LEARNING FROM COMPLEX NEUROIMAGING DATASETS

Mahmoud Mostapha

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2020

Approved by:

Martin A. Styner

Stephen M. Pizer

Jan-Michael Frahm

Leonard McMillan

Mark D. Shen

ABSTRACT

Mahmoud Mostapha: Learning from Complex Neuroimaging Datasets
(Under the direction of Martin A. Styner)

Advancements in Magnetic Resonance Imaging (MRI) allowed for the early diagnosis of neurodevelopmental disorders and neurodegenerative diseases. Neuroanatomical abnormalities in the cerebral cortex are often investigated by examining group-level differences of brain morphometric measures extracted from highly-sampled cortical surfaces. However, group-level differences do not allow for individual-level outcome prediction critical for the application to clinical practice.

Despite the success of MRI-based deep learning frameworks, critical issues have been identified: *(1)* extracting accurate and reliable local features from the cortical surface, *(2)* determining a parsimonious subset of cortical features for correct disease diagnosis, *(3)* learning directly from a non-Euclidean high-dimensional feature space, *(4)* improving the robustness of multi-task multi-modal models, and *(5)* identifying anomalies in imbalanced and heterogeneous settings.

This dissertation describes novel methodological contributions to tackle the challenges above. First, I introduce a Laplacian-based method for quantifying local Extra-Axial Cerebrospinal Fluid (EA-CSF) from structural MRI. Next, I describe a deep learning approach for combining local EA-CSF with other morphometric cortical measures for early disease detection. Then, I propose a data-driven approach for extending convolutional learning to non-Euclidean manifolds such as cortical surfaces. I also present a unified framework for robust multi-task learning from imaging and non-imaging information. Finally, I propose a semi-supervised generative approach for the detection of samples from untrained classes in imbalanced and heterogeneous developmental datasets.

The proposed methodological contributions are evaluated by applying them to the early detection of Autism Spectrum Disorder (ASD) in the first year of the infant's life. Also, the aging human brain is examined in the context of studying different stages of Alzheimer's Disease (AD).

To my wife, family, and friends ...

ACKNOWLEDGMENTS

I want to express my sincere thanks to everyone who has made my time at UNC and in Chapel Hill so special. First, I thank my research advisor, Martin Styner, for his mentorship and support over these past several years, without which this work would not have been possible. I was fortunate to have Stephen Pizer as my academic advisor who helped me achieve all the doctoral milestones and inspired me with his outstanding passion and dedication in research. I also thank the other members of my dissertation committee. I thank Mark Shen for sharing his clinical expertise and collaboration. I also thank Jan-Michael Frahm and Leonard McMillan for their insightful suggestions on my research. I appreciate the financial support provided by the departments of Computer Science and Psychiatry, with special mention of Beatriz Paniagua for her support of my research through the SlicerSALT project. I would also like to thank the Infant Brain Imaging Study (IBIS) network, the Alzheimer's Disease Neuroimaging Initiative (ADNI), the Baby Connectome Project (BCP), and the Early Brain Development Study (EBDS) for kindly providing their data. I want to give a special mention to my collaborators and colleagues: Juan Prieto, Sun-Hyung Kim, Guorong Wu, Joseph Blocher, Mark Foster, Roza Vlasova, Ahsan Mahmood, Jared Vicory, Junpyo Hong, Ilwoo Lyu, and all other members of Neuro Image Research and Analysis Laboratories (NIRAL). I also want to acknowledge the support and collaboration of Mariappan Nadar, Boris Mailhe, Xiao Chen, Pascal Ceccaldi, Youngjin Yoo during my internship at Siemens Healthineers. Lastly, and most importantly, I would like to thank my wife, family, and friends for their endless love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACC	Accuracy
AD	Alzheimer's Disease
ADI-R	Autism Diagnostic Interview-Revised
ADNI	Alzheimer's Disease Neuroimaging Initiative
ADOS	Autism Diagnostic Observation Schedule-G
AEs	Autoencoders
AI	Artificial Intelligence
ASD	Autism Spectrum Disorder
AUC	Area Under the ROC Curve
BCP	Baby Connectome Project
CN	Control
CNNs	Convolutional Neural Networks
CSF	Cerebrospinal Fluid
CT	Cortical Thickness
CV	Coefficient of Variation
DNN	Deep Neural Network
DSC	Dice Similarity Coefficient
EA-CSF	Extra-Axial CSF
EBDS	Early Brain Development Study
EMD	Earth Mover's Distance
FCN	Fully Convolutional Neural Network

