks [280].

While other studies have shown evidence of small-world brain networks at different ages, the fine temporal sampling of the dataset in our study exposed the clear trend of increasing SW and allowed detection of a significant positive correlation of SW with age [20]. This result suggests that the human connectome is less small-world at an early stage in development, and perhaps at a very young age, not small-world at all.

In contrast, it was found that for mean-FA connectomes, SW was greater than 1 but did not increase with PMA. While many of the region-pair connections which exhibited a consistent increase in tract count also exhibited an increase in mean FA, there were a number of connections which exhibited an increase in FA but not in tract count (Section 3.3.2). Thus, the mean-FA connectomes seemed to have extra connections, not present in the tract-count connectomes, that gained edge weights over time. It is likely that these extra connections with increasing edge weights are responsible for preventing increasing SW by inhibiting network segregation since the two connectome types seem to otherwise change similarly with age.

We also saw that not every subject's tract-count connectome increased in SW across age (Section 3.3.4). This implies that either there is some error in our pipeline introducing a certain inaccuracy in connection weights, or that individual subjects develop somewhat distinctly and that the trend in increased small-worldness with age is not true for every person across every time-increment of early development. There is increased potential for this second option, given that the subjects studied here are all born preterm and, while were screened to be healthy, are likely affected in some way by the early age of birth. Nevertheless, it is impossible not to introduce some error in the registration, segmentation and tractography processes and thus it is probably the case that both explanations of decreased SW are, in part, responsible. Moving forward, there are certainly aspects of the connectome pipeline which can and should be improved, particularly through the use of more advanced tractography [35, 42, 108, 107] and registration techniques [34].

One subject in particular exhibited greatly decreased CC (<-0.5) and was an outlier in this respect. This subject also exhibited a relatively large increase in total tract count given the time between scans (i.e., it is well above the linear regression on the delta MWD plot in Figure 3.7). This large increase in number of tracts enables a large change in network topology between scans. However, the reason why, in this subject in particular, these new tracts are distributed across the network in a way that is causing reduced clustering is not clear since no other aspects of this subject, for which we have data, are notable.

One area not explored in this work is the effect of the birth age on the brain network properties examined here. While the cohort in this study was screened thoroughly to be normative, we still expect known effects of preterm birth, like reduced overall FA and increased MD, to have some effect on brain network connectivity [8, 17, 128]. A future direction is to examine the relationships between birth age, age at time of scan and connectome topology.

Also, it has been noted in the literature that when performing inference on multiple edges in a graph, like we did in Section 3.3.2, the topology of the graph can be used to more appropriately control the family-wise error rate (FWER) [314]. Network based statistics (NBS) by Zalesky et al. is a standard approach for leveraging network structure when controlling FWER but assumes a generalized linear model at each edge. Unfortunately, linear mixed-effects models are not within the class of generalized linear models. Future work may include extending NBS to linear mixed-effects models in order to extract more accurate inferences from datasets where each subject is scanned multiple times.

3.5 Conclusions

In this chapter, we analyzed the topology and longitudinal change in the structural connectomes of a cohort of young normative preterm neonates. To do this we constructed a pipeline to count the number of white matter tracts, and compute the mean FA, between pairs of anatomically defined regions in each scan. Individual connections and the high-level topology of the resulting brain networks were analyzed as a group and individually across age. Certain region-pair connections showed a particularly high rate of tract growth compared to others, particularly in the frontal and occipital lobes. However, discrepancies between tract-count and mean-FA connectome longitudinal trends exposed the importance of examining both measures of structural brain connectivity. Finally, we found that the preterm group in this study consistently exhibited high network integration and segregation and, most interestingly, significantly increasing small-worldness across age.

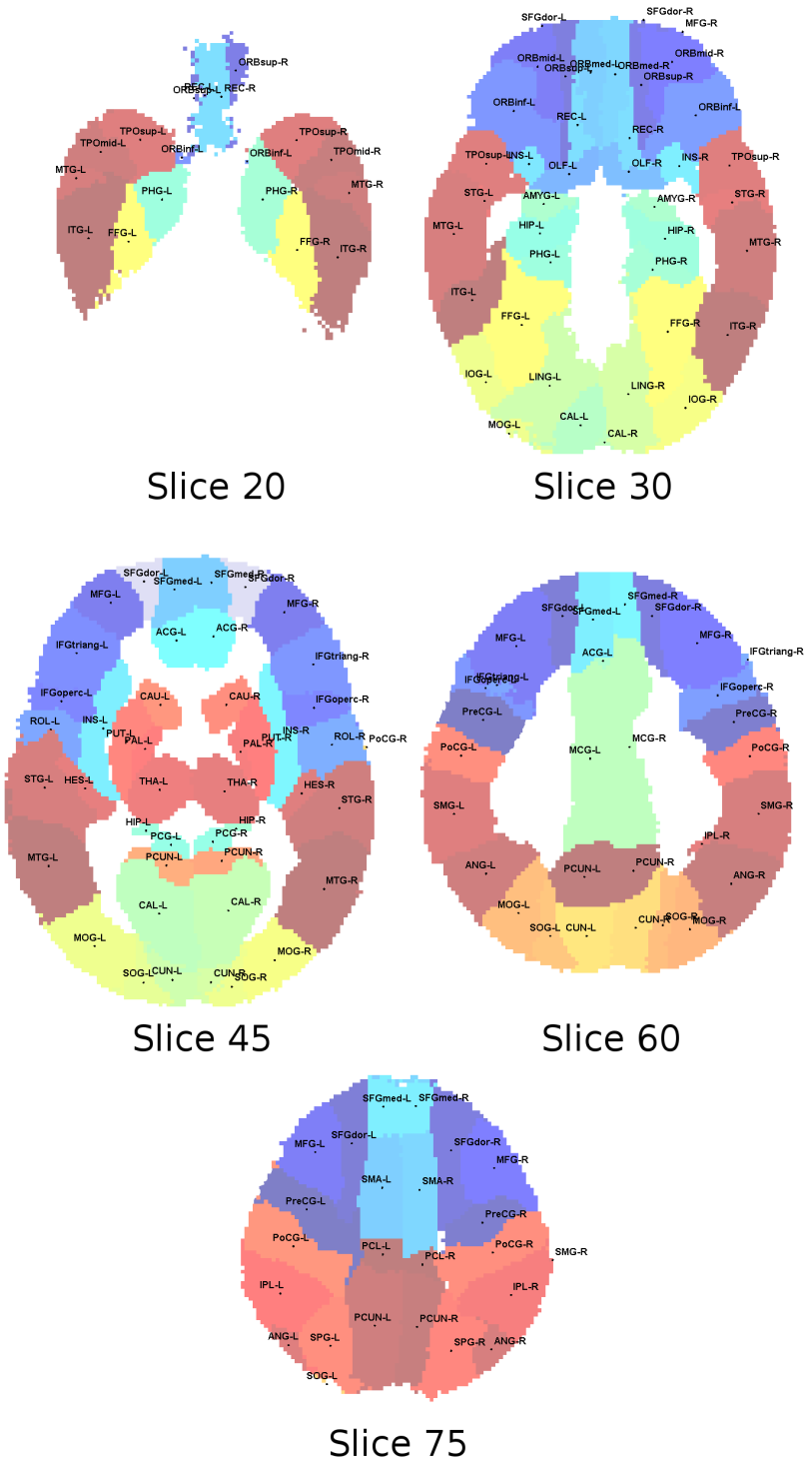

Figure 3.8: Axial slices of the UNC atlas, numbered from inferior to superior. Each atlas region is coloured uniquely and labelled using a label from Table 3.4.

Table 3.4: List of anatomical region names and abbreviations in UNC brain region atlas.

ID	Region	Abrv.	ID	Region	Abrv.
1	Precentral gyrus left	PreCG-L	46	Cuneus right	CUN-R
2	Precentral gyrus right	PreCG-R	47	Lingual gyrus left	LING-L
3	Superior frontal gyrus (dorsal) left	SFGdor-L	48	Lingual gyrus right	LING-R
4	Superior frontal gyrus (dorsal) right	SFGdor-R	49	Superior occipital gyrus left	SOG-L
5	Orbitofrontal cortex (superior) left	ORBsup-L	50	Superior occipital gyrus right	SOG-R
6	Orbitofrontal cortex (superior) right	ORBsup-R	51	Middle occipital gyrus left	MOG-L
7	Middle frontal gyrus left	MFG-L	52	Middle occipital gyrus right	MOG-R
8	Middle frontal gyrus right	MFG-R	53	Inferior occipital gyrus left	IOG-L
9	Orbitofrontal cortex (middle) left	ORBmid-L	54	Inferior occipital gyrus right	IOG-R
10	Orbitofrontal cortex (middle) right	ORBmid-R	55	Fusiform gyrus left	FFG-L
11	Inf. frontal gyrus (opercular) left	IFGoperc-L	56	Fusiform gyrus right	FFG-R
12	Inf. frontal gyrus (opercular) right	IFGoperc-R	57	Postcentral gyrus left	PoCG-L
13	Inf. frontal gyrus (triangular) left	IFGtriang-L	58	Postcentral gyrus right	PoCG-R
14	Inf. frontal gyrus (triangular) right	IFGtriang-R	59	Superior parietal gyrus left	SPG-L
15	Orbitofrontal cortex (inferior) left	ORBinf-L	60	Superior parietal gyrus right	SPG-R
16	Orbitofrontal cortex (inferior) right	ORBinf-R	61	Inferior parietal lobule left	IPL-L
17	Rolandic operculum left	ROL-L	62	Inferior parietal lobule right	IPL-R
18	Rolandic operculum right	ROL-R	63	Supramarginal gyrus left	SMG-L
19	Supplementary motor area left	SMA-L	64	Supramarginal gyrus right	SMG-R
20	Supplementary motor area right	SMA-R	65	Angular gyrus left	ANG-L
21	Olfactory left	OLF-L	66	Angular gyrus right	ANG-R
22	Olfactory right	OLF-R	67	Precuneus left	PCUN-L
23	Superior frontal gyrus (medial) left	SFGmed-L	68	Precunesu right	PCUN-R
24	Superior frontal gyrus (medial) right	SFGmed-R	69	Precentral lobule left	PCL-L
25	Orbitofrontal cortex (medial) left	ORBmed-L	70	Precentral lobule right	PCL-R
26	Orbitofrontal cortex (medial) right	ORBmed-R	71	Caudate left	CAU-L
27	Rectus gyrus left	REC-L	72	Caudate right	CAU-R
28	Recuts gyrus right	REC-R	73	Putamen left	PUT-L
29	Insula left	INS-L	74	Putamen right	PUT-R
30	Insula right	INS-R	75	Pallidum left	PAL-L
31	Anterior cingulate gyrus left	ACG-L	76	Pallidum right	PAL-R
32	Anterior cingulate gyrus right	ACG-R	77	Thalamus left	THA-L
33	Middle cingulate gyrus left	MCG-L	78	Thalamus right	THA-R
34	Middle cingulate gyrus right	MCG-R	79	Heschl gyrus left	HES-L
35	Posterior cingulate gyrus left	PCG-L	80	Heschl gyrus right	HES-R
36	Posterior cingulate gyrus right	PCG-R	81	Superior temporal gyrus left	STG-L
37	Hippocampus left	HIP-L	82	Superior temporal gyrus right	STG-R
38	Hippocampus right	HIP-R	83	Temporal pole (sup.) left	TPOsup-L
39	ParaHippocampal gyrus left	PHG-L	84	Temporal pole (sup.) right	TPOsup-R
40	ParaHippocampal gyrus right	PHG-R	85	Middle temporal gyrus left	MTG-L
41	Amygdala left	AMYG-L	86	Middle temporal gyrus right	MTG-R
42	Amygdala right	AMYG-R	87	Temporal pole (middle) left	TPOmid-L
43	Calcarine cortex left	CAL-L	88	Temporal pole (middle) right	TPOmid-R
44	Calcarine cortex right	CAL-R	89	Inferior temporal gyrus left	ITG-L
45	Cuneus left	CUN-L	90	Inferior temporal gyrus right	ITG-R

Chapter 4

Prediction of Motor Function in Very Preterm Infants using Connectome Features and Local Synthetic Instances

4.1 Introduction

Every year, an estimated 2.2 million babies worldwide are born very preterm (born at 32 weeks gestation or younger) [52]. Very preterm birth puts newborns at a high risk of long-term motor dysfunction (e.g. cerebral palsy), which places significant burdens on the child, the family and the community [14, 182]. Early detection of motor dysfunction could enable more rapid identification of infants who would benefit from rehabilitative interventions. While motor outcomes can be assessed in preterm-born infants at 18 months of age using the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III) [26], we desire earlier identification of infants at risk in order to inform care and ongoing monitoring.

Certain brain pathologies, such as white matter injury (WMI) and intraventricular hemorrhaging (IVH) are detectable in a structural MRI scan of an infant's brain. It is also known that some of these pathologies, as well as findings from more advanced MR methods such as diffusion tensor imaging (DTI), are associated with later neurodevelopmental outcomes [18, 59]. However, most studies to date have focused on group differences or correlations between specific DTI measures and motor outcomes. For example, Chau et al. recently reported that the trajectory of brain maturation from early in life to term-equivalent age, using region-of-interest-based DTI measures, was associated with neurodevelopmental outcomes [59]. Ball et al. examined the relationship between connectivity in the thalamocortical connectome in preterm infants and Bayley-III scores [18]. They found that the strength of certain connections were significantly correlated with outcomes.

In comparison to finding correlations, prediction is a harder problem. In order to predict accurately, the complete set of factors contributing to outcome must be modelled. Prediction of motor function from brain structure at birth is thus very challenging due to the large number of confounding factors affecting brain development, including potentially unknown genetic and environmental factors. It is especially difficult in young infants due to the combination of limited image resolution and small brain sizes, artifacts due to motion and rapid structural change across a small temporal window (discussed in Chapter 3) [44]. Furthermore, datasets are often class-imbalanced, containing fewer abnormal cases. This last issue is of particular importance to prediction since many prediction models are highly sensitive to imbalanced training data [132].

Strategies exist in the literature to alleviate this class imbalance problem by augmenting the training set. For example, the synthetic minority over-sampling technique (SMOTE) finds K nearest neighbours to each training instance and interpolates new instances randomly along lines connecting neighbours [60]. Another method is to sample synthetic instances from an approximate distribution of positive instances, learned using kernel density estimation (KDE). Alternatively, the dominant modes of variation for positive training instances can be learned using principal component analysis (PCA), then sampled to generate new instances. The method that generates the most realistic synthetic instances (i.e., those which improve prediction accuracy the most) is likely application dependent, making class imbalance challenging for prediction problems.

Despite the challenges, prediction of long-term motor dysfunction within the first few weeks of birth remains a desirable goal as it would enable better treatment planning and a more informed assessment of patient outcome. Recently, Ziv et al. used connectome based features from term infants scanned at 6 months of age to predict general neurological outcome at 12 months [323]. Here we set the goal of predicting motor outcome at 18 months from scans taken within the first weeks after birth. In our study, the earlier post-menstrual age (PMA) at scan and the larger temporal gap between scan and outcome makes our task even more challenging. Similar to Ziv et al. we use dimensionality-reduced connectome features and a support vector machine (SVM) classifier to achieve this goal. However, we also introduce a novel method for generating synthetic training samples designed to tackle the specific challenges in our data. Our training data contains only a small number of positive instances existing in a high-dimensional space, likely constrained to a complex manifold. In order to only generate realistic data, we propose a data interpolation technique that generates synthetic instances which are restricted to be more similar to individual known instances.

In this chapter, we show for the first time that 18-month motor outcomes, assessed with the Bayley-III, can be predicted from an MRI taken in the first weeks of a preterm neonate's life. Our method achieves an accuracy of $> 70\%$, which establishes a baseline level of accuracy for this important yet very challenging task. Further, we demonstrate

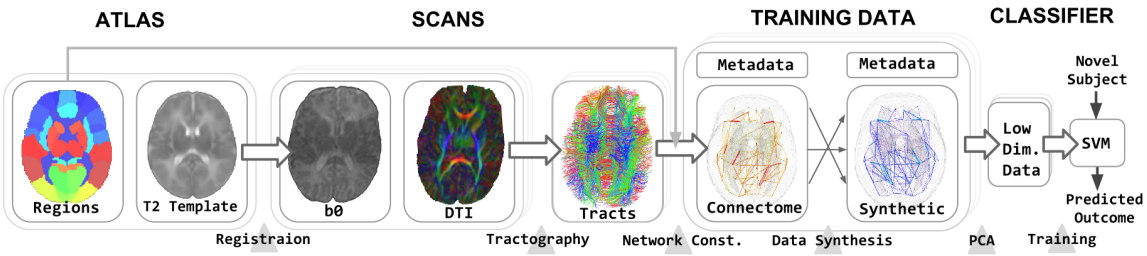

Figure 4.1: High level schematic representation of connectome and training pipeline.

that data augmentation can reliably improve prediction scores and that our novel method for generating local synthetic instances outperforms competing methods.

4.2 Methods

In Fig. 4.1, we present a schematic diagram of our supervised learning framework for predicting motor function at 18 months from DTI scans acquired in the first weeks of life. Details for this pipeline are presented below.

4.2.1 Dataset

The cohort used in this study was a group of 115 preterm infants born between 24 and 32 weeks PMA (from the same dataset that was examined in Chapter 3) . Neonates underwent a brain MRI between 27 and 45 weeks PMA on a Siemens (Erlangen, Germany) 1.5T Avanto using VB 13A software. Each scan was a multi-slice 2D axial EPI diffusion MR acquisition (TR 4900 ms; TE 104 ms; FOV 160 mm; slice thickness 3 mm; no gap) with 3 averages of 12 non-colinear gradient directions, with an isotropic in-plane resolution of 0.625 mm. Two such acquisitions were performed: one at $b = 600$ s/mm^2 and one at $b = 700$ s/mm^2. The combined diffusion weighted image set was preprocessed using the FSL Diffusion Toolbox (FDT) pipeline[1] and tensors were fit using RESTORE [57]. Nearly half of the subjects (53) were scanned twice for a total of 168 diffusion tensor images.

At 18 months, each subject was evaluated using the Bayley-III test which produces three composite scores of cognitive, language and motor skills [26]. The scores are normalized with mean of 100 and standard deviation of 15; we considered adverse motor outcomes as scores below 85 (i.e., -1 std.). In our cohort, 146 scans were of infants with normal motor function at 18 months, and 22 scans were of infants with abnormal motor function.

4.2.2 Connectome Construction and Analysis

Each DTI was segmented using a neonatal atlas of 90 brain regions from the IDEA group at University of North Carolina (UNC) School of Medicine, Chapel Hill [247]. The atlas'

[1] http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FDT

associated T2 template was aligned to the b0 image of each DTI scan in order to segment each brain. Note that this method is an improvment (i.e., achieved better registration results) over the method used in Chapter 3, in which the T1 template was aligned to the T1 scan of each infant. Alignment was performed as a rigid registration using FMRIB's Linear Image Registration Tool (FLIRT) [133] followed by a deformable registration using the MATLAB Image Registration Toolbox (MIRT)[2]. Full-brain tractography was performed on each scan using TrackVis [292]. Each connectome was constructed by grouping tracts according to their endpoint regions. Following our work in Chapter 3, we constructed three connectome types for each scan, including a mean-FA weighted connectome, a tract-count connectome and a normalized tract-count connectome. Network measures summarizing the connectome's topological properties were computed for each connectome type. Ten individual network measures were extracted: 1) mean nodal degree, 2) transitivity, 3) global efficiency, 4,5) raw and normalized modularity, 6,7) raw and normalized clustering coefficient, 8,9) raw and normalized characteristic path length and 10) small-worldness. Each measure was computed for mean-FA, tract-count and normalized tract-count connectomes giving a total of $30 = 10 \times 3$ measures. For a comprehensive summary of network measures and their meanings, see [234] or our discussion of them in Chapter 2).

4.2.3 Classification

For each DTI scan, metadata and connectome features were extracted. See Table 4.1 for a full list of feature types. Metadata features included gender, age-at-scan and age-at-birth. Age-at-scan is included because we expect it to be an important co-variate given the rapid development of the brain across the age range of our cohort. Connectome features included mean FA across each of the $4005 = 90 \times 89/2$ edges and other high-level network features as described above. WMI severity (graded $[0, 3]$) and IVH severity (graded $[0, 4]$), assessed from a T1 MRI by an experienced neuroradiologist, were included as additional features. All features were concatenated into one feature vector.

Feature vectors from the training set were processed using PCA to extract, at most, the top m modes of variation. At test time, instances were projected into this PCA space. A linear dimensionality reduction method was used instead of, for instance, a kernel based method in order to prevent over-fitting to our sparsely sampled, high-dimensional training set. An SVM classifier was then trained on the instances in this reduced space.

4.2.4 Local Synthetic Instances (LSI)

Due to the limited number of cases in our cohort that show motor dysfunction (22 out of 168), our training set has a severe class imbalance ($>$ 6:1). We compensate for this by both replicating existing positive instances, $\mathbf{t}_i$ (i.e., feature vectors), and by generating

[2]https://sites.google.com/site/myronenko/research/mirt

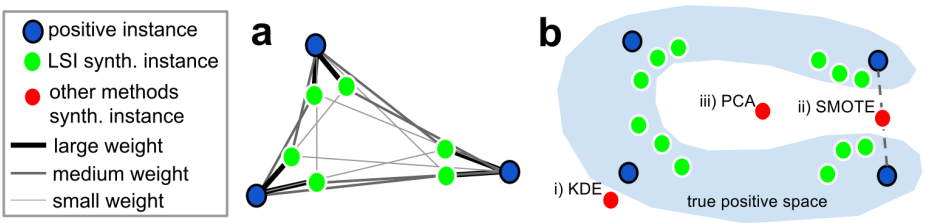

Figure 4.2: a) Schematic representation of LSI weights for 6 synthesized instances. b) Possible failure cases for other data augmentation methods.

new synthetic positive instances, $\mathbf{s}_j$. New instances are generated by interpolating instance feature vectors within the training set of N positive instances.

We do not know the true distribution of positive instances but we assume that it is locally smooth. Under this assumption, we can generate reasonable synthetic instances as long as they are near existing positive training instances. To this end, we seek an interpolant that satisfies two conditions: (i) gives the majority of weight to one training instance and (ii) gives some non-zero weight to other instances to ensure some variation. We achieve this by randomly assigning elements from a normalized p-series as weights to each instance. Let P^j be a random $N \times N$ permutation matrix and $\mathbf{r}^j = P^j[1, 2, ..., N]^T$. Then, the elements, r_i^j, of $\mathbf{r}^j$ are numbers from 1 to N and the vector $\mathbf{r}^j$ is randomly ordered and,

$$\mathbf{s}_j = \sum_{i=1}^{N} w(r_i^j)\mathbf{t}_i \qquad \text{where} \qquad w(i) = \frac{i^{-p}}{\sum_{n=1}^{N} n^{-p}}. \tag{4.1}$$

Thus, for the jth synthetic instance, $\mathbf{s}_j$, $\mathbf{r}^j$ randomly assigns weights to training instances without replacement. Note that $\sum_{i=1}^{N} w(i) = 1$. Also, note that $\forall i > 1, w(1) \geq 2^p w(i)$ and for $p = 2$ (and in fact, any $p > \ 1.7287$), $w(1) > \sum_{i=2}^{N} w(i)$ for any N. In other words, the weight on one instance dominates the others and each synthetic instance will be generated local to one training instance as desired. Fig. 4.2a is a schematic example of synthetic instances generated from three real instances.

The proposed LSI method has desirable properties compared to other existing methods in the case of high-dimensional data and small number of training instances, as we have here. For instance, while samples from a KDE-estimated distribution are likely to be near training instances, a fixed kernel is used, resulting in new instances that may not vary away from training instances in a realistic way. Instead, LSI encourages new instances to be near training instances but requires variation to be towards known training data. Furthermore, unlike KDE, LSI does not allow extrapolation. This is adventageous since, given a small training size, extrapolation is unlikely to yield realistic samples (Fig. 4.2b(i)). SMOTE generates instances local to a subset of known instances but fails where the manifold of positive instances is too sparsely sampled to be approximated using nearest neighbours (Fig. 4.2b(ii)). Sampling from PCA modes ensures that new instances only vary along pri-

Table 4.1: Name and description of each feature type.