FD	Fractal Dimension
FDR	False Discovery Rate
FMM	Fast Marching Method
FN	False Negatives
FP	False Positives
GANs	Generative Adversarial Networks
GI	Gyrification Index
GM	Gray Matter
GMM	Gaussian Mixture Model
GPUs	Graphics Processing Units
HR	High-familial Risk
IBIS	Infant Brain Imaging Study
ICV	Intra-cranial Volume
LR	Low-familial Risk
MAE	Mean Absolute Error
MLP	Multilayer Perceptron
MMSE	Mini-Mental State Examination
MRI	Magnetic Resonance Imaging
MTL	Multi-Task Learning
NDDs	Neurodevelopmental Disorders
NPV	Negative Predictive Value
OC-NN	One-Class Neural Network

OC-SVM	one-class SVM
PDE	Partial Differential Equation
PPV	Positive Predictive Value
PVE	Partial Volume Effects
QC	Quality Control
ReLU	Rectified Linear Unit
SA	Surface Area
SCI	Shape Complexity Index
SEN	Sensitivity
SMC	Significant Memory Concern
SMOTE	Synthetic Minority Over-sampling
sMRI	Structural MRI
SPC	Specificity
T1w	T1-weighted
T2w	T2-weighted
TN	True Negatives
TP	True Positives
VAEs	Variational Autoencoders
VBM	Voxel-based Morphometry
WM	White Matter

CHAPTER 1: INTRODUCTION

1.1 Overview

Neurodevelopmental Disorders (NDDs) are a diverse group of conditions that are characterized by deficiencies in brain functions. Such defects include cognition, communication, behavior, and motor skills resulting from atypical brain development. Attention-deficit/Hyperactivity Disorder (ADHD), Autism Spectrum Disorder (ASD), and Fragile X syndrome are examples of disorders that fall under the umbrella of NDDs (Association et al., 2010). Such conditions can be chronically disabling, resulting in significant long-term impairment lasting to the adult life of those children who suffer them. Due to the lack of biomarkers to diagnose NDDs or to differentiate among them, these disorders are still currently identified based on clinical interviews and behavioral observations. Therefore, NDDs share common needs that include *(1)* the development of a clinically-useful, early presymptomatic test for identifying infants who will develop neurodevelopmental disorders, thereby enabling more efficient early intervention in infancy, *(2)* the assessment of disease severity through inspecting associated dysfunctional brain systems and contrasting them to typically functional ones, leading to accurate long-term disease prognosis and expedited evaluation of applied interventions.

Advances in Magnetic Resonance Imaging (MRI) enabled the non-invasive visualization of the infant's brain through acquired high-resolution images. The increasing availability of large-scale datasets of detailed infant brain multi-modal MR images, e.g., T1-weighted (T1w), T2-weighted (T2w), and Diffusion-weighted MRI images, affords unique opportunities to accurately study early postnatal brain development leading to insights into the origins and abnormal developmental trajectories of NDDs. However, the processing and analysis of MR infant brain images are typically far more challenging as compared to the adult brain setting. Notably, an infant's brain MRI suffers from reduced tissue contrast, large within-tissue inhomogeneities, regionally-heterogeneous image

appearance, considerable age-related intensity changes, and severe partial volume effect due to the small brain size. Since most of the existing tools were designed for adult brain MRI data, infant-specific computational neuroanatomy tools are a relatively recent development.