Name	Description (# features)	Name	Description (# features)
Meta	Birth age, scan age and gender (3)	MRI	WMI and IVH scores at birth (2)
Edge	Connectome edge FA values (4005)	Network	High-level network measures (30)

Table 4.2: Mean training (Tr) and test accuracy, sensitivity (Sn) and specificity (Sp) for 1000 rounds of leave-2-out cross validation. Tests marked * are plotted in Fig 4.3. Best test accuracy is in **bold**.

	No Synth. Data				With Synth. Data			
Feature Types Used	Tr	Test	Sn	Sp	Tr	Test	Sn	Sp
Meta, MRI	69.0	64.2*	0.46	0.83	69.9	64.4	0.47	0.82
Meta, Edge	85.0	58.5	0.33	0.84	86.1	58.5	0.34	0.83
Meta, Edge, Network	85.5	61.2	0.34	0.82	85.7	62.8	0.43	0.83
Meta, Edge, Network, MRI	84.4	64.7	0.51	0.79	85.0	64.7	0.49	0.81
Meta, Network, MRI	77.5	69.0*	0.56	0.82	79.3	**72.3***	0.66	0.79

mary modes of variation but may generate samples that are anatomically implausible since they may not be near any training instance (Fig. 4.2b(iii)). In comparison, the proposed method offers a balance between trusting local instances and using global information.

4.3 Results

We evaluated our method using a variety of feature subsets and instance synthesis methods. We assessed classification accuracy for different sets of feature types (Table 4.1) via 1000 rounds of cross-validation. In each round, one positive and one negative instance were left out for testing. Test subjects were omitted from training data and, since some subjects were scanned twice, the total training set varied between 164 and 166 scans. For these tests, we set $m = 20$, which explains $> 99\%$ of the variance and set $p = 2$. SVM misclassification penalty was set empirically to $C = 2^{-7}$ but test accuracy was relatively insensitive to this value, varying only about 1% for $2^{-8} < C < 2^{-6}$. Classes were balanced by weighting positive instances via replication. Table 4.2 shows prediction results for different sets of features and compares those tests run with synthetic training data versus those without. In the tests with synthetic data, the training set was first doubled by generating synthetic instances, then real positive training instances were replicated until the classes were balanced.

Note that the addition of synthetically-generated instances improved our highest test accuracy by 3%. Furthermore, as expected, the inclusion of high-level connectome features and MRI-based gradings (i.e. WMI & IVH grades) consistently boosted accuracy on average by about 3% each. Interestingly, *excluding* edge FA features causes the classification accuracy to *improve* when network measures and MRI-based scores are included. This is likely because the edge features are noisy and that much of the relevant structural information from the edge features is captured more succinctly in the network measures, WMI grade,

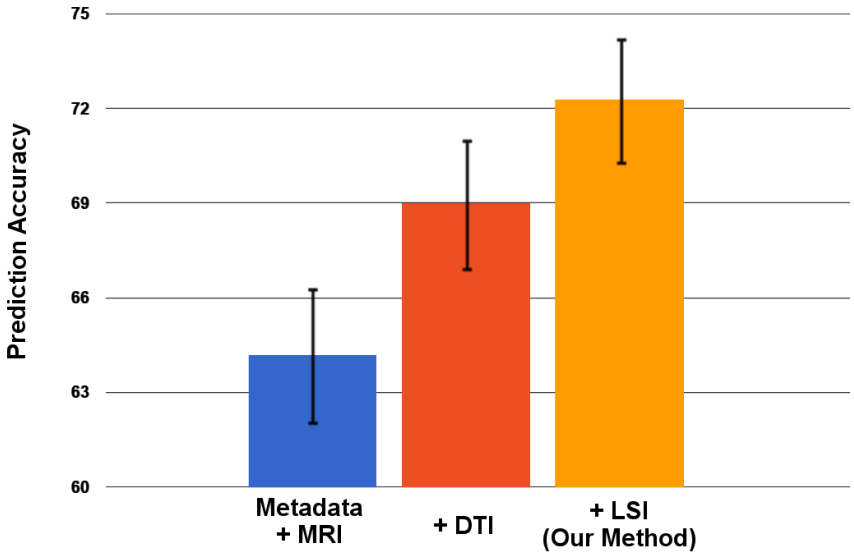

Figure 4.3: Test accuracy and 95% CIs for select tests from Table 4.2.

Table 4.3: Comparison between class balancing methods.

Method	Tr	Test	Sn	Sp	Method	Tr	Test	Sn	Sp
WLT	74.2	54.3	0.23	0.85	PCA	79.0	68.5	0.57	0.80
SSO	88.8	62.1	0.56	0.68	KDE	79.7	68.9	0.57	0.81
SDC	78.3	67.6	0.56	0.80	LSI	79.3	**72.3**	0.66	0.79

and IVH grade. It may also suggest that no single white-matter fibre bundle is strongly tied to motor outcome and instead that more widespread factors are at work. This finding is consistent with recent work in [14] which showed that the causes of poor motor outcome are multi-faceted. Note also that while metadata and MRI based information alone provides reasonable predictive power, it is clearly advantageous to include connectome information derived from DTI. This is shown in Fig. 4.3 with 95% confidence intervals (CI) for each result.

For the test using the subset of features that gave the highest test accuracy (Meta, Network, MRI), test accuracy was 69% across an infant's first scans and 74% across their second scans. This suggests that the later scans may be at some advantage in their ability to predict outcome. This is not surprising since the temporal gap between scan-time and assessment for these scans is smaller.

Again using the same features and ratio of synthetic to training instances that achieved the highest accuracy in the above experiment, we tested four different instance synthesis methods (Table 4.3 and with 95% confidence intervals in Fig. 4.4). For KDE, a Gaussian kernel was used and, after checking a range of scales, a standard deviation equal to the mean distance between samples was found to give the highest accuracy. SMOTE with different costs (SDC) is a variant of SMOTE, improved to better handle imbalanced classes [2]. For

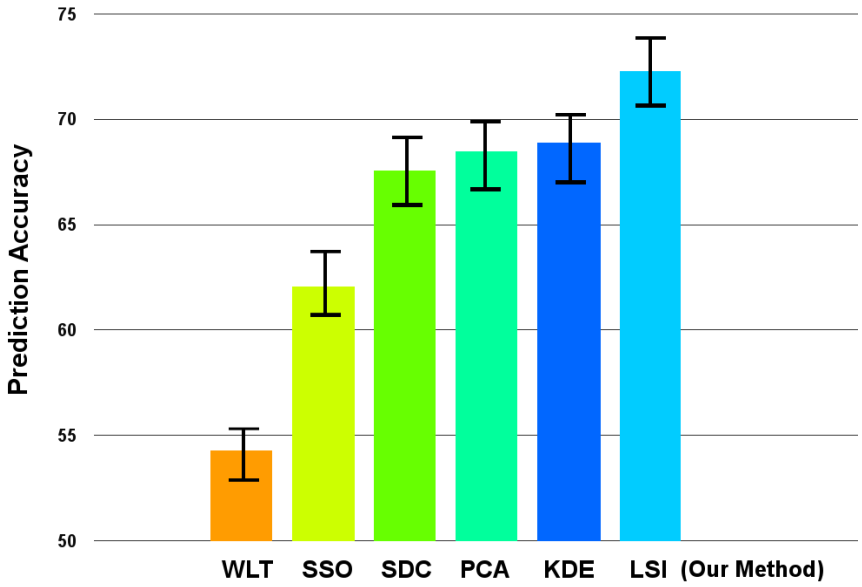

Figure 4.4: Test accuracies for different class balancing methods with 95% confidence intervals.

SDC, we set $K = 1$ as it was found to give highest accuracy over range $[1, 5]$. For PCA, once the variation modes were learned, new instances were generated by sampling the Gaussian distribution defined by the variation modes. To be consistent with the PCA step used for classification, again the top $m = 20$ modes of variation were used. We also compared these synthesis methods against subset sampling optimization (SSO) [304], a state-of-the-art undersampling method, and weighted Lagrangian twin SVM (WLT), a classifier designed to natively deal with class imbalance [243]. No replicated or synthetic instances were used with these methods. Note that since we use non-image data in our feature vector, data augmentation methods which modify the images directly are not applicable here, and so were not tested. Table 4.3 shows that our proposed method outperforms the competing methods. These findings agree with our hypothesis at the end of Section 4.2, that LSI is well suited to a small sample size from a complex manifold in high-dimensional space.

4.4 Conclusions

In this chapter, we predicted preterm infant motor outcomes at 18 months using structural connectome features from DTI scans taken at birth. In doing so, we established a baseline accuracy of over 70% for this challenging but important task. We also proposed a novel method to mitigate the effects of small positive sample sizes common to normal/abnormal datasets. Our approach improved prediction accuracy and outperformed a variety of other methods for this application.

However, while the proposed method enables accurate prediction of neuromotor outcomes, the feature weights learned by the model are not easily interpretable in terms of particular brain regions (due to the global nature of the features and the PCA step, which combines these global features together in different combinations). In subsequent chapters, we will explore ways (e.g., via different machine learning techniques) to further improve prediction accuracy as well as ways to determine which connections in the brain are most important for predicting infant outcomes.

Chapter 5

Predictive Subnetwork Extraction with Structural Priors for Infant Connectomes

5.1 Introduction

5.1.1 Motivation

Very preterm birth ($\leq$ 32 weeks post-menstrual age, PMA) affects brain development and puts a child at a high risk, not only for altered neuromotor development, but also for altered cognitive development. It is known from studies of diffusion MR images, that the development of white matter plays a critical role in the function of a child's brain, and that white matter injury is associated with poorer outcomes [14, 59, 323, 44].

Representing a diffusion tensor image (DTI) of the brain as a network (i.e., connectome) defined between regions of interest (ROIs) allows an anatomically informed reduction of dimensionality from millions of tensor-valued voxels down to thousands of connections (edges). However, for the purposes of prediction, thousands of features may still be too many and cause over-fitting when limited numbers (e.g. only hundreds) of scans are available [188].

In Chapter 4, we showed that a small set of features (i.e., 30 features) of network topology could be used to predict abnormal neuromotor function [45]. One drawback of using topological network measures as features is that the learned model weights for each feature are difficult to interpret and do not correspond to individual regions or connections in the brain. Furthermore, ROI based studies suggest that structural abnormalities related to poor neurodevelopmental outcomes are not spread evenly across the entire brain, but instead are localized to particular anatomy [59]. Thus, there is motivation to discover which particular subnetworks (group of connections or edges) in the brain network best predict different brain functions or different aspects of development.

5.1.2 Previous Works

Some previous works have explored the use of brain subnetworks for predicting outcomes [322, 106, 188, 166]. For instance, Zhu et al. used t-tests at each edge in a dataset of functional connectomes for group discriminance, followed by correlation-based feature selection (CFS) and training of a support vector machine (SVM), to find subnetworks that were predictive of schizophrenia [322]. This multi-stage feature selection and model training is not ideal, however, because it precludes simultaneous optimization of all model parameters.

Munsell et al. used an Elastic-Net based subnetwork selection for predicting the presence of temporal lobe epilepsy and the success of corrective surgery in adults [188]. This method encourages sparse selection of stable features, useful for identifying those edges most important for prediction [114], but fails to leverage the underlying structure of the brain networks that might inform the importance or the relationships between edges. In order to capture dependencies between neighbouring edges, Li et al. employed a Laplacian-based regularizer (in a framework similar to GraphNet [114]) that encouraged their subnetwork weights to smoothly vary between neighbouring edges [166]. However, this smoothing may reduce sparsity by promoting many small weights and blur discontinuities between the weights of neighbouring edges that should be preserved. An ideal regularizer would encourage a well connected subnetwork while preserving sparsity and discontinuities.

Ghanbari et al. used non-negative matrix factorization to find a sparse set of non-negative basis subnetworks in structural connectomes [106]. However, rather than trying to predict specific outcomes (as we propose below), Ghanbari et al. introduced age-regressive, group-discriminative, and reconstructive regularization terms on groups of subnetworks, encouraging each group to covary with a particular factor. They argued that non-negative subnetwork edge weights are more anatomically interpretable, especially in the case of structural connectomes which have only non-negative edge feature values.

In this chapter, we present our novel approach to identifying anatomical subnetworks of the human white-matter connectome that are optimally predictive of a preterm infant's cognitive and motor neurodevelopmental scores assessed at 18 months of age, adjusted for prematurity. Similar to Munsell et al., our method is based on a regularized linear regression on the outcome score of choice. Here, however, we introduce a constraint that ensures the non-negativity of subnetwork edge weights. We further propose two novel informed priors designed to find predictive edges that are both anatomically plausible and well integrated into a connected subnetwork. We demonstrate that these priors effectuate the desired effect on the learned subnetworks and that, consequently, our method outperforms a variety of other competing methods on this very challenging outcome prediction task. Finally, we discuss the structure of the learned subnetworks in the context of the underlying neuroanatomy.

5.2 Method

5.2.1 Preterm Infant Data

In this chapter, we use the same validation dataset as was used in Chapter 4. Briefly, the dataset contains 168 scans taken between 27 and 45 weeks PMA from a cohort of 115 preterm infants (nearly half of the infants were scanned twice), born between 24 and 32 weeks PMA. Connectomes were generated for each scan by aligning an infant atlas of 90 anatomical brain regions with each DTI. Full-brain streamline tractography was then performed in order to count the number of tracts (i.e., edge strength) connecting each pair of regions. Previous chapters provide details on the scanning and connectome construction processes (Chapter 4) and a discussion on interpreting infant connectomes (Chapter 3). Cognitive and neuromotor function of each infant was assessed at 18 months of age, corrected for prematurity, using the Bayley Scales of Infant and Toddler Development 3rd edition [26]. The scores are normalized to 100 ± 15; adverse outcomes are those with scores at or below 85 (i.e., ≤ -1 std.). Note that the definition of adverse outcomes used here (≤ 85) is slightly different from the definition used in Chapter 4 (< 85) because here we examine cognitive outcomes in addition to motor outcomes and we found that including scans with cognitive outcome scores of exactly 85 considerably helped to balance the classes for this label.

Despite this modification to the class definitions, our dataset is imbalanced, containing few scans of infants with high and low outcome scores. In order to flatten this distribution, the number of connectomes in each training set was doubled by synthesizing instances with high and low outcome scores, using the synthetic minority over-sampling technique (SMOTE) [60]. Note that we did not use local synthetic instances (LSI, proposed in Chapter 4) because our primary task here was to perform regression on the outcome scores directly (with classification of abnormality as a secondary prediction step). Whereas LSI is able to synthesize new instances in particular classes, it is not designed to flatten a distribution of continuous labels as was required for this task.

5.2.2 Subnetwork Extraction

Given a set of preterm infant connectomes, our goal is to find a subnetwork that is:

a) predictive (i.e., contains edges that accurately predict a neurodevelopmental outcome),
b) anatomically plausible (i.e., edges correspond to valid axon bundles),
c) well connected (i.e., high network integration as described in Chapter 3)
d) reasonably sparse and
e) non-negative.

Each connectome is represented as a graph $G(V, E)$ comprising a set of 90 vertices, V, and $M = 90 \times 89/2 = 4005$ edges, E. The tract counts associated with the edges are represented as a single feature vector $\mathbf{x} \in \mathbb{R}^{1\times M}$ and the entire training set of N subjects is represented as $X \in \mathbb{R}^{N\times M}$ with outcome scores $\mathbf{y} \in \mathbb{R}^{N\times 1}$. To find a subnetwork that fits the above criteria, we optimize an objective function over a vector of subnetwork edge weights, $\mathbf{w} \in \mathbb{R}^{M\times 1}$:

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} ||\mathbf{y} - X\mathbf{w}||^2 + \lambda_{L1}||\mathbf{w}||_1 + \lambda_B(\mathbf{w}^T B\mathbf{w}) + \lambda_C(\mathbf{w}^T C\mathbf{w}) \tag{5.1}$$

$$\text{such that } \mathbf{w} \geq 0, \tag{5.2}$$

where $||\mathbf{w}||_1$ is a sparsity regularization term, B is the network backbone prior matrix (see Section 2.3), and C is the connectivity prior matrix (see Section 2.4). Hyper-parameters, λ_B, λ_C and λ_{L1} are used to weight each of the regularization terms. Given a set of learned weights, $\mathbf{w}^*$, the outcome score of a novel infant connectome, $\mathbf{x}_{new}$ can be predicted as $y_{pred} = \mathbf{x}_{new}\mathbf{w}^*$.

Note that since X is non-negative and since $\mathbf{w}$ is required to be non-negative, we also require $\mathbf{y}$ to be non-negative, as they should since the true Bayley scores range between 45 and 155. To perform this optimization we used the method (and software) of Schmidt et al. [239]. While the first three terms in Eq. 5.1 are convex, the fourth term (i.e., the connectivity prior term) is not. However, we found empirically that for random initialization values of $\mathbf{w} \in [0, 1]$, our optimizer found a consistent non-degenerate solution.

5.2.3 Network Backbone Prior

Many of the 4005 possible connectome edges are anatomically unlikely (i.e., between regions not connected by white matter fibers) but may be non-zero in certain scans due to imaging noise and accumulated pipeline error (i.e. due to atlas registration, tractography, and tract counting) [64]. With many more edges than training samples, some edges may appear discriminative by pure chance, when in fact they are just noise. Therefore, we propose a network backbone prior term that encodes a penalty discouraging the subnetwork from including edges with a low signal-to-noise ratio (SNR) in the training data. The SNR of the j-th edge can be computed as the ratio $\text{MEAN}(X_{:,j})/\text{SD}(X_{:,j})$. However, this may falsely declare an edge as noisy when the variability (c.f. denominator) in the edge value is not due to noise but rather due to the edges' values changing in a manner that correlates with the outcome of the subject. To counteract this problem, we divide the scans into two classes: scans with normal outcomes, H, and scans with adverse outcomes, U. The SNR is then computed separately for each class. Let X_Ω represent a matrix with a subset of the rows in X where $\Omega \in \{U, H\}$. The SNR for each edge, j, in each class, Ω, is computed

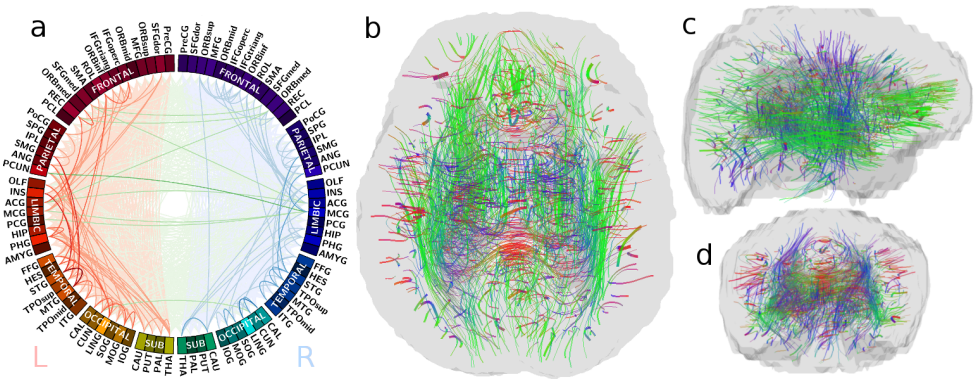

Figure 5.1: **a)** A sample backbone prior network (i.e., all edges where $B_{j,j} = 0$) mapped on to a Circos ideogram (http://circos.ca/). Inter-hemispherical connections are in green and intra-hemispherical connections are in red (left) and blue (right). Opacity of each link is computed as $\text{SNR}(X_{U,j}) \times \text{SNR}(X_{H,j})$. **b)** Axial, **c)** sagittal and **d)** coronal views of the same network rendered as curves representing the mean shape of all tracts between those connected regions (from one infant's scan).

as $\text{SNR}(X_{\Omega,j}) = \frac{\text{MEAN}(X_{\Omega,j})}{\text{SD}(X_{\Omega,j})}$. In order not to favour the strongest fiber bundles over weak yet important bundles, we threshold the SNR at each edge conservatively, to exclude only the least anatomically likely edges. An edge, j, is only penalized if both $\text{SNR}(X_{U,j})$ and $\text{SNR}(X_{H,j})$ are less than or equal to 1 (i.e., signal is weaker than noise in both classes). In particular, B is an $M \times M$ diagonal matrix, such that,

$$B_{j,j} = \begin{cases} 1, & \text{if } \text{SNR}(X_{H,j}) \leq 1 \text{ and } \text{SNR}(X_{U,j}) \leq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{5.3}$$

So $\mathbf{w}^T B \mathbf{w}$ only penalizes edges that do not pass the SNR threshold among either instances with normal outcomes or abnormal outcomes, and thus are likely noisy. Fig. 5.1 shows an example of B. Note that, especially for infant connectomes, even edges with high SNR may not represent white matter fibers but instead high FA from other causes (Chapter 3). Nevertheless, such high-SNR edges are not likely due to noise but instead to some real effect and thus may aid prediction.

5.2.4 Connectivity Prior

We also want to encourage the subnetwork to be highly integrated as opposed to being a set of scattered, disconnected edges. This is motivated by the fact that functional brain network activity is generally constrained to white matter structure [126] and white matter structure is organized into well connected link communities [76]. Thus, we do not expect there to be many, disconnected sub-parts of the brain that are all highly responsible for any particular neurodevelopmental outcome type. To embed this prior, we incentivize pairs of

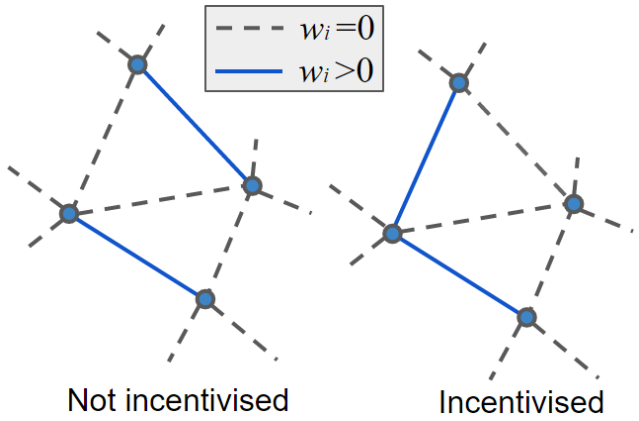

Figure 5.2: Edge connectivity prior: by including this prior, the model objective function encourages strong weights on pairs of edges that share a node (right) but not pairs that do not share a node (left).

edges in the target subnetwork to share common nodes (Fig. 5.2). For edge $e_{i,j}$, between nodes i and j, and edge $e_{p,q}$ between nodes p and q, we construct the matrix,

$$C(e_{i,j}, e_{p,q}) = \begin{cases} -1, & \text{if } i = p \text{ or } i = q \text{ or } j = p \text{ or } j = q \\ 0, & \text{otherwise,} \end{cases} \tag{5.4}$$

such that the term $\mathbf{w}^T C \mathbf{w}$ becomes smaller (i.e., more optimal) for each pair of non-zero weighted subnetwork edges sharing a node. This term places a priority on retaining edges in the subnetwork that are connected to hub nodes. This is desirable since subnetwork hub nodes indicate regions that join many connections (i.e., edges) predictive of outcome. In contrast to a Laplacian based regularizer which would encourage subnetwork weights to become locally similar, reducing sparsity, our proposed term simply rewards subnetworks with stronger hubs.

5.3 Results

We compare the proposed subnetwork-driven predicted outcomes for the preterm infant cohort (N=168) with competing outcome prediction techniques. Methods are evaluated using i) Pearson's correlation between ground truth and predicted scores, and ii) the area over the regression error characteristic curve (AOC), which provides an estimate of regression error [32]. Some previous studies have focused on predicting a binary abnormality label instead of predicting actual scalar outcome scores [45, 323]. Thus, to compare more directly to these works, we also evaluate the accuracy of our models as a binary classifier for predicting scores above or below 85. Similar to in Chapter 4, an SVM was used to classify normal from abnormal instances as it was found to perform better than thresholding the

Table 5.1: Correlation (r) between ground-truth and predicted scores, area over REC curve (AOC) values and classification accuracy of scores at or below 85 (acc.) for each model, assessed via 1000 rounds of leave-2-out cross validation. Note that our approach presented in Chapter 4 performs binary classification only.