As acquiring such multi-modal MR infant brain images becomes more common, the collections of medical imaging data available to researchers are increasing in number, size, and complexity. For instance, the Baby Connectome Project (BCP) is aiming to acquire longitudinal, high-resolution, multi-modal MRI data from more than 500 typically-developing children from birth to 5 years of age (Howell et al., 2019). This necessitates the development of methods that can utilize information extracted from these large datasets. When analyzing big data, Artificial Intelligence (AI) and machine learning are often used interchangeably. However, the term AI has a broader notion than machine learning, which tackles the use of computers to mimic humans' cognitive functions. Machine learning is a subset of AI that focuses on the design and development of algorithmic techniques that allow computer systems to evolve function or behavior by learning from extensive data. Classical machine learning algorithms have a variety of applications that have been tailored to the medical imaging field (Wernick et al., 2010; Wang and Summers, 2012; Deo, 2015). However, most necessitate the creation of "handcrafted" image features through careful engineering and specific domain expertise. Also, such traditional machine learning algorithms tend to not generalize well to new, previously unseen data (LeCun et al., 2015). This problem is amplified in medical imaging applications due to the inherent anatomical variability in brain morphology, discrepancies in acquisition settings and MRI scanners, and variations in the appearance of pathological tissues.

In recent years, a machine learning technique referred to as deep learning (LeCun et al., 2015; Schmidhuber, 2015) has gained wide popularity in many artificial intelligence applications due to its ability to overcome limitations of classical machine learning algorithms. Deep network architecture enables the extraction of a complex hierarchy of features from input data via self-learning as contrasted with the engineered feature extraction methods used by classical machine learning algorithms. Deep learning shows impressive performance and generalizability through training on a large amount of data (Krizhevsky et al., 2012; He et al., 2016; Huang et al., 2017). This

success is due mostly to the rapid progress in computational power, in particular through Graphics Processing Units (GPUs), which enabled the fast development of complex deep learning algorithms. Several types of deep learning architectures have been developed for different tasks, including object detection, speech recognition, and classification. In turn, the success of deep learning in computer vision leads to its use in medical image analysis: for image segmentation (Akkus et al., 2017), image registration (Yang et al., 2017), image fusion (Suk et al., 2014), lesion detection (Kooi et al., 2017), and computer-aided diagnosis (Hoo-Chang et al., 2016). While deep learning-based methods have made significant strides in medical imaging applications, there are still significant open problems, and relatively few methods have been applied to infant MRI datasets. Therefore, there is still a need to develop deep learning methods that can learn from such complex datasets in a way that tackles open issues including low data size restrictions, class imbalance problems, and lack of interpretation of the resulting deep learning solutions.

There has been a limited effort to utilize deep learning techniques for early prediction of NDDs using complex infant neuroimaging datasets. This dissertation identifies key challenges that need to be tackled when performing such an analysis:

[1] Extracting cost-effective, accurate, and reliable, presymptomatic, local imaging biomarkers for NDDs diagnosis. Deriving disease-specific features is critical for identifying high-risk infants who would benefit from very early intervention.

[2] Determining optimal strategies to learn from a diverse high-dimensional feature space efficiently. This includes adopting dimensionality reduction techniques that avoid losing relevant information and identifying a parsimonious subset of features needed for optimal performance.

[3] Extending current deep learning methods to irregular non-Euclidean domains as NDDs-related biomarkers are often extracted from cortical surfaces. The ability to learn directly from non-Euclidean features in a way that takes into account the underlying geometry would lead to performance improvements as well as providing meaningful insights to the investigated disease.

[4] Improving the robustness of multi-task deep learning models utilizing multi-modal imaging datasets. The ability to predict disease-related dimensional scores in addition to categorical diagnostic outcomes would facilitate the development of targeted interventions. The generalization performance of such a multi-task learning model can further be improved by leveraging multi-modal datasets (i.e., includes imaging and non-imaging information).

[5] Detecting samples from untrained classes in imbalanced and heterogeneous settings typically encountered when studying NDDs. This challenge can be tackled by formulating the problem as anomaly detection with generative deep learning models trained to recognize samples from the well-represented classes. Such a framework would help improve the robustness of deep learning models, increasing the confidence of the clinical decisions made using such models.

This dissertation describes novel methodological contributions to tackle the challenges above (Figure 1.1). First, I introduce a Laplacian-based method for quantifying a local measure of Cerebrospinal Fluid (CSF) from structural MRI. Next, I describe an approach for combining morphometric cortical measurements for early disease diagnosis. Then, I propose a data-driven approach for extending Convolutional Neural Networks (CNNs) for use on non-Euclidean cortical surfaces. I also present a framework for multi-task learning utilizing multi-modal information to improve the robustness of the learned models. Finally, I propose a generative approach to identify samples from untrained classes in imbalanced and heterogeneous settings.