	Motor			Cognitive		
Method	r	AOC	acc.	r	AOC	acc.
t-test + CFS [322]	0.159	27.39	45.1	0.021	28.05	49.7
Elastic-Net [188]	0.270	24.58	58.8	0.207	24.83	54.8
Network Measures (Chapter 4)	-	-	62.9	-	-	52.6
Linear Regression	0.270	24.78	58.8	0.245	24.72	55.2
+ L1 regularization	0.314	18.55	64.0	0.244	24.75	55.2
+ Non-neg. Constraint	0.433	14.53	68.8	0.317	17.73	57.7
+ Backbone Prior	0.436	14.47	68.6	0.327	17.82	58.5
+ Connectivity Prior (Ours)	**0.442**	**14.25**	**70.8**	**0.343**	**17.38**	**59.5**

predicted scores at 85. SVM learns a max-margin threshold for the predicted scores (i.e., one input feature), optimal for classification over the training set.

For each method (both proposed and competing), coarse grid searches were performed in powers of two over the method's hyper-parameters to find the best performance for both cognitive and motor outcomes independently. For the proposed method, this search was over $\lambda_{L1}, \lambda_C, \lambda_B \in \{2^0, ..., 2^9\}$. A finer grid search was not performed to avoid over-fitting to the dataset. For each setting of the parameters, a leave-2-out, 1000-round cross validation test was performed. If two scans were of the same infant, those scans were not split between test and training sets. Table 5.1 shows a comparison of the different methods tested on the preterm infant connectomes for prediction of motor and cognitive scores.

Our proposed method with backbone and connectivity priors achieved the highest correlations, lowest AOCs and best 2-class classification accuracies for both motor and cognitive scores (for parameter settings, $[\lambda_{L1}, \lambda_C, \lambda_B]$ of $[2^2, 2^1, 2^6]$ and $[2^5, 2^2, 2^5]$, respectively). For 2-class classification in particular, this method outperformed using network measure features (Chapter 4) by 7.4%, Elastic-Net [188] by 8.4% and t-tests followed by CFS [322] by 17.6% higher accuracy on average (Fig. 5.3). Note that the motor prediction accuracy reported here for the method described in Chapter 4 is lower than what was reported there due to a different definition of classes into normal and abnomal outcomes (described above in Section 5.2.1). Using a two-proportion z-test, we found all these differences to be statistically significant ($p < 0.05$). Also, note that, beginning with standard linear regression, the correlation values improved as each regularization term was added. All tested methods had statistically significant ($p < 0.05$) correlations since, for $1000 \times 2 = 2000$ total predictions, the threshold for 95% significance is $r \geq 0.0439$.

Fig. 5.4 displays the predictive subnetworks learned by our proposed method (averaged over all rounds of cross validation). Subnetworks were stable across rounds: 93.6% of all edges were consistently in or out of the subnetwork 95% of the time. We examined the

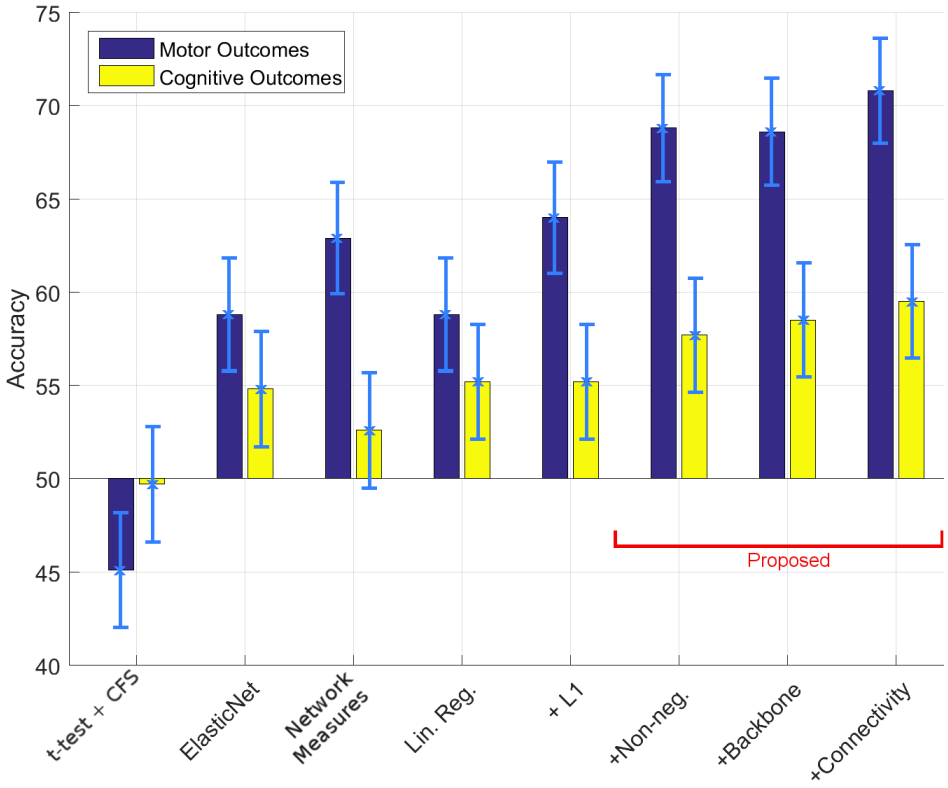

Figure 5.3: Two-class (normal/abnormal) prediction accuracies and 95% confidence intervals for motor and cognitive neurodevelopmental outcomes across all tested models.

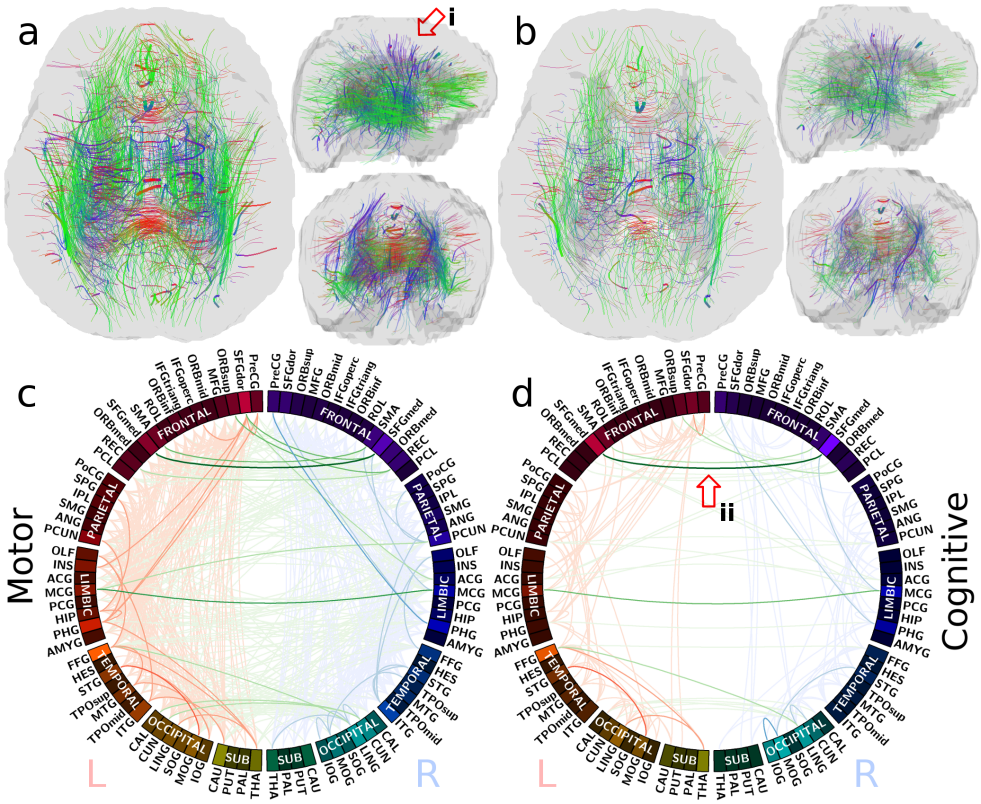

Figure 5.4: **(Top)** Optimal weighted subnetworks for prediction of **a)** motor and **b)** cognitive outcomes. Stronger edge weights are represented with more opaque streamlines. **(Bottom)** Circos ideograms for the **c)** motor and **d)** cognitive subnetworks.

structure of the selected subnetworks to analyse the effect of the proposed regularization terms. By including the L1 regularization term, the learned subnetworks were very sparse, having an average of 71.6% and 98.2% of edge weights set to zero for motor and cognitive scores, respectively, up from only 6.7% (for either score) without the L1 term. Adding the backbone network prior reduced the number of low SNR edges (i.e., $B_{j,j} = 1$) by 18.6% percent for motor score prediction and 11.2% for cognitive score prediction. Adding the connectivity prior improved subnetwork efficiencies (a measure of network integration [44]) by a factor of 6.8 (from 0.0059 to 0.0403) and 2.2 (from 0.2807 to 0.6215) for subnetworks predictive of motor and cognitive scores, respectively.

As expected, the predictive motor subnetwork clearly includes the cortico-spinal tracts (Fig. 5.4a.i). The predictive cognitive subnetwork was more sparse and had generally lower weights than the motor subnetwork (as visualized by less dense, more transparent streamlines), due to the larger L1 weight used for best prediction of the cognitive scores. However, the left and right medial superior frontal gyri (SFGmed) and the connection between these two regions that had stronger weights (factor of 2.1) in the cognitive network than in the motor network, (Fig. 5.4d.ii). This is not surprising as these regions contain the

presupplementary motor area which is thought to be responsible for a range of cognitive functions [317].

5.4 Conclusions

To better understand neurodevelopment and to allow for early intervention when poor outcomes are predicted, we proposed a framework for learning subnetworks of structural connectomes that are predictive of neurodevelopmental outcomes for infants born very preterm. We found that by introducing our novel network backbone prior, the learned subnetworks were more robust to noise by including fewer edges with low SNR weights. By including our connectivity prior, the subnetworks became more highly integrated, a property we expect for subnetworks pertinent to specific functions. Compared to other methods, our approach achieved the best accuracies for predicting both cognitive and motor scores of preterm infants, 18 months into the future.

While our highly regularized linear model performed well, in the next chapter, we move to a non-linear, stacked model and investigate if it can better model the relationship between an infant's connectome at birth and its neurodevelopmental outcome. In future works, we will address the non-convexity of the proposed connectivity prior term.

Chapter 6

BrainNetCNN: Convolutional Neural Networks for Brain Networks; Towards Predicting Neurodevelopment

6.1 Introduction

6.1.1 Motivation

In Chapters 4 and 5 we showed that it is possible to predict neurodevelopmental outcomes of infants born preterm by training predictive learning models on dMRI-based connectome features of each infant's brain. However, whereas these previous approaches employed linear models, the complexity of the processes governing brain development and injury as well as the large number of confounding factors which may influence neurodevelopment suggest that a nonlinear model may be more appropriate [115, 182]. Thus, in this chapter we propose an alternative deep learning approach which inherently allows a high degree of nonlinearity.

Artificial Neural Networks (ANNs),[1] specifically Convolutional Neural Networks (CNNs), have had much success lately in performing prediction tasks on medical image data [68, 69, 232]. CNNs are especially useful when important features are too complex to be designed or even easily recognized by human observers [160]. In this chapter, we propose BrainNetCNN, a novel type of CNN with specially-designed *edge-to-edge*, *edge-to-node* and *node-to-graph* convolutional layer types for brain network data. These novel layer types are actually specific cases of more general convolutional filters that have meaningful interpretations in terms

[1] We refer to two types of networks in this chapter: The artificial neural networks (e.g., CNN) and the human brain network (connectomes). To avoid possible confusion between the two, we have endeavoured to make the distinction clear from the context and use of qualifiers such as 'brain' or 'convolutional'.

of network topology. BrainNetCNN is the first deep learning framework with architecture *designed specifically* for brain network data.

We validate our BrainNetCNN on both synthetic graph data and dMRI-derived structural brain networks of preterm infants. For this, we use the same 168 dMRI images of very preterm born infants used in Chapters 4 and 5. Due to the relatively few number of training instances available, a problem common to many neuroimaging applications, CNNs are advantageous as they share weights within layers which can reduce the number of free parameters to learn when compared to fully connected neural networks. We first demonstrate this in controlled experiments on synthetic graph data by showing that BrainNetCNN outperforms a fully connected neural-network with the same number of model parameters.

On the preterm infant connectome data, we first test BrainNetCNN with the task of predicting infant PMA at the time of scan. BrainNetCNN is able to predict an infant's age with an average error of about 2 weeks, demonstrating that it can learn relevant topological features from the connectome data. Finally, we apply BrainNetCNN to the much more challenging task of predicting neurodevelopmental scores. We were able to achieve statistically significant correlations between predicted scores and true scores, with an average prediction error of around 11%. Furthermore, we show that BrainNetCNN achieves significantly higher correlation values than other competing prediction methods on this task.

Finally, we explore the high-level features learned by the CNN by visualizing which connections in the brain are most predictive of age, cognitive outcomes and motor outcomes. We find that edges important for predicting age are well distributed across the brain network. Also, we find that edges important for motor score prediction are connected to regions known to be responsible for motor function, and that other unique connections are important to predict cognitive scores.

6.1.2 Related Works

The usefulness of representing the brain as a structural brain network for inference or prediction of injury and disease in adults has been widely recognized [74, 106, 188, 322]. However, only a very limited number of studies (other than our previous works in Chapters 4 and 5) have applied these techniques to scans of infants (Chapter 2). Ziv et al. examined if it were possible to predict general neurological health of infants at 6 months after birth using brain networks derived from DTIs. They employed a support vector machine (SVM) trained on high-level topological features [323].

While the application of ANNs to medical image analysis is well established for some clinical applications, its use for neurological applications has only lately become more popular [313, 307, 173, 165, 40, 262, 261, 86]. For instance, ANNs have recently been used to segment brain lesions in multiple sclerosis patients [313], segment brain tumors in multimodal MRI volumes [86], and classify different types of cerebellar ataxia [307]. Various deep architectures have also recently been used to predict stages of Alzheimer's disease

progression [173, 165, 262, 261]. Similarly, Brosch and Tam employed deep belief networks to learn a manifold describing variation in a population of Alzheimer's patients [40]. The networks in these studies, however, were all trained over standard grid-like MR images of brain structure as opposed to graph or network representations of brain structure.

Very few papers have applied ANNs to brain connectivity data. Munsell et al. used a fully connected deep auto-encoder to extract features from connectomes, but did not explicitly consider the structure of the brain network in the fully connected layers [188]. Plis et al. explored the use of deep belief networks for a variety of classification tasks over functional MR (fMRI) and standard MR brain data, but collapsed the spatial dimensions of each input image to a single vector of voxels [213].

Recently, Bruna et al. and Henaff et al. showed that CNNs could be applied to data over a graphical domain (as opposed to grid-like data such as images) [48, 122]. Their work followed work by Shuman et al. who showed how to generalize convolutions to graph structured domains [250]. In those works the input signal was given over the nodes of the graph with a single set of edge weights fixed for all samples. In contrast, for the case of structural brain networks, the input signal is given as weights over the edges (reflecting, e.g., connectivity strength), implying a different set of edge weights for each sample. Thus, the techniques described by those works are not immediately applicable to brain network data and so, here, we introduce specialized filters for the task. There is, however, a relationship between convolutions over graphs as defined by Shuman et al. and the edge-to-edge filters we propose in this chapter (detailed in Section 6.2.1).

Finally, some recent works have leveraged graph kernels to facilitate kernel based learning on connectome data [141, 79]. In contrast to graph convolutions, graph kernels do not explicitly extract graph features but instead define an inner product between graphs. As far as we are aware, however, none of these works have applied graph kernels to infant structural brain networks nor incorporated them into a deep learning framework. We know of no other work, to date, that has adapted CNNs for edge-weighted networks and applied them to the human connectome.

6.2 Method

Here, we present our novel CNN layer types, designed specifically for network data input (Section 6.2.1), the dataset used in this study (Section 6.2.2), the overall architecture of BrainNetCNN (Section 6.2.3), how we implemented BrainNetCNN (Section 6.2.4) and finally our evaluation metrics (Section 6.2.5).

6.2.1 CNN Layers for Network Data

A dMRI-derived brain network, $G = (E, V)$, is a compact representation of the white matter connections in a patient's brain, where V is a set of nodes representing regions in

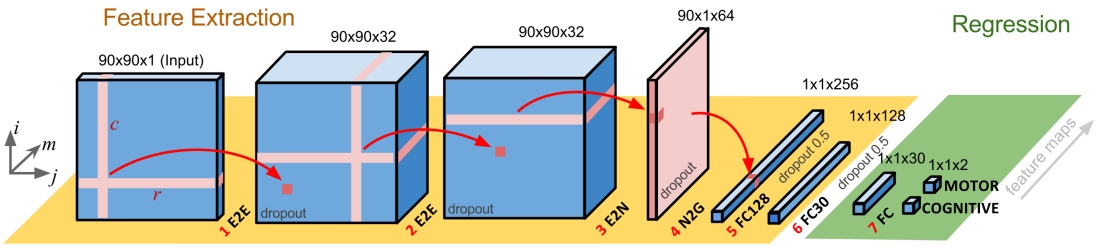

Figure 6.1: Schematic representation of the BrainNetCNN architecture. Each block represents the input and/or output of the numbered filter layers. The 3rd dimension of each block (i.e., along vector m) represents the number of feature maps, M, at that stage. The brain network adjacency matrix (*leftmost block*) is first convolved with one or more (two in this case) E2E filters which weight edges of adjacent brain regions. The response is convolved with an E2N filter which assigns each brain region a weighted sum of its edges. The N2G assigns a single response based on all the weighted nodes. Finally, fully connected (FC) layers reduce the number of features down to two output score predictions.

the brain and E is a set of weighted edges defined by an adjacency matrix, A, representing the connection strength between each pair of brain regions (typically defined as the number of white-matter tracts connecting the regions).

One way to apply ANNs to brain network data is to ignore the structure of the brain network and treat the input edge weights as a vector of features [188]. This approach, however, discards the topological relationships between edges that are intrinsic to the data. An alternative approach is to treat the adjacency matrix as an image and use established convolutional filters designed to capture the spatial 2D grid locality of images (e.g., a 5×5 filter). However, spatial locality between entries of the adjacency matrix does not directly correspond to topological locality in the brain network. For an entry located at $A_{i,j}$, only those elements within the i-th row and j-th column are topologically local and so the typical grid convolutional filters used for images are not appropriate here.

We consider these topological differences between images and brain networks as we adapt the CNN paradigm to brain network data. To leverage the structure found within the adjacency matrix, we introduce three new layer types: edge-to-edge layers, edge-to-node layers, and node-to-graph layers. Each layer type consists of one or more simple convolutional filters of a particular shape and performs a specific operation on the brain network. A BrainNetCNN layer contains one or more filters (of the same type). Each filter takes all feature maps from the previous layer as input and then outputs a distinct feature map for the next layer. Note that for all equations of the filter types below, we omit the activation function and the standard bias term for simplicity.

Edge-to-edge Layers

An *edge-to-edge* (E2E) layer is similar to a standard convolutional layer in a CNN over grid-like data in that it filters data locally. Whereas in grid-like data, filters may be defined in

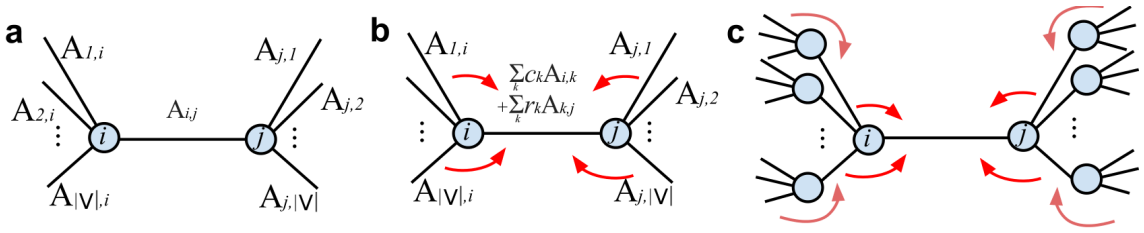

Figure 6.2: An E2E filter at edge (i, j) shown, a) before filtering, b) after being applied once, and c) after being applied twice. For simplicity, these examples assume only one input feature map and one output feature map. Accordingly, the feature map indices and layer indices are omitted.

terms of spatial locality, the E2E filter is defined in terms of topological locality, combining the weights of edges that share nodes together.

Formally, let $G^{\ell,m} = (E^{\ell,m}; V)$ represent the m-th feature map of a weighted brain network at the ℓ-th layer of the CNN, where V is the set of nodes corresponding to brain regions and $E^{\ell,m}$ is a set of weighted edges defined by adjacency matrix $A^{\ell,m} \in \mathbb{R}^{|V|\times|V|}$. Each layer takes M^ℓ feature maps as input, and for this study we assume that $M^1 = 1$ (i.e., the input feature map to the whole CNN is just a single adjacency matrix describing one connectome). Since the number of nodes do not change between input and output, V stays constant and the output of an E2E layer is a filtered adjacency matrix defined as,

$$A_{i,j}^{\ell+1,n} = \sum_{m=1}^{M^\ell} \sum_{k=1}^{|V|} r_k^{\ell,m,n} A_{i,k}^{\ell,m} + c_k^{\ell,m,n} A_{k,j}^{\ell,m} \tag{6.1}$$

where $[\mathbf{c}^{\ell,m,n}, \mathbf{r}^{\ell,m,n}] = \mathbf{w}^{\ell,m,n} \in \mathbb{R}^{2|V|}$ such that $[\mathbf{w}^{\ell,1,n}, ..., \mathbf{w}^{\ell,M^\ell,n}] \in \mathbb{R}^{2|V|\times M^\ell}$ are the learned weights of the nth filter at layer ℓ. Thus, for each pair of input and output feature maps, (m, n), at layer ℓ, the E2E layer learns a single vector of weights, $\mathbf{w}^{\ell,m,n} = [w_1^{\ell,m,n}, ..., w_{2|V|}^{\ell,m,n}]$. The set of all weights, $\{\mathbf{w}^{\ell,m,n} | m \in \{1, 2, ..., M^\ell\}\}$, that contribute to one output feature map, n, in one layer, ℓ, defines a single filter. The E2E filter is illustrated, for a single input feature map, in Fig. 6.2 and in entirety as a block diagram on the left side of Fig. 6.1.

Intuitively, for some edge (i, j) in an adjacency matrix encoded in some feature map, m, an E2E filter computes a weighted sum of edge weights over all edges connected either to node i or j, like a convolution. This implies that a single weight, $w_k^{\ell,m,n}$, is applied to all edges of a given node. This, however, does not imply that edges from a given node are all treated with equal importance. A single edge, (i, j), may be highly weighted if both $r_j^{\ell,m,n}$ and $c_i^{\ell,m,n}$ are large. Multiple distinct edges may then be weighted in this way via different network feature maps.

While this study focuses on the application of BrainNetCNN to undirected graph data, the E2E filter can, more generally, operate on directed graphs. For symmetric input, $A^{\ell,m}$, the output of an E2E filter $A^{\ell+1,n}$ may be asymmetric since, in general, it is not necessarily

true that $r_i^{\ell,m,n} + c_j^{\ell,m,n} = r_j^{\ell,m,n} + c_i^{\ell,m,n}$. The filter may weight the input asymmetrically. For undirected graphs, however, this is simply the same as having two output feature maps (one upper triangular, one lower triangular) and so it isn't necessary to enforce symmetric output. While it might be possible to design a filter similar to the E2E filter that operates only over the upper (or lower) triangular elements, it would very likely preclude the use of standard convolutional filters (i.e., the **r** and **c** components of the E2E filter). The proposed formulation of the E2E filter allows us to leverage these efficient convolutional filters and implement this filter easily in established CNN software packages (see below).

The E2E filter is similar to a $3 \times 3 \times M^\ell$ convolution filter over a stack of 2D grid data, in that, for each feature map, it combines the signal at some point with the signal from the direct neighbours, but does so with a cross shape filter instead of a box-shaped filter. Note that unlike a 2D image, the brain network has no topological boundaries and so the output of the layer can be the same size as the input without requiring any padding. Another difference, as noted above, is that whereas a convolution typically acts on a signal defined over the nodes of the grid (or over a general network as in the case of Shuman et al. [250]), here our filter acts on a signal defined over the edges (i.e., edge weights).

The connection between convolutions over the edges of a graph versus convolutions over the nodes of a graph can be understood in terms of the concept of a line graph [109]: Let $\mathcal{L}(G)$ represent the line graph of G. Briefly, $\mathcal{L}(G)$ is a graph with one node corresponding to each edge in G and one edge corresponding to each pair of edges in G that are joined by a node. The nodes of $\mathcal{L}(G)$ adopt the signal over the edges of G (i.e., edge weights) and because there is no signal over the nodes of G, the topology of $\mathcal{L}(G)$ is consistent over the entire dataset. Thus, by constructing $\mathcal{L}(G)$, the definition of convolution over a graph by Shuman et al. [250] becomes applicable to brain network data. It turns out that an E2E filter over G is equivalent to a filter over $\mathcal{L}(G)$ with a K-hop of 1, which, as demonstrated by Shuman et al. [249], can be written as a generalized convolution. Note, however, that for typical sizes of V, in the order of dozens to hundreds (e.g., 90, as is the case here), $\mathcal{L}(G)$ contains $\frac{1}{2}|V|^3 - \frac{1}{2}|V|(|V|-1) = 360{,}495$ edges versus only $\frac{1}{2}|V|(|V|-1) = 4{,}005$ for G, making operations over $\mathcal{L}(G)$ much more memory intensive. Thus, for efficiency and ease of interpretation, we chose to define the E2E filter in terms of G rather than $\mathcal{L}(G)$.

Edge-to-Node Layer

An *edge-to-node* (E2N) filter takes an adjacency matrix, $A^{\ell,m}$, (representing a, possibly filtered, brain network) from each feature map as input and outputs a vector of size $|V|$. Thus, the output of an E2N layer is defined as,

$$a_i^{\ell+1,n} = \sum_{m=1}^{M^\ell} \sum_{k=1}^{|V|} r_k^{\ell,m,n} A_{i,k}^{\ell,m} + c_k^{\ell,m,n} A_{k,i}^{\ell,m}, \tag{6.2}$$

where, similar to an E2E layer, $[\mathbf{c}^{\ell,m,n}, \mathbf{r}^{\ell,m,n}] = \mathbf{w}^{\ell,m,n} \in \mathbb{R}^{2|V|}$ such that $[\mathbf{w}^{\ell,1,n}, ..., \mathbf{w}^{\ell,M^\ell,n}] \in \mathbb{R}^{2|V| \times M^\ell}$ are the learned weights of the nth filter at layer ℓ. However, the n-th output feature map, $\mathbf{a}^{\ell+1,n}$, of an E2N layer is a vector in $R^{|V| \times 1}$, in contrast to an E2E layer whose output feature map is in $R^{|V| \times |V|}$.

An E2N filter is equivalent to convolving the adjacency matrix with a spatial 1D convolutional row filter and adding the result to the transpose of the output from a 1D convolutional column filter. This operation can be interpreted as computing a single output value for each node, i, by taking a weighted combination of the incoming and outgoing weights of each edge connected to i. Note that if we assume the input to the E2N filter is a symmetric matrix, we can drop either the term containing the row weights, $\mathbf{r}^{\ell,m,n}$, or the term containing the column weights, $\mathbf{c}^{\ell,m,n}$, since the incoming and outgoing weights on each edge will be equal. In all experiments in this chapter, we used E2N filters with only the $|V|$ row weights in $\mathbf{r}$ because we did not empirically find any clear advantage in learning separate weights for both incoming and outgoing edges when training over symmetric connectome data.

Similar to the E2E layer, the E2N layer does not necessarily discard information about distinct edges with particular importance: If weights $r_i^{\ell,m,n}$, $c_i^{\ell,m,n}$, $r_j^{\ell,m,n}$ and $c_j^{\ell,m,n}$ are all relatively large, then edge (i, j) will be weighted especially strongly and through multiple feature maps, many edges may be highly weighted in this way.

Node-to-Graph Layer

Finally, similar to the E2N layer, a node-to-graph (N2G) layer reduces the dimensionality of the input, in this case by taking a weighted combination of nodes to output a single scalar,

$$a^{\ell+1,n} = \sum_{m=1}^{M^\ell} \sum_{i=1}^{|V|} w_i^{\ell,m,n} a_i^{\ell,m}, \tag{6.3}$$

per output feature map, n. The N2G filter, also a 1D spatial convolution, is applied after an E2N filter and reduces the spatial dimensions of the original input to single scalar per feature map. In the context of being applied after an E2N filter, which summarizes the responses of neighbouring edges into a set of node responses, the N2G filter can be interpreted as getting a single response from all the nodes in the graph.

6.2.2 Preterm Data

The data used in this study is the same set of dMRI scans of very preterm infants used in Chapters 4 and 5, imaged at BC Children's Hospital in Vancouver, Canada. After excluding images for poor scan quality, as detailed in Chapter 3 (in short, first by visual inspection of the DTIs and then by examining tractography results for serious artefacts and directional bias), scans of 115 infants were used. Roughly half of the infants were scanned twice (shortly after birth and then again at about 40 weeks PMA), for a total of 168 scans. Full-brain

streamline tractography was performed on each DTI to recover the neuronal connections in each brain. Using a neonatal atlas of $|V| = 90$ anatomical regions from the University of North Carolina (UNC) School of Medicine at Chapel Hill [247], a weighted, undirected network was constructed from each scan by counting the number of tracts connecting each pair of anatomical regions. Each network is represented as a 90×90 symmetric adjacency matrix with zeros along the diagonal and is scaled to $[0, 1]$. At 18 months of age, adjusted for prematurity, the cognitive and neuromotor function of each subject was assessed using the Bayley Scales of Infant and Toddler Development (Bayley-III) [26]. Cognitive and motor scores from this test are normalized to a population mean of 100 with standard deviation of 15. See Chapter 4 for further details about assessment protocol, scanning protocol and connectome construction.

Given the small data set (dMRI of preterm infants is not standard procedure in clinical practice) and the imbalance (low numbers of preterm infants with high and low neurodevelopmental outcomes) we adopted the synthetic minority over-sampling technique (SMOTE) [60] to balance and augment each training set (as was used in Chapter 5). Training samples were binned by score (5 bins) and then SMOTE was run, repeatedly, to generate a synthetic sample from the bin with the fewest total number of real and synthetic samples, until the training set was augmented by a factor of 256. Note that in our previous work, we showed that the proposed local synthetic instances (LSI) method outperformed SMOTE for improving prediction accuracy (Chapter 4). While LSI worked well in that context, we were performing 2-class classification rather than regression. LSI is not applicable here because it augments data in individual classes, and in this chapter we are performing regression over a single training set.

6.2.3 BrainNetCNN Architecture

We base the architecture of our BrainNetCNN (for connectomes) on a common CNN (for images) where the first section of the network is composed of convolutional layers and the last section is composed of fully connected (FC) layers (e.g., [253]). Fig. 6.1 is a block diagram of a representative BrainNetCNN architecture with at least one layer of each of the proposed filter types.

The input to a BrainNetCNN model is a brain network, G^0, represented as a 90×90 adjacency matrix. The output layer of the network has two nodes where each node predicts a different neurodevelopmental outcome score (motor and cognitive). The second to last layer in the network of size $1 \times 1 \times 30$ can be interpreted as a set of high-level features learned by the previous layers. We selected a size of 30 features in order to directly compare the features learned by BrainNetCNN to the 30 network measure features used in Chapter 4.

Since E2E layers operate on a whole adjacency matrix (per feature map), they can only be applied before E2N and N2G, which reduce the input dimensionality (to a vector or a scalar per feature map). However, since E2E layers do not alter the input dimensionality,

many E2E layers can be stacked (with the trade-off of an increased number of parameters to learn). An E2N layer reduces the 90 by 90 matrix to a single matrix of 90 by 1 elements and thus must be applied before an N2G layer. The N2G layer reduces the input dimensionality down to a single feature (per feature map) and thus cannot be applied before the E2E or E2N layers.

In the experiments below (Section 6.3) we test a variety of configurations of BrainNetCNN. Each configuration of BrainNetCNN can be understood as a CNN with a subset of the layers shown in Fig. 6.1. The basic configuration (E2Enet) contains one of each type of proposed layers along with 3 fully connected layers (i.e., layers 1, 3, 4, 5, 6 and 7 in Fig. 6.1). We also tested configurations with fewer layers: One model with the E2E layer removed (E2Nnet), and two more models similar to E2Enet and E2Nnet but with two of the fully connected layers removed (E2Enet-sml and E2Nnet-sml, respectively). Finding good results with these FC layers removed, we tested a model with the same layers as E2Enet-sml but with an additional E2E layer (2E2Enet-sml).

We compare our results from these BrainNetCNN configurations to one and two layer fully connected neural networks (FC30net and FC90net, respectively), which don't contain any of the proposed convolutional layers. The input to the FC networks is a 1×4005 vector consisting of the upper triangular values of the symmetric connectome matrix. FC90net is similar to layers 5, 6 and 7 in Fig. 6.1 but with only 90 responses between layers 5 and 6 to make the number of learnable parameters approximately equal to that in E2Nnet-sml and E2Enet-sml.

Generally, the number of output feature maps from each layer, M^ℓ, is independent of other network parameters and can be set freely. In the BrainNetCNN architecture, we increased the number of feature maps with each layer to compensate for the reductions along the other dimensions (i.e., dimensions i and j in Fig. 6.1); a common strategy for CNNs (e.g.,[253]). Precisely, E2Nnet-sml has an E2N layer with 130 1×90 filters (layer 3 is increased from 64 to 130 to match the number of parameters with the other models) producing feature maps of size $1 \times 90 \times 130$. This is followed by an N2G layer with feature maps of size $1 \times 1 \times 30$ (layer 4) and a fully connected layer with an output of size 2 (layer 7). E2Enet-sml is constructed from layers 1, 3, 4, 7 (Fig. 6.1), with an E2E layer composed of 32 1×90 and 32 90×1 filters (layer 1) producing feature maps of size $90 \times 90 \times 32$. This is followed by an E2N layer with 64 $1 \times 90 \times 32$ filters (layer 3) producing feature maps of size $1 \times 90 \times 64$, an N2G layer with feature maps of size $1 \times 1 \times 30$ (layer 4), and a fully connected layer with an output of size 2 (layer 7).

Every layer in our network uses very leaky rectified linear units as an activation function, where a leaky value of $x/3$ is assigned if $f(x) < 0$, as done by Graham [112]. For training, we employed dropout using a rate of 0.5 after the N2G layer and the FC layer of 128 units as shown in Fig. 6.1 (dropout was found to slightly improve correlation by ≈ 0.01 for the fully connected model). We used momentum of 0.9, a mini-batch of size 14, a weight decay

of 0.0005, and a learning rate of 0.01. Mini-batch sizes, weight decay and learning rates were set to values that performed well over the fully connected model (see Section 6.3.2). All models minimized the training loss, which is defined as the Euclidean distance between the predicted and real outcomes plus a weighted $L2$ regularization term over the network parameters.

The ideal number of training iterations for a given model depends on the model architecture and on the training parameters. Thus, to minimize overfitting to the training data, and to ensure a fair comparison across all model types (both proposed and competing), we trained each model for a variable number of iterations, from 10K to 100K (in 10K increments) and selected the model corresponding to the number of iterations that yielded the least overfitting (i.e., best performance on the test data).

6.2.4 Implementation

We implemented our BrainNetCNN using the popular deep learning framework, Caffe [135]. While the E2N and N2G filters were straightforward to implement using 1D filters, the E2E filter required a convolution of two 1D filters, $\mathbf{c} \in \mathbb{R}^{|V|\times 1}$ and $\mathbf{r} \in \mathbb{R}^{1\times|V|}$, with the adjacency matrix, producing responses of dimensions $\mathbb{R}^{1\times|V|}$ and $\mathbb{R}^{|V|\times 1}$, respectively. These response vectors are each replicated $|V|$ times to produce two $\mathbb{R}^{|V|\times|V|}$ matrices, which are summed element-wise yielding a single matrix equivalent to Eq. 6.1.

6.2.5 Evaluation Metrics

In addition to reporting mean absolute error (MAE) and the standard deviation of absolute error (SDAE) between the predicted and the true scores, we report the Pearson correlation coefficients between the predicted and the true scores, and the corresponding p-values. As our dataset contains many scores close to the mean value, MAE may be disproportionately low for regressors that frequently predict nearer to the mean score of the training data, even if they underfit the data. The Pearson correlation coefficient, however, measures the linear dependence between predicted and true scores and so is less affected by the distribution of the inputs. MAE is still important to report, however, since Pearson's correlation does not expose if a regressor is biased towards frequently predicting too high or too low. Thus, the measures are complementary.

6.3 Experiments

6.3.1 Simulating Injury Connectomes for Phantom Experiments

Before testing BrainNetCNN on real brain networks, we assessed its ability to learn and discriminate between differing network topologies using sets of synthetically generated networks. We first examined the performance of BrainNetCNN on data with increasing levels

Table 6.1: Synthetic experiments using E2Enet-sml to predict injury parameters α and β under different levels of noise measured by the peak-signal-to-peak-noise-ratio (PSPNR=$1/\gamma$). As expected, as the noise levels decrease, the Pearson correlation r increases (r_α indicates correlation with the α parameter), and the mean absolute error (MAE) and the standard deviation of the absolute error (SDAE) decrease.

PSPNR	r_α	MAE_α	SDAE_α	r_β	MAE_β	SDAE_β
4 (12 dB)	0.554	19.949	14.497	0.588	18.356	13.967
8 (18 dB)	0.873	9.732	7.870	0.873	9.980	8.259
16 (24 dB)	0.965	6.458	5.026	0.969	5.008	4.195
∞	1.000	1.071	0.682	0.999	1.088	0.879

of noise and then compared BrainNetCNN to a fully-connected neural network with the same number of model parameters. To simulate realistic synthetic examples, each example is based on the mean connectome, X_μ (Fig. 6.3-left), of our preterm infant data, perturbed by a simulated focal brain injury using a local signature pattern S. The symmetric matrix $S \in \mathbb{R}^{|V|\times|V|}$ has non-zero elements uniformly selected between $[0, 0.1]$ (i.e., up to 10% of the values of X_μ) along the same row and column index. Thus, the simulated injury is to all connections (with varying intensity) emanating from a single brain region. We created two focal injury signatures, S^1 and S^2, with two corresponding injury regions. These two regions were chosen as the two rows in X_μ with the highest median responses in order to simulate injury to important regions (i.e., hubs) of the brain (Fig. 6.3-right). Mathematically, the i-th synthetic connectome, X_i, is formed as,

$$X_i = \frac{X_\mu}{(\mathbf{1} + \alpha_i S^1)\,(\mathbf{1} + \beta_i S^2)} + \gamma N_i \tag{6.4}$$

where $\mathbf{1}$ is a matrix composed of all ones; $N_i \in \mathbb{R}^{|V|\times|V|}$ is composed of random values simulating noise weighted by γ; and, α_i and β_i are scalar injury parameters that weight their respective signature matrices. α_i and β_i range between 50 to 140 as these are typical neurodevelopmental outcome ranges in our dataset. All operations are done element-wise and the resulting synthetic connectome X_i (Fig. 6.3-center) forms our observed example.

Predicting Injury Parameters over Varying Noise

We first tested our model's ability to predict the injury parameters (i.e., α_i and β_i) given the corresponding X_i under different level of noise, γ. The model was trained using 1000 synthetic examples and test over another 1000 examples. We chose 1000 training samples as it represents a realistic best-case scenario for a large dataset of DTI scans. As shown in Table 6.1, under moderate noise, our BrainNetCNN model (E2Enet-sml) accurately predicts α and β, indicating an ability to recognize multiple subtle, synthetically induced connectome perturbations.

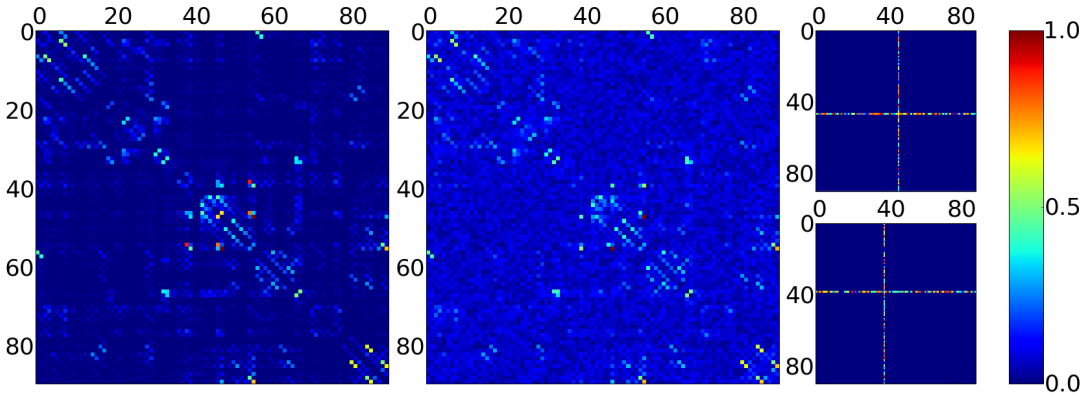

Figure 6.3: (*Left*) The averaged connectome. An example synthetic connectome (*center*) used in our focal injury phantom experiments after introducing noise and the two signatures at the 47th and 39th regions (*right*).

Table 6.2: Comparison of a fully connected model (*top row*) with two proposed BrainNetCNN models (*bottom rows*), all with similar numbers of parameters on phantom data.

Model	r_α	MAE_α	SDAE_α	r_β	MAE_β	SDAE_β
FC90net	0.648	20.583	11.609	0.688	20.080	11.513
E2Nnet-sml	0.736	16.380	10.977	0.752	16.492	9.834
E2Enet-sml	**0.812**	**13.760**	**9.494**	**0.772**	**15.021**	**9.761**

Predicting Focal Injury Parameters with Different Models

We also used the phantom data to assess the difference in predictive ability on a small training set, between a fully connected model (FC90net) and two models based on our proposed BrainNetCNN layers (E2Nnet-sml, E2Enet-sml), each with a similar number of model parameters.

To more closely approximate our real dataset, we used 112 synthetic samples to train, 56 synthetic samples to test and used relatively high, fixed PSNR of 8 (or 18 dB, where PSPNR=$1/\gamma$). The results are reported in Table 6.2.

The E2Enet-sml outperformed the FC90net model achieving an average increase in mean correlation of 15.54% and an average decrease in MAE of 29.17% over both parameters, and slightly outperformed E2Nnet-sml across all measures. The E2Nnet-sml also outperformed FC90net across all measures. As these models all have nearly the same number of parameters to learn, and E2Nnet-sml has the same number of non-linear layers as the FC90net model, these tests indicate that the BrainNetCNN convolutional filters contribute greatly to the improvements in prediction accuracy on this realistic phantom.

Predicting Diffuse Injury Parameters with Different Models

It is thought that poorer neurodevelopmental outcomes in many preterm infants, especially low cognitive scores, may be caused by diffuse white matter injuries rather than focal

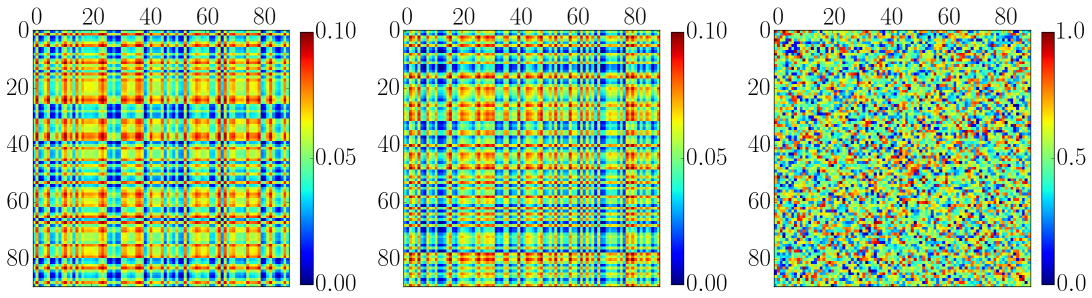

Figure 6.4: (*Left and center*) Sample diffuse whole brain injury patterns. (*Right*) Sample diffuse injury synthetic connectome with two diffuse injury patterns and noise applied.

Table 6.3: Comparison of a fully connected model (*top row*) with two proposed BrainNetCNN models (*bottom rows*), all with similar numbers of parameters on diffuse injury phantom data.

Model	r_α	MAE_α	SDAE_α	r_β	MAE_β	SDAE_β
FC90net	0.129	22.614	**11.946**	0.217	20.796	13.838
E2Nnet-sml	**0.398**	**19.570**	12.476	**0.326**	**19.724**	**13.223**
E2Enet-sml	0.386	19.712	12.483	0.315	19.938	13.531

lesions [14]. Thus, in addition to simulating focal injuries, we also test our method on a phantom dataset with simulated diffuse injuries, spread across the whole brain. The diffuse injury synthetic connectomes are created using the same method described above, in Section 6.3.1, except that the focal injury pattern matrices, S^1 and S^2 are replaced with diffuse injury pattern matrices, D^1 and D^2. Diffuse injury patterns are simulated by selecting a random injury weight (again in $[0, 0.1]$) for each region. Given a 90×1 vector, d^k, of injury weights, a symmetric diffuse injury pattern is computed as $D^k_{i,j} = \frac{1}{2}(d^k_i + d^k_j)$. Examples of diffuse injury patterns and a diffuse injury synthetic connectome are shown in Fig. 6.4. While the same level of noise (PSPNR of 8) was applied to this dataset as for the focal injury phantoms, the broader injury pattern produces a weaker overall connectivity signal, causing the noise to appear more pronounced.

As with the experiment on focal injury phantoms, here we test the ability of FC90net, E2Nnet-sml and E2Enet-sml models to predict two independent injury patterns. On this more challenging phantom data, the BrainNetCNN models again outperform the FC model in terms of both MAE and correlation (Table 6.3). Here, however, the E2Nnet-sml model slightly outperforms the E2Enet-sml.

6.3.2 Infant Age and Neurodevelopmental Outcome Prediction

To test the BrainNetCNN on the preterm infant data, we performed 3-fold cross-validation. The data was split randomly into three folds of 56 scans with the constraint that scans of the same subject were in the same fold. We chose three folds because, despite giving a larger training set size, more folds would require an increased number of (deep) models to

be trained. In each round, two folds were selected as a training set, augmented (as described in Section 6.2.2) and then used to train a model. As ANNs can find different local minima and thus produce different solutions, for each test involving an ANN, we trained each model with five different random initializations and averaged the predicted scores [68, 69, 253].

Model Sensitivity to Initialization and Number of Iterations

As was mentioned above, for a fair comparison, the reported correlation values (i.e., capturing the prediction accuracy) for each architecture were the best achieved for that architecture across different numbers of training iterations. Fig. 6.5 compares the correlation values across increasing numbers of training iterations (from 10K to 100K) for both FC90net and E2Enet-sml architectures. For each type of architecture, predicting each neurodevelopmental outcome type, the correlation values increase rapidly and then roughly plateau after about 30K training iterations. So, while we chose the best number of iterations for each method to be fair to each type of architecture, we observe that the correlation value is fairly insensitive to this parameter. Furthermore, Fig. 6.5 validates that 100K is a good upper limit for number of training iterations, as no model appears like it would greatly improve given more training. In the case of cognitive score prediction using the E2Enet-sml model, the correlation values appear to slightly decrease after 80K iterations, potentially indicating that the model is beginning to over-fit to the training data past this point. Results for both E2Enet-sml and FC90net models are reported in Table 6.4 at the 60K mark since it is the peak of each of their combined correlations. Note that the mean correlation value slightly differs from what is reported in Table 6.4 since Table 6.4 averages the predictions together over the five models before taking the correlation (instead of computing the mean of the correlations for each model as is displayed in Fig. 6.5).

Fig. 6.5 shows the mean and the standard deviation of the correlation values across the predictions of the five different randomly initialized models. Furthermore, the standard deviation decreases with the number of iterations, meaning that the different independently initialized models converge to similar performance after training.