The proposed methodological contributions are mainly evaluated in the application of the early detection of ASD in the first year of the infant's life. Also, the adult human brain will be investigated in the context of studying different stages of Alzheimer's Disease (AD). In subsequent sections, I provide an overview of the driving application problems, outline the prior work highlighting current limitations, and provide the intuition behind each of my contributions. Then I make my thesis statement and describe the organization of the remaining chapters.

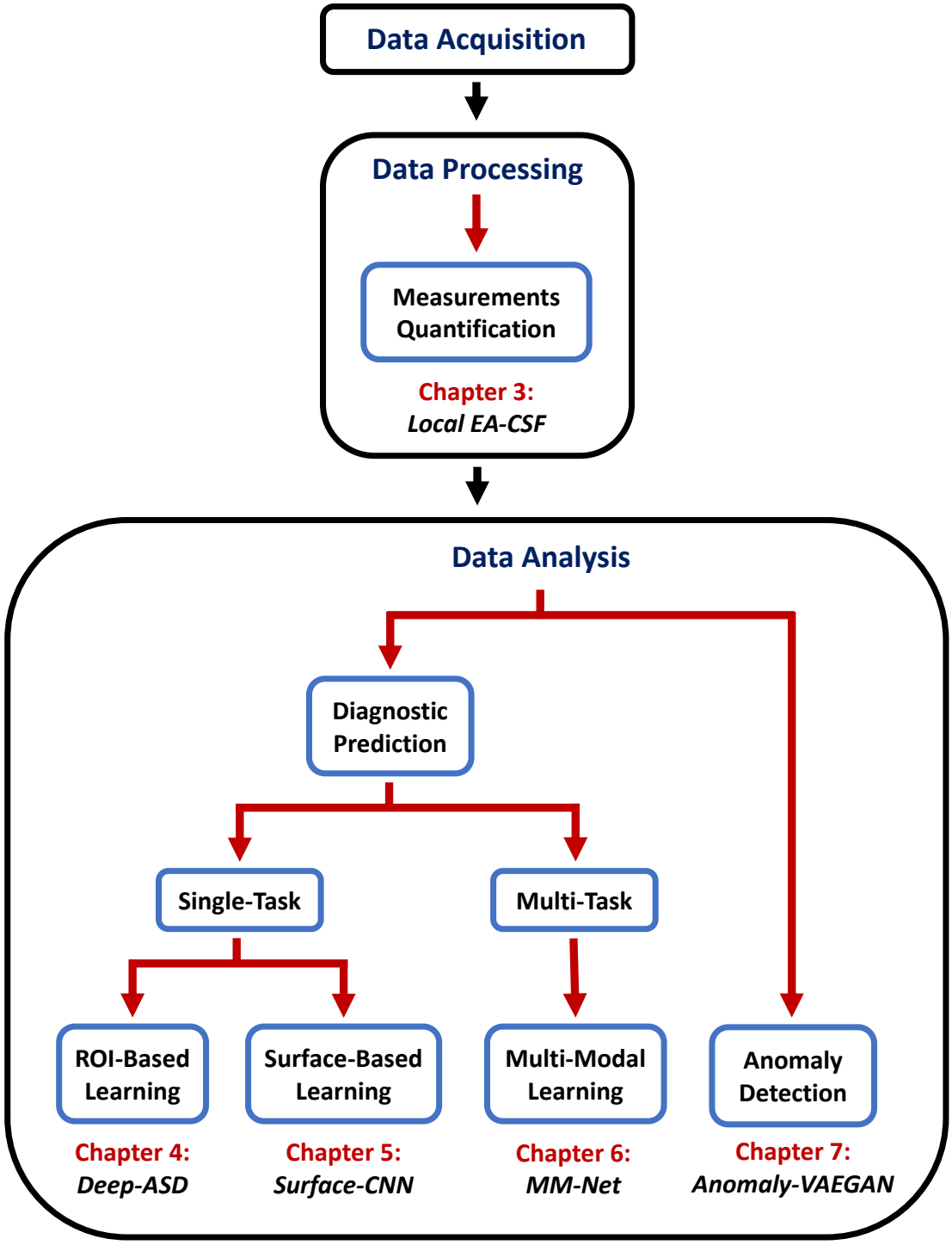


Figure 1.1: A schematic overview of the proposed contributions in the MRI processing and analysis pipeline. A detailed discussion of the each components is made for each corresponding chapter.