Age Prediction

Before applying BrainNetCNN to the very difficult task of predicting neurodevelopmental outcomes, we first trained it to predict infant PMA at the time of scan. We performed this test to establish an upper-bound on the predictive performance of BrainNetCNN, as there are perhaps fewer complicating factors in predicting age compared to predicting neurodevelopmental outcomes (which we discuss in Section 6.4). Using E2Enet-sml, we were able to accurately predict PMA, with an MAE of 2.17 weeks (or 11.1% of the total age range) and an SDAE of 1.59 weeks. The correlation between predicted and ground-truth age was 0.864. While the purpose of this test is only to show the ability of our model to

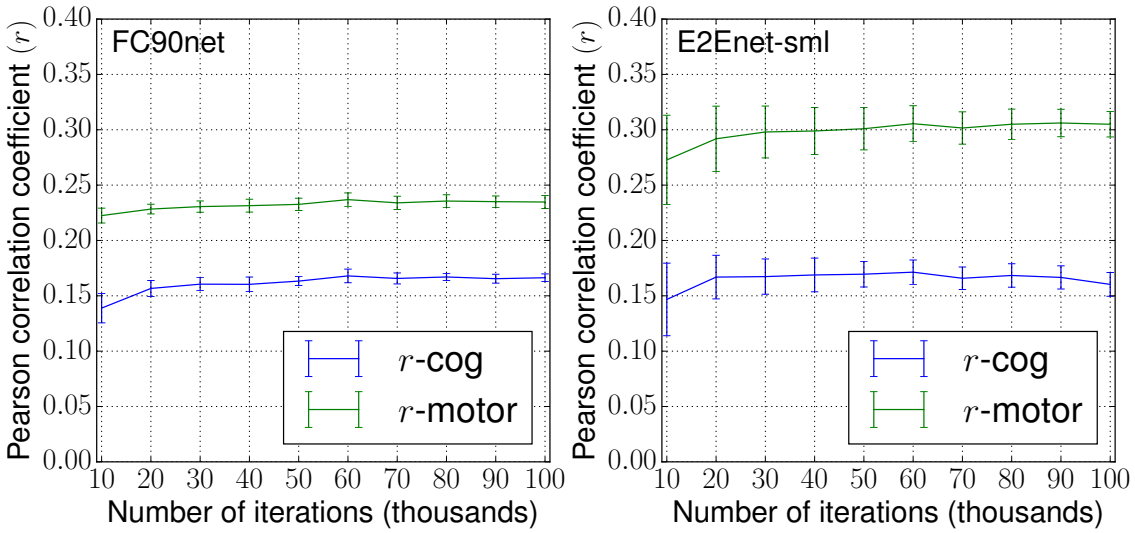

Figure 6.5: The effect of the number of training iterations on correlations between predicted and ground truth outcome scores (on test data) for FC90net (*left*) and E2Enet-sml (*right*) architectures. The standard deviation of each of the five randomly initialized models is shown at each 10K iterations as vertical error bars.

learn some clinical parameters given the connectome data, for completeness, we also tested the FC90net model and the E2Nnet-sml model. On this baseline task, the FC90net model performed slightly worse than E2Enet-sml, achieving an MAE of 2.29 weeks, SDAE of 1.65 weeks and a correlation of 0.858. Similarly, the E2Nnet-sml model slightly underperformed E2Enet-sml, achieving an MAE of 2.377, SDAE of 1.72 and a correlation of 0.843.

We found that absolute error of age prediction (using the E2Enet-sml model) was correlated with PMA ($r = 0.224$), implying that age predictions were more accurate for younger infants. In Section 6.3.2, we visualize and discuss which edges and regions of the infant connectomes BrainNetCNN determined to be most important for predicting age.

Neurodevelopmental Outcome Prediction

We explored the more challenging outcome prediction task using different configurations of BrainNetCNN and competing methods (see Table 6.4). We compared the ANN models (i.e., FC and BrainNetCNN models) to linear regressors trained on features from **i)** the raw edge weights (Raw Edges), **ii)** 30 principal components of the edge features using PCA (PCA30), **iii)** high-level network features (Network), as used in Chapter 4, and **iv)** 6 clinically relevant metadata features (Clinical) including age at birth, age at scan, gender and ratings of brain white matter injury [182], ventriculomegaly [55] and intraventricular hemorrhaging [206] that are used by clinicians to assess risks to preterm infants' neurodevelopmental outcomes. As with the size of the last layer of BrainNetCNN, we chose 30 PCA features in order to provide the most direct comparison to Chapter 4.

Table 6.4: Correlation, r, corresponding p-values, MAE and standard deviation of absolute error (SDAE) between true and predicted Bayley-III motor and cognitive scores. Results for different configurations of BrainNetCNN (with different subsets of the layers shown in Fig. 6.1) and for competing models trained on different features. Our proposed, full BrainNetCNN model with one E2E layer for motor score and two E2E layers for cognitive outperform all other methods in terms of correlation.

		Layers	Motor				Cognitive			
	Model		r	p	MAE	SDAE	r	p	MAE	SDAE
competing	Clinical	-	0.106	0.170	16.139	13.737	0.086	0.271	15.339	12.053
	Network	-	0.227	0.003	13.345	9.761	0.143	0.064	13.564	9.722
	PCA30	-	0.181	0.019	12.186	8.259	0.069	0.374	11.682	8.809
	Raw Edges	7	0.176	0.023	27.399	27.273	0.063	0.420	27.502	26.529
	FC30net	6,7	0.231	0.003	10.915	8.075	0.158	0.041	10.583	8.572
	FC90net	5,6,7	0.237	0.002	11.142	7.986	0.169	0.029	10.545	8.631
proposed	E2Nnet	3,4,5,6,7	0.271	0.0004	11.095	7.797	0.154	0.046	10.845	8.902
	E2Enet	1,3,4,5,6,7	0.281	0.0002	11.506	7.833	0.182	0.018	11.132	8.964
	E2Nnet-sml	3,4,7	0.263	0.0006	**10.640**	8.075	0.162	0.0355	**10.493**	8.459
	E2Enet-sml	1,3,4,7	**0.310**	**<0.0001**	10.761	7.734	0.174	0.0239	11.231	**8.424**
	2E2Enet-sml	1,2,3,4,7	0.290	0.0001	11.153	**7.686**	**0.188**	**0.0148**	11.077	8.574

Table 6.4 reports MAE, SDAE, correlations and correlation p-values between ground-truth and predicted scores. The statistical significance ($p < 0.05$), reports the very small likelihood that the positive correlation obtained is coincidental.

In terms of MAE, many models performed similarly well over motor and cognitive outcomes. PCA30 performed nearly as well as the neural network based models which all achieved average absolute errors of $< 11\%$ (based on a range of scores between 50 and 155). This result, alone, appears to suggest that the simplest models can perform with similar accuracy to more complex models. However, the correlation results contradict this and suggest that the PCA model has actually underfit the data, predicting a similar output for every input, resulting in comparatively low r values.

Different configurations of our BrainNetCNN produce the highest prediction correlation values for both motor and cognitive scores. Despite having the same number of trainable model parameters as FC90net (and significantly less parameters than E2Nnet and E2Enet) the E2Enet-sml model results in the highest motor correlation. Similarly for cognitive scores, a model with an additional E2E layer, 2E2net-sml, attains the highest prediction correlation. The E2Nnet-sml yields the lowest MAE for both motor and cognitive scores.

Paired t-tests were used to check the significance of the improvement of the BrainNetCNN models over FC90net, the next best model. To do this, 1000 random subsets of 56 instances (i.e., the size of each fold) were selected. For each model, the correlation between scores predicted by that model and the ground truth scores were computed within each subset. (Note that for all models, the distributions of correlation values across the 1000 subsets were found to be normal using Kolmogorov-Smirnov tests.) Each paired t-test was performed between a pair of models with the null hypothesis that the mean of the distri-

bution of correlation values were equal. The paired t-tests showed that all models with an E2E layer performed significantly better, on average, than the FC90net model on both prediction tasks with $p < 0.05$ except for the E2Enet-sml model which did not perform significantly better at predicting cognitive scores. For the 2E2Enet-sml model, correlations improved over FC90net an average of 8.44% for motor scores and 10.4% for cognitive scores.

To ensure that BrainNetCNN was not consistently predicting too high or too low (i.e., prediction bias), a t-test on the prediction errors of E2Enet with respect to each score type was performed. The mean difference between predicted and ground truth values for cognitive and motor scores were not found to be statistically significantly different from zero (p-values of 0.6817 and 0.9731 respectively), meaning that our model was unbiased.

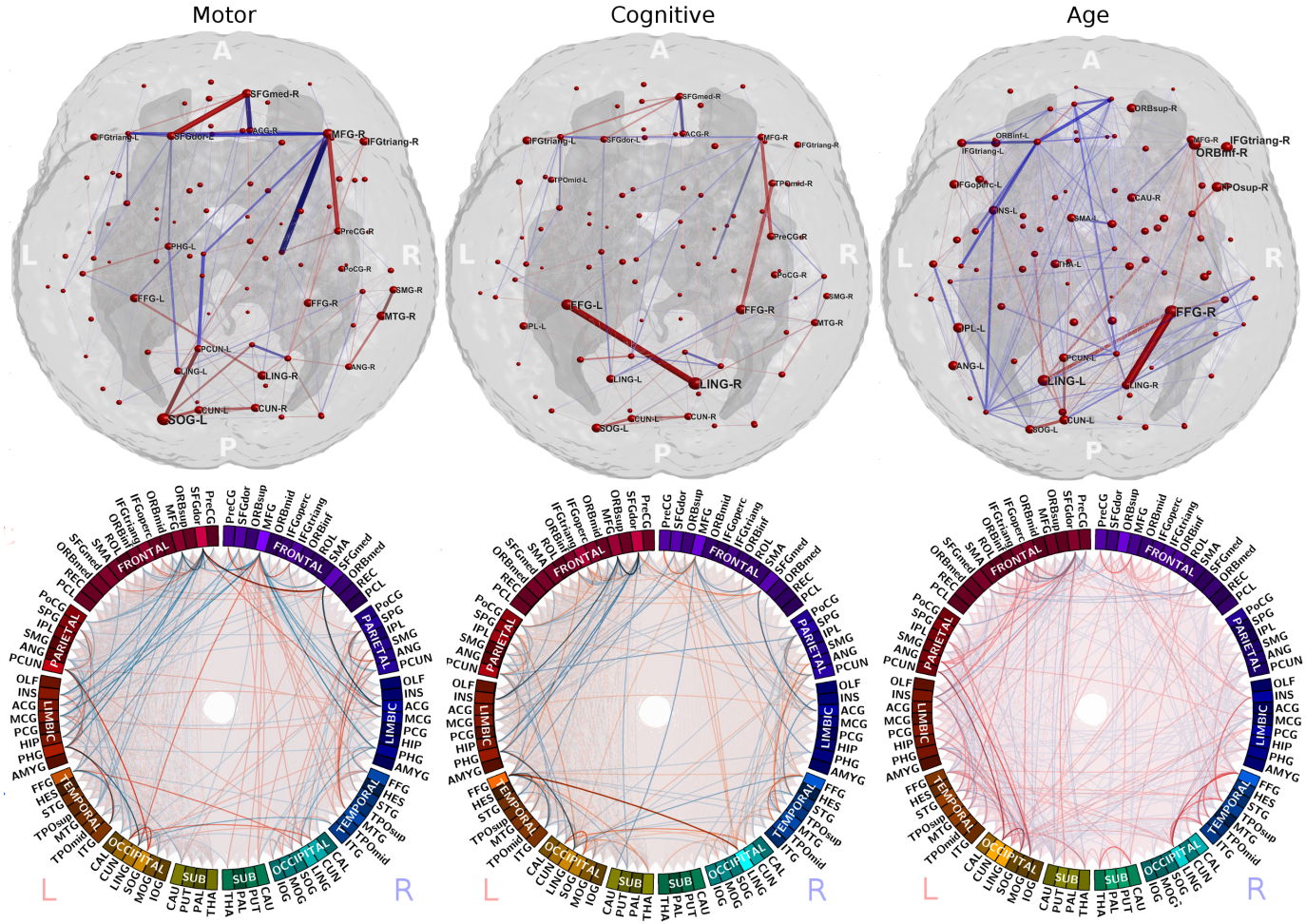

Figure 6.6: Connections learned by BrainNetCNN to be most predictive of outcomes and ages. **Top Row**: Edges with positive (red) and negative (blue) partial derivatives with respect to motor outcomes (left), cognitive outcomes (middle) and ages (right). Edge thickness and opacity represent the magnitude of each partial derivative. Very small magnitudes (< 0.001) were omitted for clarity. Node sizes represent the sum of partial derivative magnitudes of all neighbouring edges with positive derivatives. **Bottom row**: The same partial derivatives mapped on to Circos ideograms. Brightness of the color of the regions in each ring denotes the sum of positive partial derivative magnitudes.

Maps of Predictive Edges

In order to uncover which connections were learned by BrainNetCNN to be predictive of age, cognitive outcome and motor outcome, we used the method of Simonyan et al. [252], which computes the partial derivatives of the outputs of an ANN with respect to the input features. For each outcome y_s (i.e., either age or motor or cognitive scores), Simonyan et al.'s method computes $\frac{\partial y_s}{\partial A_{i,j}^{1,m}}$ for every input edge (i, j). In Fig. 6.6, the partial derivatives of motor and cognitive scores, averaged over the entire dataset, are plotted for all connectome edges, both spatially on line segments connecting centroids of the UNC atlas regions and schematically in Circos plots [156]. A complete list of region names corresponding to the region codes used in the Circos plots can be found in Chapter 3.

While many of the partial derivatives are positive (red) indicating connections that, when strong (i.e., high tract count), contribute to better outcomes there are also many negative partial derivatives (blue). We see that many brain connections (edges) from the right middle frontal gyrus (MFG) are selected as being predictive of positive outcomes for both motor and cognitive scores. The left precuneus (PCUN), fusiform gyrus (FFG), superior frontal gyrus (SFGdor) and right lingual gyrus (LING) also appear to be prominent hubs of important connections for both scores. For motor scores, the connection between the two superior frontal gyri appears to be of particular importance. In contrast, the connection between the left FFG and right LING is highlighted as being relatively more important for cognitive scores than for motor scores. However, there is considerable overlap between the two sets of edges.

Compared to the sets of edges found to be important for predicting neurodevelopmental outcomes, those found to be important for predicting age are much more widely distributed across the brain network (Fig. 6.6). Only the connection between the right LING and the right FFG appears to stand out as being a particularly strong predictor. We discuss possible anatomical reasons for these observations below.

6.4 Discussion

Broadly, the proposed BrainNetCNN performed well, predicting motor and cognitive scores with the highest correlations to the ground truth scores. Furthermore, it was found that, with respect to most accuracy measures, our convolution based models (e.g., E2Enet-sml, 2E2Enet-sml) were able to outperform other models without relying on the large fully connected layers. This increased accuracy was found for both real connectome data and carefully controlled phantom data. These results validate that our novel E2E, E2N and N2G filters, are able to learn important structures for prediction with a relatively small number of parameters. As well, it suggests that an alternative to learning larger models

with more layers is to employ convolutional layers that leverage the topology of the input data.

It was also found that for prediction of cognitive scores, it was helpful to stack E2E layers as seen by the comparatively high correlation value for 2E2Enet-sml. This stacking of E2E layers enables learning of complex structural patterns while requiring the optimization of relatively few additional parameters.

When BrainNetCNN was used to predict age, it was found that prediction was more accurate for the younger infants. One factor that likely contributed to this result is that there are more scans of younger infants in our data set (60% of scans are below the age range mid point), which provided more training data for these cases. If true, it suggests that larger training sets could further improve prediction results.

Despite the discrepancy in prediction error between younger and older infants, our E2Enet-sml model was able to predict PMA with high accuracy. However, when predicted with the same model, the correlation values for neurodevelopmental outcome scores were relatively low (e.g., 0.310 for motor scores versus 0.866 for age). While statistically significant, these values for prediction of outcome scores entail only weak to moderate correlations. Nevertheless, note that the correlation values and relative improvement of BrainNetCNN over FC models were only slightly lower for this real data than for the simulated diffuse injury phantoms. Fig. 6.4 (right), especially as compared to Fig. 6.3 (center), gives a sense of the level of difficulty of the prediction problem to result in correlation values in this range.

A number of factors contribute to the increased difficulty of predicting outcome scores compared to predicting age. Probably the most significant factor is the $\approx$18 months of intervening time between scan and Bayley-III assessment. This task of predicting neurodevelopmental outcomes of infants 18 months into future is made more difficult by the fact that, shortly after birth, infants are developing very rapidly and environmental and genetic factors will affect the course of this development. The infant brain may also be impacted by preterm birth and postnatal illness through mechanisms that do not alter DTI metrics of diffusivity. Furthermore, the amount of available data for training remains relatively small compared to the dimensionality of the input networks, especially for the minority of cases with very high and low outcome scores. While data augmentation can be use to expand and balance a dataset, it is not a substitute for real data.

In all of the experiments on real connectome data, we trained each ANN model on motor and cognitive outcomes jointly. This was done because the scores are strongly correlated ($r = 0.68$ in our dataset) and we expected that prediction of two outcomes would help regularize our underdetermined models. We did explore training motor and cognitive outcomes separately but found little difference in our metrics compared to joint training. Given that joint training requires only a single model trained for both tasks, significantly reducing the computational burden and training times, it was our adopted choice for all

experiments. While it's possible that low cognitive outcomes and low motor outcomes do not share a common aetiology, the 30 high-level features of the last layer of the proposed models provides these models with high flexibility to identify injury patterns of different types.

In comparison to our BrainNetCNN learned features, the network measure features used in Chapter 4 performed poorly. This was somewhat surprising, as they were shown to perform well on the similar task of motor outcome abnormality classification. However, in that work, these network measure features were combined with several meta-data features, including information about age, gender and brain white matter injury grade, then dimensionality reduced using PCA before performing prediction. It is possible then that the network measures are much more powerful only in combination with meta-data. In contrast, our prediction model based on regularized subnetworks, presented in Chapter 5, acheived better correlations between predicted and ground truth motor (0.442 versus 0.310 here) and cognitive (0.343 versus 0.188 here) outcomes than BrainNetCNN. The highly regularized linear subnetwork model is perhaps better suited to the relatively small dataset than the deep model presented here. Note however that the performance difference may also be in part due to the difference in cross-validation schemes in the two chapters. Unfortunately, we could not match the cross-validation scheme used in Chapter 5 here, as the 1000-fold cross validation would require 1000 BrainNetCNN models to be trained, which was not computationally feasible.

Generally, prediction results were more accurate for motor scores than for cognitive scores. It is likely that this is mainly due to motor scores having a higher accuracy of assessment at 18 months of age; our ability to accurately assess cognitive function improves over time, as more functions can be assessed with age. The disparity in prediction accuracy could also be partly due to motor scores having a simpler dependence on the input features compared to the cognitive scores. This is plausible since a few particular regions (e.g., primary motor and premotor cortices) are well known to be responsible for many motor functions [179] whereas cognitive function likely relies on a complex network of many regions (which may be unique per cognitive task) [38]. Furthermore, compared to motor outcomes, cognitive outcomes may be more sensitive to environmental factors not captured by imaging such as maternal education and socioeconomic status [115].

We regard our work as only a proof of concept, showing that filters designed to leverage the structure of the input brain networks can outperform other models on this prediction task. Consequently, as with other non-medical applications of deep learning, given the large number of parameters to be learned, the full potential of CNNs like BrainNetCNN would be realized when applied to applications with much larger neuroimaging datasets, which in turn will require further time and effort to explore a wide array of architectures and parameter settings. To accelerate this exploration, we make our BrainNetCNN publicly

available, downloadable at `http://BrainNetCNN.cs.sfu.ca`. Additionally, here we identify three important avenues for future investigation.

First, while it was found that our connectome based models performed better than the models learned from clinical features, it is likely that these features may contribute complementary information to that derived from the connectomes. If the features from both sources could be combined intelligently, the prediction accuracies would likely increase.

Second, as noted, a lack of training data is a major challenge for complex models like the ones proposed. However, other works have shown that transfer learning can occur after pre-training a deep convolutional neural network over larger, similar datasets [80]. Since diffusion tensor images of preterm infants are difficult to acquire, perhaps pre-training BrainNetCNN with connectomes from infants born at term or other similar data could improve its predictive ability.

Finally, how to generate the most realistic synthetic training data is still an open research question. We were motivated to attempt to perform data augmentation here because it was clear that even with convolutional filters, the number of parameters to learn in a deep network is high. It is possible that a more advanced data augmentation strategy than SMOTE could perform better. We plan to extend LSI (Chapter 4) from the context of binary classification to regression in the hope that it would perform better than SMOTE for this sparsely sampled, high-dimensional data. Another more dMRI-specific way to augment our dataset would be to use a bootstrap approach to tensor fitting and/or tractography, in order to generate multiple connectomes per image [143, 267]. We expect that by improving our approach in these ways, we will move towards achieving clinically useful predictive power.

When visualizing which edges BrainNetCNN selected as most predictive of positive cognitive and motor outcomes, it was found that many edges were common to both tasks. This is not surprising since, as mentioned above the two scores are well correlated and since BrainNetCNN was trained to predict both scores simultaneously. However, it may also indicate that some of these common connections in the brain are ones which are at higher risk for damage from the external factors that can lead to poor neurodevelopmental outcomes (e.g., white matter injury and infection) and thus are good common predictors of healthy outcomes. The right middle-frontal gyrus (MFG), in particular, was connected to many strong predictors of both outcomes (Fig. 6.6). This region is known to be associated with spatial memory [163], recognition and recall [236], among other functions, and so may be of particular importance for high Bayley-III scores. However, we note that 18 month outcomes have limited sensitivity to distinguish specific motor and cognitive skills. A longer term follow-up of this cohort is underway and will be helpful to examine specificity of these connections.

Fig. 6.6 also showed that the most predictive connections of both outcomes had clear laterality. Ratnarajah et al. found asymmetric functional specializations in the structure of

the neonatal connectome [222]. Our finding of laterality may then be due to connections between specific asymmetrical functional regions of the brain that are important for the Bayley-III cognitive and motor tests.

In terms of motor outcomes specifically, we found that the right precentral gyri (PreCG) was highly predictive of motor function, which is expected since the PreCG houses the primary motor cortex. Similarly, the premotor cortex is located, at least in part, within the superior frontal gyri (SFGdor), which were found to be connected to many strongly predictive edges, especially in the left hemisphere. One connection very predictive of cognitive outcomes and not motor outcomes was that between the left FFG and right LING. Both regions have been found to be associated with reasoning about sequences of events [49], however exactly why this particular link is important for prediction of the cognitive outcomes, is unclear. Again, a longer term follow-up may help answer these kinds of questions.

Edges found to be important for prediction of PMA were much more widespread across the brain network compared to those for predicting neurodevelopmental outcomes. This is expected since the whole brain is developing during this early period of development (i.e., many connections changing with age) whereas motor or cognitive functions depend predominantly on specific subnetworks [31]. One connection that stood out as being especially positively predictive of age was between the right LING and FFG. This result is consistent with our analysis of the development of healthy preterm infants in Chapter 3.

Generally, extracting the important features from the trained network provides candidate regions and connections for further investigation. This is especially important given the complexity of the brain and what remains to be fully understood about its function and development.

6.5 Conclusions

In this chapter we presented BrainNetCNN, the first CNN regressor for connectome data. We introduced three specialized convolutional layer types, designed to leverage the structure inherent in weighted brain networks. We first demonstrated the ability of our framework to learn multiple independent injury patterns to brain networks by predicting the input parameters of each instance in a realistic phantom dataset. We then tested BrainNetCNN on a set of 168 preterm infant brain networks and showed that our method was able to predict Bayely cognitive and motor scores 18 months into the future. Cognitive and motor scores predicted by BrainNetCNN had significantly higher correlations to the ground truth scores than those predicted by comparable methods (though it did not outperform the subnetwork model presented in Chapter 5). Finally, those edges that were learned by BrainNetCNN to be important for each neurodevelopmental outcome were visualized. We found that, as expected, connections from the premotor and primary motor cortices were

found to be predictive of better motor outcomes. We also found a general asymmetry in the important connections consistent with other reports in the literature.

Chapter 7

Conclusions and Future Directions

7.1 Summary and Conclusions

Through the studies presented in this thesis, we explored the early development of the preterm infant brain network and then used that brain network representation in novel machine learning frameworks to predict neurodevelopmental outcomes. To begin, analysis of normative preterm infant brain networks both allowed us to observe a number of interesting aspects of early development (e.g., increasing small-worldness with age) and also confirmed our hypothesis that global network measures of the structural connectome may be sensitive biomarkers of neurodevelopment (Chapter 3).

With these insights, we used the same network measures as features for training a machine learning model, designed to predict neurodevelopmental motor outcomes (Chapter 4). By introducing a novel data augmentation technique (LSI) to correct for the severe normal/abnormal class imbalance in our dataset, we found that we were able to predict neuromotor outcomes 18 months into the future with $> 70\%$ accuracy.

One drawback of using global network measures, however, is that it is impossible to determine which specific regions and connections in the brain were most predictive of the outcome scores. We addressed this issue by formulating a new learning model specifically designed to extract the most predictive subnetwork of edges in the brain network (Chapter 5). This subnetwork model outperformed the initial network-measure-based model on neuromotor outcome prediction and, for the first time, enabled predicition of cognitive outcomes (i.e., a much more difficult task), while exposing those regions and connections in the brain that were most important for prediciting each outcome.

Despite the success of this model, we recognized that a limiting factor might be its linearity, since we expect the relationship between brain network features (acquired shortly after birth) and 18-month neurodevelopmental outcomes to be nonlinear. Thus, in our final contribution in this thesis, we devised the first stacked, nonlinear learning model for predicting outcomes (Chapter 6). Our proposed BrainNetCNN model was adapted from

the highly successful convolutional neural network class of models by introducing special convolutional layer types designed for brain network data. This model was also able to predict cognitive and motor neurodevelopmental outcomes and outperformed comparable models with the same number of learnable parameters. However, due to the complexity of the model and the relatively small dataset available for validation, this model did not outperform the simpler subnetwork model introduced in Chapter 5. Nevertheless, as was reported in Chapter 2, connectome dataset sizes are growing and we expect that with larger training datasets, deep models like our BrainNetCNN will eventually outperform simpler models that are unable to capture nonlinear relationships in the data.

While our three proposed learning models represent promising progress for the specific application of predicting neurodevelopmental outcomes of preterm infants, these models are generally applicable to a much broader set of prediction problems over connectome data (many of which are discussed in detail in Chapter 2). Furthermore, the network measure features used in Chapter 4, the novel prior terms proposed in Chapter 5 and the E2E, E2N and N2G layers proposed in Chapter 6 could generally be applied to any kind of network data, provided that each network is defined with weighted edges and is defined over a constant number of non-weighted nodes (e.g., protein-protein networks [177]).

Towards the original goal of predicition of neurodevelopmental outcomes for early intervention, the work in this thesis is only preliminary. At minimum, larger validation sets and clinical trials will be necessary before its clear if the methods presented here are viable for practical use.

7.2 Future Directions

Directly following the work in this thesis, there are a number of exciting possible directions of continued exploration. For instance, in Chapter 3, a linear model of brain network connectivities over age was built for infants betwen 27 and 45 weeks PMA and revealed how the brain network develops rapidly during this period. This model, however, was unable to analyse whether certain infants' brain networks develop more quickly than others and, if so, what factors contribute to these differences. One way to explore these questions would be to train a model to predict the age of each infant from its connectome and then compute a developmental delay, defined by subtracting the an infants predicted age from its true age [257]. It would then be interesting to examine if the computed developmental delay is correlated with a any clinical or anatomical variables (e.g., neurodevelopmental outcome scores or brain network edge connectivities) in order better understand the factors associated with neurodevelopmental delay in very preterm infants.

Another interesting direction would be to try to improve the performance of deep models like BrainNetCNN on small connectome datasets via regularization. The main challenge in training a deep model on a limited dataset is the large number of model parameters

that must be learned. This can amount to a highly overdetermined system with no unique solution causing over-fitting to training data and poor generalization. One approach to tackle this problem would be to combine BrainNetCNN with the subnetwork regularization terms proposed in Chapter 5 to produce a stacked, flexible learning model that is also highly regularized. The network backbone prior, connectivity prior and L1 sparsity prior terms could be added to the standard BrainNetCNN loss function and act on an additional input layer that would filter brain networks being passed in. This filtering would reduce noisy edges and encourage only anatomically plausible, well connected edges to be retained, likely improving performance.

Additionally, as was mentioned above, while the methods presented in this thesis are largely applicable to a wide variety of connectome data, they have only been validated on preterm infant connectomes. Thus, it would be fascinating to validate these methods on different brain network data. For instance, testing BrainNetCNN on the autism brain imaging data exchange (ABIDE) [78] connectome dataset, would be especially interesting as it provides an order of magnitude more samples than the preterm infant dataset (e.g., $\sim$1000 scans versus only 168) and may help to demonstrate the advantages of using a stacked learning model for prediction from connectome data.

Beyond these immediate follow-ups, the work in this thesis has helped to elucidate some broader challenges in 1) representing the brain as a network and also in 2) using connectome data for machine learning that give rise to exciting future research questions:

- How can brain network data from heterogeneous sources (i.e. different scanners, hospitals, protocols, preprocessing) be combined to form large databases for training deep(er) models?
- How else can deep learning models be properly trained when data is expensive or limited? Are data augmentation and/or model regularization the solution?
- How can learning models be trained with connectome data from multiple neuroimaging modalities (including diffusion MRI, functional MRI and structural MRI) for better predictions of disease?
- Can different scales of brain networks, including large scale whole-brain connectomes and fine scale neuronal networks be combined in a meaningful way that improves analysis of different structures and systems in the brain?
- What are the limitations of network representations of the brain and how can we improve or augment brain network data to overcome these limitations?
- Can new machine learning models better mimic the human brain's unmatched ability to learn and in doing so, teach us more about how we think and learn?

As we found in Chapter 2, the field of research into machine learning on brain network data is growing rapidly. The community is discovering that the network representation of the brain naturally coincides with the brain's structure and enables models that can predict a wide range of diseases, disorders and clinical variables. This is important given emerging health challenges such as the rising rate of preterm birth and the epidemic of neurodegenerative diseases (e.g. Alzheimer's, Parkinson's) affecting a growing aging population[5, 201]. Continuing to develop advanced analytical and predictive models to better understand the human brain and to better explain and predict disease states will be a major effort of the coming decades.

Bibliography

[1] A. Abraham, M. Milham, A. Di, and R. C. Craddock. Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. *NeuroImage*, 147:736–745, 2016.

[2] R. Akbani, S. Kwek, and N. Japkowicz. Applying support vector machines to imbalanced datasets. In *Machine Learning: ECML 2004*, pages 39–50. Springer, 2004.

[3] D. Alnæs, T. Kaufmann, G. Richard, E. P. Duff, M. H. Sneve, T. Endestad, J. E. Nordvik, O. A. Andreassen, S. M. Smith, and L. T. Westlye. Attentional load modulates large-scale functional brain connectivity beyond the core attention networks. *NeuroImage*, 109:260–272, 2015.

[4] D. G. Altman and J. M. Bland. Statistics notes: Diagnostic tests 2: predictive values. *British Medical Journal*, 309(6947):102, 1994.

[5] Alzheimer's Association. 2016 alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 12(4):459–509, 2016.

[6] A. Anderson, P. K. Douglas, W. T. Kerr, V. S. Haynes, A. L. Yuille, J. Xie, Y. N. Wu, J. a. Brown, and M. S. Cohen. Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD. *NeuroImage*, 102:207–219, 2014.

[7] J. S. Anderson, J. A. Nielsen, A. L. Froehlich, M. B. Dubray, T. J. Druzgal, A. N. Cariello, J. R. Cooperrider, B. A. Zielinski, C. Ravichandran, P. T. Fletcher, A. L. Alexander, E. D. Bigler, N. Lange, and J. E. Lainhart. Functional connectivity magnetic resonance imaging classification of autism. *Brain*, 134(12):3739–3751, 2011.

[8] M. Anjari, L. Srinivasan, J. M. Allsop, J. V. Hajnal, M. A. Rutherford, A. D. Edwards, and S. J. Counsell. Diffusion tensor imaging with tract-based spatial statistics reveals local white matter abnormalities in preterm infants. *NeuroImage*, 35:1021–1027, 2007.

[9] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander. Optics: Ordering points to identify the clustering structure. *ACM Sigmod Record*, pages 49–60, 1999.

[10] M. R. Arbabshirani, K. A. Kiehl, G. D. Pearlson, and V. D. Calhoun. Classification of schizophrenia patients based on resting-state functional network connectivity. *Frontiers in Neuroscience*, 7:133, 2013.

[11] S. Arslan, S. Parisot, and D. Rueckert. Joint Spectral Decomposition for the Parcellation of the Human Cerebral Cortex Using Resting-State fMRI. In *Clinical Nuclear Medicine*, volume 17, pages 85–97. 2015.

[12] V. Ashikh, G. Deshpande, D. Rangaprakash, and D. N. Dutt. Clustering of dynamic functional connectivity features obtained from functional Magnetic Resonance Imaging data. In *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 308–312. IEEE, aug 2015.

[13] H.-E. Assemlal, D. Tschumperlé, L. Brun, and K. Siddiqi. Recent advances in diffusion MRI modeling: angular and radial reconstruction. *Medical image analysis*, 15(4):369–396, 2011.

[14] S. A. Back and S. P. Miller. Brain injury in premature neonates: A primary cerebral dysmaturation disorder? *Annals of Neurology*, 75(4):469–486, 2014.

[15] G. Ball, P. Aljabar, T. Arichi, N. Tusor, D. Cox, N. Merchant, P. Nongena, J. V. Hajnal, A. D. Edwards, and S. J. Counsell. Machine-learning to characterise neonatal functional connectivity in the preterm brain. *NeuroImage*, 124:267–275, 2016.

[16] G. Ball, P. Aljabar, S. Zebari, N. Tusor, T. Arichi, N. Merchant, E. C. Robinson, E. Ogundipe, D. Rueckert, a. D. Edwards, and S. J. Counsell. Rich-club organization of the newborn human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 111(20):7456–61, 2014.

[17] G. Ball, J. P. Boardman, P. Aljabar, A. Pandit, T. Arichi, N. Merchant, D. Rueckert, a. D. Edwards, and S. J. Counsell. The influence of preterm birth on the developing thalamocortical connectome. *Cortex*, 49(6):1711–1721, 2013.

[18] G. Ball, L. Pazderova, A. Chew, N. Tusor, N. Merchant, T. Arichi, J. M. Allsop, F. M. Cowan, a. D. Edwards, and S. J. Counsell. Thalamocortical Connectivity Predicts Cognition in Children Born Preterm. *Cerebral Cortex*, pages 1–9, 2015.

[19] P. J. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of magnetic resonance*, 213(2):560–570, 2011.

[20] D. S. Bassett and E. Bullmore. Small-world brain networks. *The neuroscientist*, 12(6):512–523, 2006.

[21] D. S. Bassett and E. T. Bullmore. Human brain networks in health and disease. *Current Opinion in Neurology*, 22(4):340–347, 2009.

[22] D. S. Bassett, A. N. Khambati, and S. T. Grafton. Emerging frontiers of neuroengineering: A network science of brain connectivity. *Annual Review of Biomedical Engineering*, 19(1), 2017.

[23] D. S. Bassett and M. E. Lynall. Network methods to characterize brain structure and function. *Cognitive Neurosciences: The Biology of the Mind*, pages 1–27, 2013.

[24] D. S. Bassett, B. G. Nelson, B. A. Mueller, J. Camchong, and K. O. Lim. Altered resting state complexity in schizophrenia. *NeuroImage*, 59(3):2196–2207, 2012.

[25] M. Bastiani, N. J. Shah, R. Goebel, and A. Roebroeck. Human cortical connectome reconstruction from diffusion weighted MRI: the effect of tractography algorithm. *Neuroimage*, 62(3):1732–1749, 2012.

[26] N. Bayley. *Manual for the Bayley Scales of Infant Development*. Harcourt, San Antonio, 3rd edition, 2006.

[27] L. M. Bear. Early identification of infants at risk for developmental disabilities. *Pediatric Clinics of North America*, 51(3):685–701, 2004.

[28] C. Beaulieu. The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR in Biomedicine*, 15:435–455, 2002.

[29] Y. Benjamini and D. Yekutieli. False Discovery Rate-Adjusted Multiple Confidence Intervals for Selected Parameters. *Journal of the American Statistical Association*, 100(469):71–81, 2005.

[30] M. Bertero, T. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, 1988.

[31] R. F. Betzel, A. Griffa, A. Avena-Koenigsberger, J. Goñi, J.-P. Thiran, P. Hagmann, and O. Sporns. Multi-scale community organization of the human structural connectome and its relationship with resting-state functional connectivity. *Network Science*, 1(03):353–373, 2014.

[32] J. Bi and K. P. Bennett. Regression Error Characteristic Curves. *International Conference on Machine Learning (ICML)*, pages 43–50, 2003.

[33] S. L. Bonifacio, H. C. Glass, V. Chau, J. I. Berman, D. Xu, R. Brant, a. J. Barkovich, K. J. Poskitt, S. P. Miller, and D. M. Ferriero. Extreme premature birth is not associated with impaired development of brain microstructure. *Journal of Pediatrics*, 157(5):726–732.e1, 2010.

[34] B. G. Booth and G. Hamarneh. Consistent information content estimation for diffusion tensor MR images. In *Proceedings of IEEE Healthcare Informatics, Imaging and Systems Biology (IEEE HISB)*, pages 166–173, 2011.

[35] B. G. Booth and G. Hamarneh. Multi-region competitive tractography via graph-based random walks. In *Mathematical Methods in Biomedical Image Analysis (MMBIA), 2012 IEEE Workshop on*, pages 73–78, 2012.

[36] B. G. Booth and G. Hamarneh. Diffusion MRI for brain connectivity mapping and analysis. In A. Majumdar and R. K. Ward, editors, *MRI: Physics, Image Reconstruction, and Analysis*, chapter 7, pages 137–171. 2015.

[37] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[38] S. L. Bressler and V. Menon. Large-scale brain networks in cognition: emerging methods and principles. *Trends in cognitive sciences*, 14(6):277–290, 2010.

[39] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann. The balanced accuracy and its posterior distribution. In *Pattern recognition (ICPR), 2010 20th international conference on*, pages 3121–3124. IEEE, 2010.

[40] T. Brosch and R. Tam. Manifold learning of brain MRIs by deep learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8150 LNCS(PART 2):633–640, 2013.

[41] T. Brosch, L. Y. W. Tang, Y. Yoo, D. K. B. Li, A. Traboulsee, and R. Tam. Deep 3D Convolutional Encoder Networks With Shortcuts for Multiscale Feature Integration Applied to Multiple Sclerosis Lesion Segmentation. *IEEE Transactions on Medical Imaging (TMI)*, 35(5):1229–1239, 2016.

[42] C. J. Brown, B. G. Booth, and G. Hamarneh. K-confidence: Assessing uncertainty in tractography using K optimal paths. In *IEEE International Symposium on Biomedical Imaging (IEEE ISBI)*, pages 250–253, 2013.

[43] C. J. Brown and G. Hamarneh. Machine Learning on Human Connectome Data from MRI. *arXiv*, 2016.

[44] C. J. Brown, S. P. Miller, B. G. Booth, S. Andrews, V. Chau, K. J. Poskitt, and G. Hamarneh. Structural network analysis of brain development in young preterm neonates. *NeuroImage*, 101:667–680, 2014.

[45] C. J. Brown, S. P. Miller, B. G. Booth, K. J. Poskitt, V. Chau, A. R. Synnes, J. G. Zwicker, R. E. Grunau, and G. Hamarneh. Prediction of Motor Function in Very Preterm Infants Using Connectome Features and Local Synthetic Instances. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 69–76. Springer Berlin Heidelberg, 2015.

[46] C. J. Brown, S. P. Miller, B. G. Booth, J. G. Zwicker, R. E. Grunau, A. R. Synnes, V. Chau, and G. Hamarneh. Predictive Subnetwork Extraction with Structural Priors for Infant Connectomes. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 175–183. Springer International Publishing, 2016.

[47] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *Journal of Machine Learning Research*, 13(Jan):27–66, 2012.

[48] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun. Spectral Networks and Locally Connected Networks on Graphs. *arXiv:1312.6203*, (1312.6203):14, 2013.

[49] E. Brunet, Y. Sarfati, M.-C. Hardy-Baylé, and J. Decety. A pet investigation of the attribution of intentions with a nonverbal task. *Neuroimage*, 11(2):157–166, 2000.

[50] C. R. Buchanan, C. R. Pernet, K. J. Gorgolewski, A. J. Storkey, and M. E. Bastin. Test-retest reliability of structural brain networks from diffusion MRI. *NeuroImage*, 86:231–243, 2014.

[51] E. T. Bullmore, O. Sporns, and S. A. Solla. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews. Neuroscience*, 10(3):186–98, 2009.

[52] J. E. L. C P Howson, M V Kinney. Born too soon: The global action report on preterm birth. *World health organization. Geneva*, 2012.

[53] K. Caeyenberghs, A. Leemans, I. Leunissen, K. Michiels, and S. P. Swinnen. Topological correlations of structural and functional networks in patients with traumatic brain injury. *Frontiers in Human Neuroscience*, 7(Nov):726, 2013.

[54] E. Candes and T. Tao. The Dantzig selector: Statistical estimation when p is much larger than n. *Annals of Statistics*, 35(6):2313–2351, 2007.

[55] J. D. Cardoza, R. B. Goldstein, and R. A. Filly. Exclusion of fetal ventriculomegaly with a single measurement: the width of the lateral ventricular atrium. *Radiology*, 169(3):711–714, 1988.

[56] F. X. Castellanos, A. Di Martino, R. C. Craddock, A. D. Mehta, and M. P. Milham. Clinical applications of the functional connectome. *NeuroImage*, 80:527–540, 2013.

[57] L.-C. Chang, D. K. Jones, and C. Pierpaoli. RESTORE: Robust estimation of tensors by outlier rejection. *Magnetic Resonance in Medicine*, 53:1088–1095, 2005.

[58] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014.

[59] V. Chau, A. Synnes, R. E. Grunau, K. J. Poskitt, R. Brant, and S. P. Miller. Abnormal brain maturation in preterm neonates associated with adverse developmental outcomes. *Neurology*, 81(24):2082–2089, 2013.

[60] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(1):321–357, 2002.

[61] G. Chen, B. D. Ward, C. Xie, W. Li, Z. Wu, J. L. Jones, M. Franczak, P. Antuono, and S.-J. Li. Classification of Alzheimer Disease, Mild Cognitive Impairment, and Normal Cognitive Status with Large-Scale Network Analysis Based on Resting-State Functional MR Imaging. *Radiology*, 259(1):213–221, apr 2011.

[62] H. Chen, A. Iraji, X. Jiang, J. Lv, Z. Kou, and T. Liu. Longitudinal Analysis of Brain Recovery after Mild Traumatic Brain Injury Based on Groupwise Consistent Brain Network Clusters. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9349, pages 194–201. Springer International Publishing, 2015.

[63] H. Chen, K. Li, D. Zhu, L. Guo, X. Jiang, and T. Liu. Group-wise optimization and individualized prediction of structural connectomes. *IEEE International Symposium on Biomedical Imaging (IEEE ISBI)*, pages 742–745, 2014.

[64] H. Cheng, Y. Wang, J. Sheng, W. G. Kronenberger, V. P. Mathews, T. a. Hummer, and A. J. Saykin. Characteristics and variability of structural networks derived from diffusion tensor imaging. *NeuroImage*, 61(4):1153–1164, 2012.

[65] W. Cheng, X. Ji, J. Zhang, and J. Feng. Individual classification of ADHD patients by integrating multiscale neuroimaging markers and advanced pattern recognition techniques. *Frontiers in Systems Neuroscience*, 6(Aug):58, 2012.

[66] A. W. Chung, M. D. Schirmer, M. L. Krishna, G. Ball, P. Aljabar, a. D. Edwards, and G. Montana. Characterising brain network topologies: a dynamic analysis approach using heat kernels. *arXiv:1603.06790*, 2016.

[67] F. Chung. *Spectral Graph Theory*. Number no. 92 in CBMS Regional Conference Series. Conference Board of the Mathematical Sciences, 1997.

[68] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *NIPS*, pages 2843–2851, 2012.

[69] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *International Conference on Medical Image Computing and Computer Assisted Invervention (MICCAI)*, pages 411–418. Springer International Publishing, 2013.

[70] G. Collin and M. P. van den Heuvel. The ontogeny of the human connectome: development and dynamic changes of brain connectivity across the life span. *Neuroscientist*, 19(6):616–28, 2013.

[71] M.-A. Côté, G. Girard, A. Boré, E. Garyfallidis, J.-C. Houde, and M. Descoteaux. Tractometer: towards validation of tractography pipelines. *Medical Image Analysis*, 17(7):844–857, 2013.

[72] R. C. Craddock, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg. Disease state prediction from resting state functional connectivity. *Magnetic Resonance in Medicine*, 62(6):1619–1628, 2009.

[73] R. Cuingnet, H. Benali, and M. Chupin. Spatial and anatomical regularization of SVM for brain image analysis. *Advances in Neural Information Processing Systems*, pages 460–468, 2010.

[74] R. Cuingnet, C. Rosso, M. Chupin, S. Lehéricy, D. Dormont, H. Benali, Y. Samson, and O. Colliot. Spatial regularization of SVM for the detection of diffusion alterations associated with stroke outcome. *Medical Image Analysis*, 15(5):729–737, 2011.

[75] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1:886–893, 2005.

[76] M. a. de Reus, V. M. Saenger, R. S. Kahn, and M. P. van den Heuvel. An edge-centric perspective on the human connectome: link communities in the brain. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1653):20130527–, 2014.

[77] M. A. de Reus and M. P. van den Heuvel. The parcellation-based connectome: Limitations and extensions. *NeuroImage*, 80:397–404, 2013.

[78] A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, 19(6):659–667, 2014.

[79] L. Dodero, H. Q. Minh, M. S. Biagio, V. Murino, and D. Sona. Kernel-based classification for brain connectivity graphs on the riemannian manifold of positive definite matrices. *IEEE International Symposium on Biomedical Imaging (IEEE ISBI)*, pages 42–45, 2015.

[80] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. DeCAF: A deep convolutional activation feature for generic visual recognition. *ICML*, 32:647–655, 2014.

[81] N. U. F. Dosenbach, B. Nardos, A. L. Cohen, D. a. Fair, D. Power, J. a. Church, S. M. Nelson, G. S. Wig, A. C. Vogel, C. N. Lessov-schlaggar, K. A. Barnes, J. W. Dubis, E. Feczko, R. S. Coalson, J. R. P. Jr, D. M. Barch, S. E. Petersen, and B. L. Schlaggar. Prediction of Individua Brain Maturity Using fMRI. *Science*, 329(5997):1358–1361, 2011.

[82] B. Draganski and A. May. Training-induced structural changes in the adult human brain. *Behavioural Brain Research*, 192(1):137–142, 2008.

[83] A. Drobyshevsky, S.-K. Song, G. Gamkrelidze, A. M. Wyrwicz, M. Derrick, F. Meng, L. Li, X. Ji, B. Trommer, D. J. Beardsley, et al. Developmental changes in diffusion anisotropy coincide with immature oligodendrocyte progression and maturation of compound action potential. *The Journal of neuroscience*, 25(25):5988–5997, 2005.

[84] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, 2012.

[85] J. Dudink, J. L. Kerr, K. Paterson, and S. J. Counsell. Connecting the developing preterm brain. *Early Human Development*, 84(12):777–782, 2008.

[86] P. Dvorak and B. Menze. Structured prediction with convolutional neural networks for multimodal brain tumor segmentation. *Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge*, pages 13–24, 2015.

[87] R. Eglese. Simulated annealing: a tool for operational research. *European Journal of Operational Research*, 46(3):271–281, 1990.

[88] F. Fei, B. Jie, L. Wang, and D. Zhang. Discriminative subnetwork mining for multiple thresholded connectivity-networks-based classification of mild cognitive impairment. *Pattern Recognition in Neuroimaging, 2014 International Workshop on*, pages 1–4, 2014.

[89] A. H. Fielding and J. F. Bell. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(01):38–49, 1997.

[90] E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris, and R. Todd Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(October):1–11, 2015.

[91] M. R. Folio and R. R. Fewell. *Peabody Developmental Motor Scales*. PRO-Ed, 2nd edition, 2000.

[92] E. W. Forgy. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21:768–769, 1965.

[93] P. Fransson, U. Åden, M. Blennow, and H. Lagercrantz. The functional architecture of the infant brain as revealed by resting-state fmri. *Cerebral Cortex*, 21(1):145–154, 2011.

[94] P. Fransson, B. Skiöld, S. Horsch, A. Nordell, M. Blennow, H. Lagercrantz, and U. Aden. Resting-state networks in the infant brain. *Proceedings of the National Academy of Sciences of the United States of America*, 104(39):15531–15536, 2007.

[95] Y. Freund and R. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. *Computational learning theory*, 55(1):119–139, 1995.

[96] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[97] J. H. Friedman. Regularized Discriminant Analysis. *Journal of the American Statistical Association*, 84(405):165–175, 1989.

[98] O. Friman, G. Farneback, and C.-F. Westin. A bayesian approach for stochastic white matter tractography. *IEEE Transactions on Medical Imaging*, 25(8):965–978, 2006.

[99] K. Friston. Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS Biology*, 7(2):0220–0225, 2009.