1.2 Motivating Problems and Proposed Solutions

1.2.1 Current Findings on Early Prediction of ASD from Neuroimaging Studies

Motivation: Behavioral symptoms that are characteristic of ASD are generally not evident until the second year of life (Rogers, 2009; Ozonoff et al., 2010; Landa et al., 2013; Zwaigenbaum et al., 2005) and then consolidate into the clinical diagnosis of ASD around 2-3 years of age (Rice, 2009). Observable behavioral markers in the first year have not yet been shown to be sensitive or specific enough for accurate prediction of later ASD diagnosis, with Positive Predictive Value (PPV)s of 50% or lower (Ozonoff et al., 2009; Chawarska et al., 2014; Bussu et al., 2018). The current clinical practice is to initiate treatment only after behavioral symptoms arise in the second year of life, or after diagnosis is made between 2-3 years of age. Early interventions have been shown to be more effective than later interventions (Rogers and Vismara, 2008; Clark et al., 2018; Dawson et al., 2010; Vivanti et al., 2016; Rogers et al., 2014; Green et al., 2015; Jones et al., 2017; Cidav et al., 2017) and treatment in infancy, during the period of maximal postnatal brain plasticity, may be even more efficacious (Green et al., 2015). Thus, there is a strong rationale for identifying early, pre-symptomatic markers of ASD to accurately identify those children at highest risk who may benefit most from the pre-symptomatic intervention for ASD.

Prior Work: Efforts to identify early markers of ASD have benefited from the prospective study of younger siblings of ASD patients, who are at 15 to 20-fold higher risk of developing ASD than the general population (Ozonoff et al., 2011). Recent MRI studies had identified brain differences as early as 6 months of age, before the onset of diagnosable behavioral symptoms, in high-familial risk infants who were later diagnosed with ASD. It was first reported that cortical surface area generated from Structural MRI (sMRI) scans at 6 and 12 months could predict ASD diagnosis at 24 months (Hazlett et al., 2017). Subsequently, it was reported that functional connectivity MRI at 6 months could predict ASD diagnosis in a small sample of eleven high-risk infants who developed ASD (Emerson et al., 2017). Finally, in two separate samples, an excessive volume of CSF in the subarachnoid space surrounding the cortical surface (i.e., Extra-Axial CSF or EA-CSF)

at 6 months of age was found in infants who developed ASD two years later (Shen et al., 2013, 2017). Collectively, these previous studies demonstrated that pre-symptomatic brain markers of ASD are present in the first year of life before the full manifestation of the disorder, suggesting that a combination of cortical surface anatomy and CSF volume may be useful in predicting later ASD.

My Contributions: The approaches taken to date had several notable limitations. In the previous report using sMRI (Hazlett et al., 2017) of cortical surface area required predictors at both 6 and 12 months of age; the global measurement of EA-CSF volume lacked sufficient specificity to serve as a single predictor of later diagnosis; and the prediction based on functional connectivity MRI was conducted on a small sample and required acquisition and processing methods that may be more complex to implement for widespread clinical applications. To address these limitations, *I first propose a novel framework, called **Local EA-CSF**, for the automatic computation of EA-CSF measurements in a way that is suitable for localized analysis*. The proposed processing relies on probabilistic brain tissue segmentation, cortical surface reconstruction, as well as streamline-based quantification to produce local EA-CSF measurements. Then, *I propose a new deep learning-based framework, namely **Deep-ASD**, with the objective of optimally combining multiple measures of infant brain shape and CSF from a conventional sMRI scan at 6 months to improve the accuracy of predicting ASD diagnosis at 24 months*. While an accurate prediction can be clinically relevant: it is also crucially important to provide information on how the prediction is made and how that generates a better understanding of the disease. Hence, the proposed framework additionally aims to identify which input features have the most significant impact on the prediction performance.