[100] J. Galvis, A. Mezher, A. Ragothaman, J. E. Villalon-Reina, P. T. Fletcher, P. M. Thompson, and G. Prasad. Effects of EPI distortion correction pipelines on the connectome in Parkinson's Disease. *SPIE Medical Imaging*, 9784(3):97843D, 2016.

[101] J. Gama. Functional trees. *Machine Learning*, 55(3):219–250, 2004.

[102] H. Gao, C. Cai, J. Yan, L. Yan, J. G. Cortes, Y. Wang, F. Nie, J. West, A. Saykin, L. Shen, and H. Huang. Identifying Connectome Module Patterns via New Balanced Multi-graph Normalized Cut. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9349, pages 371–378. Springer International Publishing, 2015.

[103] W. Gao, W. Lin, Y. Chen, G. Gerig, J. Smith, V. Jewells, and J. Gilmore. Temporal and spatial development of axonal maturation and myelination of white matter in the developing brain. *American journal of neuroradiology*, 30(2):290–296, 2009.

[104] W. Gao, H. Zhu, K. S. Giovanello, J. K. Smith, D. Shen, J. H. Gilmore, and W. Lin. Evidence on the emergence of the brain's default network from 2-week-old to 2-year-old healthy pediatric subjects. *Proceedings of the National Academy of Sciences*, 106(16):6790–6795, 2009.

[105] D. Gellerup. *Discriminating Parkinson's Disease Using Functional Connectivity and Brain Network Analysis*. PhD thesis, 2016.

[106] Y. Ghanbari, A. R. Smith, R. T. Schultz, and R. Verma. Identifying group discriminative and age regressive sub-networks from DTI-based connectivity via a unified framework of non-negative matrix factorization and graph embedding. *Medical Image Analysis*, 18(8):1337–1348, 2014.

[107] A. Ghosh and R. Deriche. A survey of current trends in diffusion mri for structural brain connectivity. *Journal of neural engineering*, 13(1):011001, 2015.

[108] G. Girard, A. Daducci, K. Whittingstall, R. Deriche, D. Wassermann, and M. Descoteaux. Microstructure driven tractography in the human brain. In *Organization for Human Brain Mapping (OHBM).*, 2016.

[109] C. Godsil and G. F. Royle. *Algebraic graph theory*, volume 207 of *Graduate Texts in Mathematics*. Springer Science & Business Media, 2013.

[110] G. H. Golub, P. C. Hansen, and D. P. O'Leary. Tikhonov regularization and total least squares. *SIAM Journal on Matrix Analysis and Applications*, 21(1):185–194, 1999.

[111] G. Gong, Y. He, L. Concha, C. Lebel, D. W. Gross, A. C. Evans, and C. Beaulieu. Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cerebral Cortex*, 19(3):524–536, 2009.

[112] B. Graham. Spatially-sparse convolutional neural networks. *arXiv Preprint:1409.6070*, pages 1–13, 2014.

[113] C. Granger. Some recent development in a concept of causality. *Journal of Econometrics*, 39(1-2):199–211, 1988.

[114] L. Grosenick, B. Klingenberg, K. Katovich, B. Knutson, and J. E. Taylor. Interpretable whole-brain prediction analysis with GraphNet. *NeuroImage*, 72(2):304–321, 2013.

[115] R. E. Grunau, M. F. Whitfield, J. Petrie-Thomas, A. R. Synnes, I. L. Cepeda, A. Keidar, M. Rogers, M. MacKay, P. Hubber-Richard, and D. Johannesen. Neonatal pain, parenting stress and interaction, in relation to cognitive and motor development at 8 and 18 months in preterm infants. *Pain*, 143(1):138–146, 2009.

[116] H. Guo, X. Cao, Z. Liu, H. Li, J. Chen, and K. Zhang. Machine learning classifier using abnormal brain network topological metrics in major depressive disorder. *Neuroreport*, 23(17):1006–11, 2012.

[117] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1-3):389–422, 2002.

[118] P. Hagmann, M. Kurant, X. Gigandet, P. Thiran, V. J. Wedeen, R. Meuli, and J. P. Thiran. Mapping human whole-brain structural networks with diffusion MRI. *PLoS ONE*, 2(7), 2007.

[119] M. Hall and L. a. Smith. Feature Selection for Machine Learning : Comparing a Correlation-based Filter Approach to the Wrapper CFS : Correlation-based Feature. *International FLAIRS Conference*, 1999:235–239, 1999.

[120] K. Han, C. L. Mac Donald, A. M. Johnson, Y. Barnes, L. Wierzechowski, D. Zonies, J. Oh, S. Flaherty, R. Fang, M. E. Raichle, and D. L. Brody. Disrupted modular organization of resting-state cortical functional connectivity in U.S. military personnel following concussive 'mild' blast-related traumatic brain injury. *NeuroImage*, 84:76–96, 2014.

[121] D. R. Hardoon and J. Shawe-Taylor. Sparse canonical correlation analysis. *Machine Learning*, 83(3):331–353, 2011.

[122] M. Henaff, J. Bruna, and Y. Lecun. Deep Convolutional Networks on Graph-Structured Data. *arXiv*, pages 1–10, 2015.

[123] J. M. Henderson. "Connectomic surgery": diffusion tensor imaging (DTI) tractography as a targeting modality for surgical modulation of neural networks. *Frontiers in integrative neuroscience*, 6(April):15, 2012.

[124] G. E. Hinton and R. R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786):504–507, jul 2006.

[125] A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[126] C. J. Honey, C. J. Honey, O. Sporns, O. Sporns, L. Cammoun, L. Cammoun, X. Gigandet, X. Gigandet, J. P. Thiran, J. P. Thiran, R. Meuli, R. Meuli, P. Hagmann, and P. Hagmann. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, 106(6):2035–40, 2009.

[127] D. W. Hosmer Jr and S. Lemeshow. *Applied logistic regression*. John Wiley & Sons, 2004.

[128] P. S. Hüppi, S. E. Maier, S. Peled, G. P. Zientara, P. D. Barnes, F. A. Jolesz, and J. J. Volpe. Microstructural development of human newborn cerebral white matter assessed in vivo by diffusion tensor magnetic resonance imaging. *Pediatric research*, 44(4):584–590, 1998.

[129] P. R. Huttenlocher and A. S. Dabholkar. Regional differences in synaptogenesis in human cerebral cortex. *Journal of comparative Neurology*, 387(2):167–178, 1997.

[130] T. Iidaka. Resting state functional magnetic resonance imaging and neural network classified autism and control. *Cortex*, 63:55–67, 2015.

[131] N. Jahanshad, T. M. Nir, A. W. Toga, C. R. Jack, M. A. Bernstein, M. W. Weiner, and P. M. Thompson. Seemingly unrelated regression empowers detection of network failure in dementia. *Neurobiology of Aging*, 36(S1):S103–S112, 2015.

[132] N. Japkowicz and S. Stephen. The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 6(5):429–449, 2002.

[133] M. Jenkinson, P. Bannister, M. Brady, and S. Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002.

[134] X. Ji, W. Cheng, J. Zhang, T. Ge, L. Sun, Y. Wang, and J. Feng. Increased Coupling in the Saliency Network is the main cause/effect of Attention Deficit Hyperactivity Disorder. *arXiv:1112.3496*, 2011.

[135] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe : Convolutional Architecture for Fast Feature Embedding. *ACM Conference on Multimedia*, pages 675–678, 2014.

[136] S. Jiang, H. Xue, S. J. Counsell, M. Anjari, J. Allsop, M. A. Rutherford, D. Rueckert, and J. V. Hajnal. In-utero three dimension high resolution fetal brain diffusion tensor imaging. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 18–26. Springer International Publishing, 2007.

[137] X. Jiang, T. Zhang, Q. Zhao, J. Lu, L. Guo, and T. Liu. Fiber Connection Pattern-Guided Structured Sparse Representation of Whole-Brain fMRI Signals for Functional Network Inference. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9349, pages 133–141. Springer International Publishing, 2015.

[138] B. Jie, D. Shen, and D. Zhang. Brain Connectivity Hyper-Network for MCI Classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 8674 LNCS, pages 724–732. Springer International Publishing, 2014.

[139] B. Jie, D. Zhang, B. Cheng, and D. Shen. Manifold Regularized Multi-Task Feature Selection for Multi-Modality Classification in Alzheimer's Disease. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 275–283. Springer International Publishing, 2013.

[140] B. Jie, D. Zhang, W. Gao, Q. Wang, C. Y. Wee, and D. Shen. Integration of network topological and connectivity properties for neuroimaging classification. *IEEE Transactions on Biomedical Engineering*, 61(2):576–589, 2014.

[141] B. Jie, D. Zhang, C.-Y. Wee, and D. Shen. Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification. *Human Brain Mapping*, 35(7):2876–97, 2014.

[142] Y. Jin, C.-y. Wee, F. Shi, K.-h. Thung, P.-t. Yap, and D. Shen. Identification of Infants at Risk for Autism Using Multi-parameter Hierarchical White Matter Connectomes. In *International Workshop on Machine Learning in Medical Imaging (MLMI)*, volume 9352, pages 170–177. 2015.

[143] D. K. Jones. Tractography gone wild: probabilistic fibre tracking using the wild bootstrap with diffusion tensor mri. *IEEE transactions on medical imaging*, 27(9):1268–1274, 2008.

[144] D. K. Jones. Challenges and limitations of quantifying brain connectivity in vivo with diffusion MRI. *Imaging in Medicine*, 2(3):341–355, 2010.

[145] D. K. Jones and P. J. Basser. "squashing peanuts and smashing pumpkins": How noise distorts diffusion-weighted MR data. *Magnetic Resonance in Medicine*, 52:979–993, 2004.

[146] K. Kamiya, S. Amemiya, Y. Suzuki, N. Kunii, K. Kawai, H. Mori, A. Kunimatsu, N. Saito, S. Aoki, and K. Ohtomo. Machine Learning of DTI Structural Brain Connectomes for Lateralization of Temporal Lobe Epilepsy. *Magnetic Resonance in Medical Sciences (MRMS)*, 15(1):121–129, 2015.

[147] B. Karrer and M. E. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.

[148] T. Kaufmann, T. Elvsåshagen, D. Alnæs, N. Zak, P. Ø. Pedersen, L. B. Norbom, S. H. Quraishi, E. Tagliazucchi, H. Laufs, A. Bjørnerud, U. F. Malt, O. A. Andreassen, E. Roussos, E. P. Duff, S. M. Smith, I. R. Groote, and L. T. Westlye. The brain functional connectome is robustly altered by lack of sleep. *NeuroImage*, 127:324–332, feb 2016.

[149] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, (July):1–9, sep 2016.

[150] M. G. Kendall. Rank correlation methods. 1948.

[151] A. Khazaee, A. Ebrahimzadeh, and A. Babajani-Feremi. Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease. *Brain Imaging and Behavior*, 2015.

[152] A. Khazaee, A. Ebrahimzadeh, and A. Babajani-Feremi. Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory. *Clinical Neurophysiology*, 126(11):2132–2141, 2015.

[153] I. Kostović and N. Jovanov-Milošević. The development of cerebral connections during the first 20–45 weeks' gestation. In *Seminars in Fetal and Neonatal Medicine*, volume 11, pages 415–422. Elsevier, 2006.

[154] A. Krishnan, L. J. Williams, A. R. McIntosh, and H. Abdi. Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. *NeuroImage*, 56(2):455–475, 2011.

[155] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, pages 1097–1105, 2012.

[156] M. I. Krzywinski, J. E. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra. Circos: An information aesthetic for comparative genomics. *Genome Research*, 19(9):1639–1645, 2009.

[157] M. B. Kursa and W. R. Rudnicki. Feature Selection with the Boruta Package. *Journal Of Statistical Software*, 36(11):1–13, 2010.

[158] S. H. Kwon, L. Vasung, L. R. Ment, and P. S. Huppi. The Role of Neuroimaging in Predicting Neurodevelopmental Outcomes of Preterm Neonates. *Clinics in Perinatology*, 41(1):257–283, 2014.

[159] V. Latora and M. Marchiori. Efficient behavior of small-world networks. *Physical review letters*, 87(19):198701, 2001.

[160] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[161] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–91, 1999.

[162] T.-W. Lee. Independent component analysis. In *Independent Component Analysis*, pages 27–66. Springer, 1998.

[163] H.-C. Leung, J. C. Gore, and P. S. Goldman-Rakic. Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. *Journal of cognitive neuroscience*, 14(4):659–671, 2002.

[164] E. Levina and P. Bickel. Maximum likelihood estimation of intrinsic dimension. *Advances in Neural Information Processing Systems*, 17:777–784, 2004.

[165] F. Li, L. Tran, K.-h. Thung, S. Ji, D. Shen, and J. Li. Robust Deep Learning for Improved Classification of AD / MCI Patients. In *International Workshop on Machine Learning in Medical Imaging (MLMI)*, pages 240–247, 2014.

[166] H. Li, Z. Xue, T. M. Ellmore, R. E. Frye, and S. T. Wong. Identification of faulty DTI-based sub-networks in autism using network regularized SVM. *IEEE International Symposium on Biomedical Imaging (IEEE ISBI)*, 6:550–553, 2012.

[167] J. Li, Y. Jin, Y. Shi, I. D. Dinov, D. J. Wang, A. W. Toga, and P. M. Thompson. Voxelwise spectral diffusional connectivity and its applications to Alzheimer's disease and intelligence prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 8149 LNCS, pages 655–662, 2013.

[168] X. Li, D. Zhu, X. Jiang, C. Jin, X. Zhang, L. Guo, J. Zhang, X. Hu, L. Li, and T. Liu. Dynamic functional connectomics signatures for characterization and differentiation of PTSD patients. *Human Brain Mapping*, 35(4):1761–1778, 2014.

[169] X. Liang, J. Wang, C. Yan, N. Shu, K. Xu, G. Gong, and Y. He. Effects of different correlation metrics and preprocessing factors on small-world brain functional networks: A resting-state functional MRI study. *PLoS ONE*, 7(3), 2012.

[170] F. Liem, G. Varoquaux, J. Kynast, F. Beyer, S. Kharabian Masouleh, J. M. Huntenburg, L. Lampe, M. Rahim, A. Abraham, R. C. Craddock, S. Riedel-Heller, T. Luck, M. Loeffler, M. L. Schroeter, A. V. Witte, A. Villringer, and D. S. Margulies. Predicting brain-age from multimodal imaging data captures cognitive impairment. *NeuroImage*, 2016.

[171] F. Liu, H.-I. Suk, C.-Y. Wee, H. Chen, and D. Shen. High-Order Graph Matching Based Feature Selection for Alzheimer's Disease Identification. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 311–318, 2013.

[172] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng. Early Diagnosis OF Alzheimer's Disease With Deep Learning. In *IEEE ISBI*, pages 677–680, Beijing, 2014. IEEE.

[173] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng. Early diagnosis of Alzheimer's disease with deep learning. In *IEEE International Symposium on Biomedical Imaging (IEEE ISBI)*, pages 1015–1018, Beijing, apr 2014. IEEE.

[174] C.-Y. Lo, P.-N. Wang, K.-H. Chou, J. Wang, Y. He, and C.-P. Lin. Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in alzheimer's disease. *The Journal of Neuroscience*, 30(50):16876–16885, 2010.

[175] D. G. Lowe. Object recognition from local scale-invariant features. *IEEE International Conference on Computer Vision (ICCV)*, 2([8):1150–1157, 1999.

[176] K. Maier-Hein, P. Neher, J.-C. Houde, M.-A. Cote, E. Garyfallidis, J. Zhong, M. Chamberland, F.-C. Yeh, Y. C. Lin, Q. Ji, et al. Tractography-based connectomes are dominated by false-positive connections. *bioRxiv*, page 084137, 2016.

[177] A. Mashaghi, A. Ramezanpour, and V. Karimipour. Investigation of a protein complex network. *The European Physical Journal B-Condensed Matter and Complex Systems*, 41(1):113–121, 2004.

[178] P. M. Matthews, G. D. Honey, and E. T. Bullmore. Applications of fmri in translational medicine and clinical practice. *Nature Reviews Neuroscience*, 7(9):732–744, 2006.

[179] J. D. Meier, T. N. Aflalo, S. Kastner, and M. S. Graziano. Complex organization of human primary motor cortex: a high-resolution fmri study. *Journal of neurophysiology*, 100(4):1800–1812, 2008.

[180] A. Melbourne, Z. Eaton-Rosen, A. Bainbridge, G. S. Kendall, M. J. Cardoso, N. J. Robertson, N. Marlow, and S. Ourselin. Measurement of myelin in the preterm brain: Multi-compartment diffusion imaging and multi-component t2 relaxometry. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 336–344. Springer International Publishing, 2013.

[181] S. Micheloyannis, E. Pachou, C. J. Stam, M. Breakspear, P. Bitsios, M. Vourkas, S. Erimaki, and M. Zervakis. Small-world networks and disturbed functional connectivity in schizophrenia. *Schizophrenia Research*, 87(1-3):60–66, 2006.

[182] S. P. Miller, D. M. Ferriero, C. Leonard, R. Piecuch, D. V. Glidden, J. C. Partridge, M. Perez, P. Mukherjee, D. B. Vigneron, and a. J. Barkovich. Early brain injury in premature newborns detected with magnetic resonance imaging is associated with adverse early neurodevelopmental outcome. *Journal of Pediatrics*, 147(5):609–616, 2005.

[183] A. B. Mitnitski, J. E. Graham, A. J. Mogilner, and K. Rockwood. Frailty, fitness and late-life mortality in relation to chronological and biological age. *BMC geriatrics*, 2(1):1, 2002.

[184] J. Mitra, K.-k. Shen, S. Ghose, P. Bourgeat, J. Fripp, O. Salvado, K. Pannek, D. J. Taylor, J. L. Mathias, and S. Rose. Statistical machine learning to identify traumatic brain injury (TBI) from structural disconnections of white matter networks. *NeuroImage*, 129:247–259, 2016.

[185] F. Mokhtari, S. K. Bakhtiari, G. A. Hossein-Zadeh, and H. Soltanian-Zadeh. Discriminating between brain rest and attention states using fMRI connectivity graphs and subtree SVM. In D. R. Haynor and S. Ourselin, editors, *Medical Imaging*, volume 8314, page 83144C, feb 2012.

[186] D. Moyer, B. Gutman, G. Prasad, J. Faskowitz, G. Ver, and P. Thompson. Blockmodels for Connectome Analysis. *SPIE*, 9681:1–9, 2015.

[187] B. A. Mueller, K. O. Lim, L. Hemmy, and J. Camchong. Diffusion mri and its role in neuropsychology. *Neuropsychology review*, 25(3):250–271, 2015.

[188] B. C. Munsell, C.-Y. Wee, S. S. Keller, B. Weber, C. Elger, L. A. T. da Silva, T. Nesland, M. Styner, D. Shen, and L. Bonilha. Evaluation of machine learning algorithms for treatment outcome prediction in patients with epilepsy based on structural connectome data. *NeuroImage*, 118:219–230, 2015.

[189] B. C. Munsell, G. Wu, Y. Gao, N. Desisto, and M. Styner. Identifying Relationships in Functional and Structural Connectome Data Using a Hypergraph Learning Method. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9900, pages 9–17. Springer International Publishing, 2016.

[190] B. Mwangi, T. S. Tian, and J. C. Soares. A review of feature reduction techniques in Neuroimaging. *Neuroinformatics*, 12(2):229–244, 2014.

[191] J. J. Neil, S. I. Shiran, R. C. McKinstry, G. L. Schefft, A. Z. Snyder, C. R. Almli, E. Akbudak, J. A. Aronovitz, J. P. Miller, B. Lee, et al. Normal brain in human newborns: apparent diffusion coefficient and diffusion anisotropy measured by using diffusion tensor mr imaging. *Radiology*, 209(1):57–66, 1998.

[192] M. E. Newman. Fast algorithm for detecting community structure in networks. *Physical review E*, 69(6):066133, 2004.

[193] M. E. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[194] B. Ng, V. Siless, G. Varoquaux, J. B. Poline, B. Thirion, and R. Abugharbieh. Connectivity-informed sparse classifiers for fMRI brain decoding. In *International Workshop on Pattern Recognition in NeuroImaging (PRNI)*, pages 101–104, 2012.

[195] B. Ng, G. Varoquaux, J. B. Poline, M. Greicius, and B. Thirion. Transport on Riemannian manifold for connectivity-based brain decoding. *IEEE Transactions on Medical Imaging (TMI)*, 35(1):208–216, 2016.

[196] C. Nimsky, O. Ganslandt, P. Hastreiter, R. Wang, T. Benner, A. G. Sorensen, and R. Fahlbusch. Preoperative and intraoperative diffusion tensor imaging-based fiber tracking in glioma surgery. *Neurosurgery*, 56(1):130–138, 2005.

[197] T. Nir, N. Jahanshad, C. R. Jack, M. W. Weiner, A. W. Toga, and P. M. Thompson. Small world network measures predict white matter degeneration in patients with early-stage mild cognitive impairment. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, number Cc, pages 1405–1408. IEEE, may 2012.

[198] G. E. Noether. The wilcoxon two-sample test. In *Introduction to Statistics*, pages 103–127. Springer, 1991.

[199] A. L. Oberg and D. W. Mahoney. Linear mixed effects models. In *Topics in biostatistics*, pages 213–234. Springer, 2007.

[200] J.-P. Onnela, J. Saramäki, J. Kertész, and K. Kaski. Intensity and coherence of motifs in weighted complex networks. *Physical Review E*, 71(6):065103, 2005.

[201] W. H. Organization. Preterm birth fact sheet no. 363. `http://www.who.int/mediacentre/factsheets/fs363/en/`. Accessed 2015-03-03.

[202] J. P. Owen, E. Ziv, P. Bukshpun, N. Pojman, M. Wakahiro, J. I. Berman, T. P. Roberts, E. J. Friedman, E. H. Sherr, and P. Mukherjee. TestâĂŞRetest Reliability of Computational Network Measurements Derived from the Structural Connectome of the Human Brain. *Brain Connectivity*, 3(2):160–176, 2013.

[203] B. Pakkenberg and H. J. G. Gundersen. Neocortical neuron number in humans: effect of sex and age. *Journal of Comparative Neurology*, 384(2):312–320, 1997.

[204] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

[205] A. Pandit, E. Robinson, P. Aljabar, G. Ball, I. Gousias, Z. Wang, J. Hajnal, D. Rueckert, S. Counsell, G. Montana, et al. Whole-brain mapping of structural connectivity in infants reveals altered connection strength associated with growth and preterm birth. *Cerebral Cortex*, page bht086, 2013.

[206] B. R. K. H. Papile LA, Burstein J. Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1,500 gm. *Pediatrics*, 92:529–534, 1978.

[207] V. Pariyadath, E. A. Stein, and T. J. Ross. Machine learning classification of resting state functional connectivity predicts smoking status. *Frontiers in Human Neuroscience*, 8(June):425, 2014.

[208] B.-Y. Park, J. Seo, J. Yi, and H. Park. Structural and Functional Brain Connectivity of People with Obesity and Prediction of Body Mass Index Using Connectivity. *PLoS ONE*, 10(11):e0141376, 2015.

[209] G. J. Parker and D. C. Alexander. Probabilistic monte carlo based mapping of cerebral connections utilising whole-brain crossing fibre information. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 684–695. Springer, 2003.

[210] R. S. Patel, F. D. Bowman, and J. K. Rilling. A Bayesian approach to determining connectivity of the human brain. *Human Brain Mapping*, 27(3):267–276, 2006.

[211] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein. A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology*, 49(12):1373–1379, 1996.

[212] H. C. Peng, F. H. Long, and C. Ding. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.

[213] S. M. Plis, D. R. Hjelm, R. Slakhutdinov, E. a. Allen, H. J. Bockholt, J. D. Long, H. Johnson, J. Paulsen, J. Turner, and V. D. Calhoun. Deep learning for neuroimaging: A validation study. *Frontiers in Neuroscience*, 8(July):1–11, 2014.

[214] S. M. Plis, M. P. Weisend, E. Damaraju, T. Eichele, A. Mayer, V. P. Clark, T. Lane, and V. D. Calhoun. Effective connectivity analysis of fMRI and MEG data collected under identical paradigms. *Computers in Biology and Medicine*, 41(12):1156–1165, 2011.

[215] J. P. Pluim, J. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging (TMI)*, 22(8):986–1004, 2003.

[216] J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59(3):2142–2154, 2012.

[217] G. Prasad, S. H. Joshi, T. M. Nir, A. W. Toga, and P. M. Thompson. Brain connectivity and novel network measures for Alzheimer's disease classification. *Neurobiology of Aging*, 36(S1):S121–S131, 2015.

[218] G. Prasad, S. H. Joshi, and P. M. Thompson. Optimizing Brain Connectivity Networks for Disease Classification Using EPIC. *IEEE International Symposium on Biomedical Imaging (IEEE ISBI)*, 2014:834–837, 2014.

[219] J. R. Pruett, S. Kandala, S. Hoertel, A. Z. Snyder, J. T. Elison, T. Nishino, E. Feczko, N. U. F. Dosenbach, B. Nardos, J. D. Power, B. Adeyemo, K. N. Botteron, R. C. McKinstry, A. C. Evans, H. C. Hazlett, S. R. Dager, S. Paterson, R. T. Schultz, D. L. Collins, V. S. Fonov, M. Styner, G. Gerig, S. Das, P. Kostopoulos, J. N. Constantino, A. M. Estes, S. E. Petersen, B. L. Schlaggar, and J. Piven. Accurate age classification of 6 and 12 month-old infants based on resting-state functional connectivity magnetic resonance imaging data. *Developmental Cognitive Neuroscience*, 12:123–133, 2015.

[220] S. Qi, S. Meesters, K. Nicolay, B. M. ter Haar Romeny, and P. Ossenblok. The influence of construction methodology on structural brain network measures: A review. *Journal of Neuroscience Methods*, 253(500):170–182, 2015.

[221] A. Qiu, A. Lee, M. Tan, and M. K. Chung. Manifold learning on brain functional networks in aging. *Medical Image Analysis*, 20(1):52–60, 2015.

[222] N. Ratnarajah, A. Rifkin-Graboi, M. V. Fortier, Y. S. Chong, K. Kwek, S. M. Saw, K. M. Godfrey, P. D. Gluckman, M. J. Meaney, and A. Qiu. Structural connectivity asymmetry in the neonatal brain. *NeuroImage*, 75:195–202, 2013.

[223] C. Reitz and R. Mayeux. Alzheimer disease: epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochemical pharmacology*, 88(4):640–651, 2014.

[224] C. Reveley, A. K. Seth, C. Pierpaoli, A. C. Silva, D. Yu, R. C. Saunders, D. A. Leopold, and Q. Y. Frank. Superficial white matter fiber systems impede detection of long-range cortical connections in diffusion mr tractography. *Proceedings of the National Academy of Sciences*, 112(21):E2820–E2828, 2015.

[225] J. Richiardi, S. Achard, H. Bunke, and D. Van De Ville. Machine learning with brain graphs: Predictive modeling approaches for functional imaging in systems neuroscience. *IEEE Signal Processing Magazine*, 30(3):58–70, 2013.

[226] J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, and D. Van De Ville. Decoding brain states from fMRI connectivity graphs. *NeuroImage*, 56(2):616–626, 2011.

[227] J. Richiardi, M. Gschwind, S. Simioni, J.-M. Annoni, B. Greco, P. Hagmann, M. Schluep, P. Vuilleumier, and D. Van De Ville. Classifying minimally disabled multiple sclerosis patients from resting state functional connectivity. *NeuroImage*, 62(3):2021–33, 2012.

[228] E. C. Robinson, A. Hammers, A. Ericsson, a. D. Edwards, and D. Rueckert. Identifying population differences in whole-brain structural networks: A machine learning approach. *NeuroImage*, 50(3):910–919, 2010.

[229] G. K. Rohde, A. S. Barnett, P. J. Basser, S. Marenco, and C. Pierpaoli. Comprehensive Approach for Correction of Motion and Distortion in Diffusion-Weighted MRI. *Magnetic Resonance in Medicine*, 51(1):103–114, 2004.

[230] M. J. Rosa, L. Portugal, T. Hahn, A. J. Fallgatter, M. I. Garrido, J. Shawe-Taylor, and J. Mourao-Miranda. Sparse network-based models for patient classification using fMRI. *NeuroImage*, 105:493–506, 2015.

[231] M. J. Rosa, L. Portugal, J. Shawe-Taylor, and J. Mourao-Miranda. Sparse Network-Based Models for Patient Classification Using fMRI. *International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pages 66–69, 2013.

[232] H. R. Roth, A. Farag, L. Lu, E. B. Turkbey, and R. M. Summers. Deep convolutional networks for pancreas segmentation in CT imaging. In S. Ourselin and M. A. Styner, editors, *SPIE Medical Imaging*, pages 94131G–94131G, mar 2015.

[233] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[234] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010.

[235] J. D. Rudie, J. A. Brown, D. Beck-Pancer, L. M. Hernandez, E. L. Dennis, P. M. Thompson, S. Y. Bookheimer, and M. Dapretto. Altered functional and structural brain network organization in autism. *NeuroImage: Clinical*, 2(1):79–94, 2013.

[236] M. Rugg, P. Fletcher, C. Frith, R. Frackowiak, and R. Dolan. Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain*, 119(6):2073–2083, 1996.

[237] S. Ryali, K. Supekar, D. A. Abrams, and V. Menon. Sparse logistic regression for whole-brain classification of fMRI data. *NeuroImage*, 51(2):752–764, 2010.

[238] M. D. Sacchet, G. Prasad, L. C. Foland-Ross, P. M. Thompson, and I. H. Gotlib. Support vector machine classification of major depressive disorder using diffusion-weighted neuroimaging and graph theory. *Frontiers in Psychiatry*, 6(Feb):1–10, 2015.

[239] M. Schmidt. *Graphical model structure learning with l1-regularization*. PhD thesis, University of British Columbia (Vancouver), 2010.

[240] A. Serag, P. Aljabar, G. Ball, S. J. Counsell, J. P. Boardman, M. A. Rutherford, A. D. Edwards, J. V. Hajnal, and D. Rueckert. Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. *NeuroImage*, 59(3):2255 – 2265, 2012.

[241] K. Setsompop, R. Kimmlingen, E. Eberlein, T. Witzel, J. Cohen-Adad, J. A. McNab, B. Keil, M. D. Tisdall, P. Hoecht, P. Dietz, et al. Pushing the limits of in vivo diffusion mri for the human connectome project. *Neuroimage*, 80:220–233, 2013.

[242] D. Shahnazian, F. Mokhtari, and G. A. Hossein-Zadeh. A method based on the granger causality and graph kernels for discriminating resting state from attentional task. *International Conference on Biomedical Engineering (ICoBE)*, (February):83–88, 2012.

[243] Y.-H. Shao, W.-J. Chen, J.-J. Zhang, Z. Wang, and N.-Y. Deng. An efficient weighted lagrangian twin support vector machine for imbalanced data classification. *Pattern Recognition*, 47(9):3158–3167, 2014.

[244] H. Shen, L. Wang, Y. Liu, and D. Hu. Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. *NeuroImage*, 49(4):3110–3121, 2010.

[245] N. Shervashidze and K. Borgwardt. Fast subtree kernels on graphs. *Advances in Neural Information Processing Systems*, (December):1660–1668, 2008.

[246] N. Shervashidze, S. V. N. Vishwanathan, T. Petri, K. Mehlhorn, and K. Borgwardt. Efficient Graphlet Kernels for Large Graph Comparison. *International Conference on Artificial Intelligence and Statistics*, 5:488–495, 2008.

[247] F. Shi, P.-T. Yap, G. Wu, H. Jia, J. H. Gilmore, W. Lin, and D. Shen. Infant brain atlases from neonates to 1-and 2-year-olds. *PLoS One*, 6(4):e18746, 2011.

[248] N. Shu, Y. Liu, K. Li, Y. Duan, J. Wang, C. Yu, H. Dong, J. Ye, and Y. He. Diffusion tensor tractography reveals disrupted topological efficiency in white matter structural networks in multiple sclerosis. *Cerebral cortex*, 21(11):2565–2577, 2011.

[249] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

[250] D. I. Shuman, B. Ricaud, P. Vandergheynst, E. P. F, and D. L. Epfl. A Windowed Graph Fourier Transform. *Statistical Signal Processing Workshop (SSP)*, (3):133–136, 2012.

[251] C. N. Silver and W. P. Dunlap. Averaging Correlations: Expected Values and Bias in Combined Pearson r s and Fisher's z Transformations. *Journal of Applied Psychology*, 72(1):146–148, 1987.

[252] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.

[253] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *ICLR*, 2015.

[254] S. M. Smith. Fast robust automated brain extraction. *Human brain mapping*, 17(3):143–155, 2002.

[255] S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich. Network modelling methods for fMRI. *NeuroImage*, 54(2):875–891, 2011.

[256] S. M. Smith, D. Vidaurre, C. F. Beckmann, M. F. Glasser, M. Jenkinson, K. L. Miller, T. E. Nichols, E. C. Robinson, G. Salimi-Khorshidi, M. W. Woolrich, D. M. Barch, K. UÄ§urbil, and D. C. Van Essen. Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences*, 17(12):666–682, 2013.

[257] C. D. Smyser, N. U. Dosenbach, T. A. Smyser, A. Z. Snyder, C. E. Rogers, T. E. Inder, B. L. Schlaggar, and J. J. Neil. Prediction of brain maturity in infants using machine-learning algorithms. *NeuroImage*, 136:1–9, 2016.

[258] S. N. Sotiropoulos, S. Jbabdi, J. Xu, J. L. Andersson, S. Moeller, E. J. Auerbach, M. F. Glasser, M. Hernandez, G. Sapiro, M. Jenkinson, et al. Advances in diffusion mri acquisition and processing in the human connectome project. *NeuroImage*, 80:125–143, 2013.

[259] O. Sporns, G. Tononi, and R. Kötter. The human connectome: A structural description of the human brain. *PLoS Computational Biology*, 1(4):0245–0251, 2005.

[260] H.-I. Suk, S.-W. Lee, and D. Shen. Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis. *Brain Structure and Function*, 2015.

[261] H. I. Suk, S. W. Lee, and D. Shen. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Structure and Function*, 220(2):841–859, 2015.

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

[263] Y. Sun, Y. Chen, S. L. Collinson, A. Bezerianos, and K. Sim. Reduced Hemispheric Asymmetry of Brain Anatomical Networks Is Linked to Schizophrenia: A Connectome Study. *Cerebral cortex (New York, N.Y. : 1991)*, pages 1–14, 2015.

[264] K. Supekar, L. Q. Uddin, K. Prater, H. Amin, M. D. Greicius, and V. Menon. Development of functional and structural connectivity within the default mode network in young children. *NeuroImage*, 52(1):290 – 301, 2010.

[265] A. Sweet, A. Venkataraman, S. M. Stufflebeam, H. Liu, N. Tanaka, J. Madsen, and P. Golland. Detecting epileptic regions based on global brain connectivity patterns. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 16, pages 98–105. Springer International Publishing, 2013.

[266] E. Takahashi, R. D. Folkerth, A. M. Galaburda, and P. E. Grant. Emerging cerebral connectivity in the human fetal Brain: An MR tractography study. *Cerebral Cortex*, 22(2):455–464, 2012.

[267] H. Takemura, C. F. Caiafa, B. A. Wandell, and F. Pestilli. Ensemble tractography. *PLoS Comput Biol*, 12(2):e1004692, 2016.

[268] S. Takerkart, G. Auzias, B. Thirion, and L. Ralaivola. Graph-based inter-subject pattern analysis of fMRI data. *PLoS ONE*, 9(8), 2014.

[269] S. Takerkart, G. Auzias, B. Thirion, D. Schön, and L. Ralaivola. Graph-based inter-subject classification of local fMRI patterns. In *International Workshop on Machine Learning in Medical Imaging (MLMI)*, volume 7588 LNCS, pages 184–192, 2012.

[270] P. N. Taylor, C. E. Han, J.-C. Schoene-Bake, B. Weber, and M. Kaiser. Structural connectivity changes in temporal lobe epilepsy: Spatial features contribute more than topological measures. *NeuroImage: Clinical*, 8:322–328, 2015.

[271] K. Teffer, D. P. Buxhoeveden, C. D. Stimpson, A. J. Fobbs, S. J. Schapiro, W. B. Baze, M. J. McArthur, W. D. Hopkins, P. R. Hof, C. C. Sherwood, et al. Developmental changes in the spatial organization of neurons in the neocortex of humans and common chimpanzees. *Journal of Comparative Neurology*, 521(18):4249–4259, 2013.

[272] C. E. Thomaz, E. C. Kitani, and D. F. Gillies. A maximum uncertainty LDA-based approach for limited sample size problems - with application to face recognition. *Journal of the Brazilian Computer Society*, 12(2):7–18, 2006.

[273] B. Tunç, B. Solmaz, D. Parker, T. D. Satterthwaite, M. A. Elliott, M. E. Calkins, K. Ruparel, R. E. Gur, R. C. Gur, and R. Verma. Establishing a link between sex-related differences in the structural connectome and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 1688(371):20150111, 2016.

[274] W. Tutte. *Graph Theory*. Cambridge Mathematical Library. Cambridge University Press, 2001.

[275] O. Tymofiyeva, C. P. Hess, E. Ziv, P. N. Lee, H. C. Glass, D. M. Ferriero, A. J. Barkovich, and D. Xu. A dti-based template-free cortical connectome study of brain maturation. *PloS one*, 8(5):e63310, 2013.

[276] O. Tymofiyeva, C. P. Hess, E. Ziv, N. Tian, S. L. Bonifacio, P. S. McQuillen, D. M. Ferriero, A. J. Barkovich, and D. Xu. Towards the "baby connectome": mapping the structural connectivity of the newborn brain. *PloS one*, 7(2):e31029, 2012.

[277] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1):273–289, 2002.

[278] K. Uğurbil, J. Xu, E. J. Auerbach, S. Moeller, A. T. Vu, J. M. Duarte-Carvajalino, C. Lenglet, X. Wu, S. Schmitter, P. F. Van de Moortele, et al. Pushing spatial and temporal resolution for functional and diffusion mri in the human connectome project. *NeuroImage*, 80:80–104, 2013.

[279] M. P. van den Heuvel and H. E. Hulshoff Pol. Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20(8):519–534, 2010.

[280] M. P. van den Heuvel, K. J. Kersbergen, M. a. de Reus, K. Keunen, R. S. Kahn, F. Groenendaal, L. S. de Vries, and M. J. N. L. Benders. The Neonatal Connectome During Preterm Brain Development. *Cerebral cortex (New York, N.Y. : 1991)*, pages 1–14, 2014.

[281] M. P. Van Den Heuvel and O. Sporns. Rich-club organization of the human connectome. *The Journal of neuroscience*, 31(44):15775–15786, 2011.

[282] K. R. A. van Dijk, M. R. Sabuncu, and R. L. Buckner. The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage*, 59(1):431–438, 2012.

[283] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, and K. Ugurbil. The WU-Minn Human Connectome Project: An overview. *NeuroImage*, 80:62–79, 2013.

[284] D. Vanderweyen, B. C. Munsell, J. E. Mintzer, O. Mintzer, A. Gajadhar, X. Zhu, G. Wu, J. Joseph, and ADNI. Identifying Abnormal Network Alterations Common to Traumatic Brain Injury and Alzheimer's Disease Patients Using Functional Connectome Data. In *International Workshop on Machine Learning in Medical Imaging (MLMI)*, volume 27, pages 25–38, 2015.

[285] G. Varoquaux and B. Thirion. Detection of Brain Functional-Connectivity Difference in Post-stroke Patients Using Group-Level Covariance Modeling. *Brain*, pages 200–208, 2010.

[286] G. Varoquaux and B. Thirion. How machine learning is shaping cognitive neuroimaging. *GigaScience*, 3(1):28, 2014.

[287] S. Vega-Pons, P. Avesani, M. Andric, and U. Hasson. Classification of inter-subject fMRI data based on graph kernels. *International Workshop on Pattern Recognition in NeuroImaging (PRNI)*, 2014.

[288] S. V. N. Vishwanathan, K. Borgwardt, and N. Schraudolph. Fast Computation of Graph Kernels. *Advances in Neural Information Processing Systems 19*, 19:1449–1456, 2007.

[289] P. Walsh, M. Elsabbagh, P. Bolton, and I. Singh. In search of biomarkers for autism: scientific, social and ethical challenges. *Nature Reviews Neuroscience*, 12(10):603–612, 2011.

[290] B. Wang, A. M. Mezlini, F. Demir, M. Fiume, Z. Tu, M. Brudno, B. Haibe-Kains, and A. Goldenberg. Similarity network fusion for aggregating data types on a genomic scale. *Nature methods*, 11(3):333–7, 2014.

[291] L. Wang, F. Fei, B. Jie, and D. Zhang. Combining Multiple Network Features for Mild Cognitive Impairment Classification. *IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 996–1003, 2014.

[292] R. Wang, T. Benner, A. Sorensen, and V. Wedeen. Diffusion toolkit: a software package for diffusion imaging data processing and tractography. In *Proc Intl Soc Mag Reson Med*, volume 15, page 3720, 2007.

[293] Z. Wang, M. Fernández-Seara, D. C. Alsop, W.-C. Liu, J. F. Flax, A. A. Benasich, and J. A. Detre. Assessment of functional development in normal infant brain using arterial spin labeled perfusion mri. *Neuroimage*, 39(3):973–978, 2008.

[294] J. H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

[295] T. Watanabe, B. Tunc, D. Parker, J. Kim, and R. Verma. Label-Informed Non-negative Matrix Factorization with Manifold Regularization for Discriminative Sub-network Detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9900, pages 166–174. Springer International Publishing, 2016.

[296] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *nature*, 393(6684):440–442, 1998.

[297] C.-Y. Wee, Y. Li, B. Jie, Z.-W. Peng, and D. Shen. Identification of MCI Using Optimal Sparse MAR Modeled Effective Connectivity Networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 16, pages 319–327. Springer International Publishing, 2013.

[298] C.-Y. Wee, S. Yang, P.-T. Yap, and D. Shen. Temporally Dynamic Resting-State Functional Connectivity Networks for Early MCI Identification. *International Workshop on Machine Learning in Medical Imaging (MLMI)*, 8184:139–146, 2013.

[299] C.-Y. Wee, P. T. Yap, W. Li, K. Denny, J. N. Browndyke, G. G. Potter, K. a. Welsh-Bohmer, L. Wang, and D. Shen. Enriched white matter connectivity networks for accurate identification of MCI patients. *NeuroImage*, 54(3):1812–1822, 2011.

[300] C.-Y. Wee, P.-T. Yap, and D. Shen. Diagnosis of Autism Spectrum Disorders Using Temporally Distinct Resting-State Functional Connectivity Networks. *CNS Neuroscience & Therapeutics*, 22(3):212–219, mar 2016.

[301] C.-Y. Wee, P. T. Yap, D. Zhang, K. Denny, J. N. Browndyke, G. G. Potter, K. A. Welsh-Bohmer, L. Wang, and D. Shen. Identification of MCI individuals using structural and functional connectivity networks. *NeuroImage*, 59(3):2045–2056, 2012.

[302] J. Xie, S. Kelley, and B. K. Szymanski. Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Computing Surveys (csur)*, 45(4):43, 2013.

[303] M. Yang, L. Zhang, X. Feng, and D. Zhang. Fisher Discrimination Dictionary Learning for Sparse Representation. *International Conference on Computer Vision (ICCV)*, pages 543–550, 2011.

[304] P. Yang, Z. Zhang, B. B. Zhou, and A. Y. Zomaya. Sample subset optimization for classifying imbalanced biological data. In *Advances in Knowledge Discovery and Data Mining*, pages 333–344. Springer, 2011.

[305] S. Yang, Z. Lu, X. Shen, P. Wonka, and J. Ye. Fused Multiple Graphical Lasso. *SIAM Journal on Optimization*, 25(2):916–943, jan 2015.

[306] X. Yang, Y. Jin, X. Chen, H. Zhang, G. Li, and D. Shen. Functional Connectivity Network Fusion with Dynamic Thresholding for MCI Diagnosis. In *International Workshop on Machine Learning in Medical Imaging (MLMI)*, pages 246–253. Springer International Publishing, 2016.

[307] Z. Yang, S. Zhong, A. Carass, S. H. Ying, and J. L. Prince. Deep Learning for Cerebellar Ataxia Classification and Functional Score Regression. *International Workshop on Machine Learning in Medical Imaging (MLMI)*, 8679:68–76, 2014.

[308] Z. Yao, B. Hu, Y. Xie, P. Moore, and J. Zheng. A review of structural and functional brain networks: small world and atlas. *Brain Informatics*, 2(1):45–52, 2015.

[309] P. T. Yap, Y. Fan, Y. Chen, J. H. Gilmore, W. Lin, and D. Shen. Development trends of white matter connectivity in the first years of life. *PLoS ONE*, 6(9), 2011.

[310] B. Yoldemir, B. Ng, and R. Abugharbieh. Coupled Stable Overlapping Replicator Dynamics for Multimodal Brain Subnetwork Identification. In *Information Processing in Medical Imaging (IPMI)*, pages 770–781. 2015.

[311] B. Yoldemir, B. Ng, R. Abugharbieh, and S. Member. Stable Overlapping Replicator Dynamics for Brain Community Detection. *IEEE Transactions on Medical Imaging (TMI)*, 35(2):529 – 538, 2015.

[312] Y. Yoo, T. Brosch, A. Traboulsee, D. Li, and R. Tam. Deep Learning of Image Features from Unlabeled Data for Multiple Sclerosis Lesion Segmentation. In *International Workshop on Machine Learning in Medical Imaging (MLMI)*, pages 117–124. Springer International Publishing, 2014.

[313] Y. Yoo, T. Brosch, A. Traboulsee, D. K. Li, and R. Tam. Deep learning of image features from unlabeled data for multiple sclerosis lesion segmentation. In *Machine Learning in Medical Imaging*, pages 117–124. Springer, 2014.

[314] A. Zalesky, A. Fornito, and E. T. Bullmore. Network-based statistic: Identifying differences in brain networks. *NeuroImage*, 53(4):1197–1207, 2010.

[315] L. Zhan, Y. Liu, Y. Wang, J. Zhou, N. Jahanshad, J. Ye, and P. M. Thompson. Boosting brain connectome classification accuracy in Alzheimer's disease using higher-order singular value decomposition. *Frontiers in Neuroscience*, 9, 2015.

[316] H. Zhang. The optimality of naive bayes. *AAAI FLAIRS*, 1(2):3, 2004.

[317] S. Zhang, J. S. Ide, and C. S. R. Li. Resting-state functional connectivity of the medial superior frontal cortex. *Cerebral Cortex*, 22(1):99–111, 2012.

[318] S. Zhong, Y. He, and G. Gong. Convergence and divergence across construction methods for human brain white matter networks: An assessment based on individual differences. *Human Brain Mapping*, 35(5):1995–2013, 2015.

[319] L. Zhou, L. Wang, and P. Ogunbona. Discriminative Sparse Inverse Covariance Matrix: Application in Brain Functional Network Classification. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3097–3104, 2014.

[320] D. Zhu, K. Li, L. Guo, X. Jiang, T. Zhang, D. Zhang, H. Chen, F. Deng, C. Faraco, C. Jin, C. Y. Wee, Y. Yuan, P. Lv, Y. Yin, X. Hu, L. Duan, X. Hu, J. Han, L. Wang, D. Shen, L. S. Miller, L. Li, and T. Liu. DICCCOL: dense individualized and common connectivity-based cortical landmarks. *Cerebral Cortex*, 23(4):786–800, 2013.

[321] D. Zhu, K. Li, D. P. Terry, a. N. Puente, L. Wang, D. Shen, L. S. Miller, and T. Liu. Connectome-scale assessments of structural and functional connectivity in MCI. *Human Brain Mapping*, 35(7):2911–2923, 2014.

[322] D. Zhu, D. Shen, X. Jiang, T. Liu, and C. A. Imaging. Connectomics Signature For Characterization of Mild Cognitive Impairment and Schizophrenia. *IEEE International Symposium on Biomedical Imaging (IEEE ISBI)*, pages 325–328, 2014.

[323] E. Ziv, O. Tymofiyeva, D. M. Ferriero, a. J. Barkovich, C. P. Hess, and D. Xu. A machine learning approach to automated structural network analysis: Application to neonatal encephalopathy. *PLoS ONE*, 8(11), 2013.

[324] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B - Statistical Methodology*, 67(2):301–320, 2005.

[325] J. G. Zwicker, R. E. Grunau, E. Adams, V. Chau, R. Brant, K. J. Poskitt, A. Synnes, and S. P. Miller. Score for neonatal acute physiology-II and neonatal pain predict Corticospinal tract development in premature newborns. *Pediatric Neurology*, 48(2):123–129, 2013.