1.2.2 Deep Convolutional Learning on Non-Euclidean Cortical Surfaces

Introduction: The development of sophisticated, noninvasive 3D medical imaging technologies such as MRI provides a means to extract biomarkers to assist in the diagnosis of NDDs and neurodegenerative diseases. Neuroanatomical abnormalities in the cerebral cortex are widely investigated by examining cortical brain morphometric measures such as cortical thickness and cortical surface area (Hazlett et al., 2017). Such cortical measures are usually extracted from

finely-sampled 2D cortical surfaces, forming a high-dimensional feature list. Based on the extracted features, machine learning classifiers can be used to predict diagnostic outcomes. Such classifiers undergo a training process in which they learn category labels for each set of features. However, with limited datasets, it is hard to train stable classifiers using such a high-dimensional feature space without overfitting. To solve this problem, a dimensionality reduction step becomes necessary for improving the prediction performance of the predictors, providing faster and more cost-effective predictors and providing a better understanding of the underlying process that generated the data.

Dimensionality reduction can be performed using machine learning techniques in a supervised or unsupervised manner (Van Der Maaten et al., 2009). However, unsupervised methods run the risk of losing relevant information while supervised methods tend to be more biased and therefore harder to generalize. Alternatively, if possible, dimensionality can be reduced by first summarizing the features via prior anatomical knowledge, such as summarizing vertex-wise cortical features via a cortical subdivision that divides the cortex into a mosaic of anatomically and/or functionally distinct, spatially adjacent areas using prior parcellation atlases (Glasser et al., 2016). For example, in the ASD prediction framework proposed in (Hazlett et al., 2017), cortical measurements were first summarized using the Brodmann-based automatic anatomic labeling atlas, namely AAL (Achard et al., 2006; Tzourio-Mazoyer et al., 2002). However, current parcellation atlases are generic and not optimized to the given age or pathology that is studied, running the risk of losing relevant information that can be critical in the classification process. The reliance on parcellation atlases is particularly problematic in infant studies as currently only adult atlases are employed due to the scarcity of infant cortical surface atlases. Hence, there is a need for data-driven classifiers that can directly learn from a high-dimensional feature space without a separate feature reduction step.

Prior Work: In recent years CNNs have become the predominant machine learning tool in computer vision and speech recognition. The main strength of CNNs is their ability to extract hierarchical abstractions directly from raw data with little need for prior knowledge. Another benefit of CNNs is that they are easier to train and have fewer parameters than fully connected networks with the same number of hidden units. However, extending the use of CNNs in applications where

the input data has an irregular non-Euclidean structure is still challenging. Challenges include the missing notion of a grid on a non-Euclidean surface and the additional need for localized kernel design. To solve this problem a generalization of the CNN paradigm to non-Euclidean manifolds was introduced based on a local geodesic system of polar coordinates to extract "patches," which are then passed through a cascade of filters and linear and non-linear operators (Masci et al., 2015). The coefficients of the filters and linear combination weights are optimization variables that are learned to minimize a task-specific cost function. Although this implementation is promising, it usually fails if the surface mesh is very irregular or if the radius of the geodesic patches is large compared to a curvature radius of the shape. To solve these drawbacks, one would need to generate highly uniform surface representations as well as to decrease the geodesic path radius, which in turn might be a problem in terms of expected computational complexity. An alternative generalization based on localized frequency analysis (a generalization of the windowed Fourier transform to manifolds) that is used to extract the local behavior of some dense intrinsic descriptor, roughly acting as an analogy to patches in images (Boscaini et al., 2015). The resulting local frequency representations are then passed through a bank of filters whose coefficients are determined by a learning procedure minimizing a task-specific cost. Such implementation addressed some of the limitations described in the previous work; however, this work was only designed to capture shape descriptors on the surface and is not suitable to be used with other surface measures.

My Contributions: In this dissertation I aim to address the limitations mentioned above by *proposing a novel data-driven approach to extend CNNs to non-Euclidean manifolds such as cortical surfaces. I accomplish that using a novel* ***Surface-CNN*** *architecture that utilizes geodesic-based kernels to learn the most powerful features and brain regions from the extracted large dimensional feature space*: thus creating a new feature space in which the dimensionality is reduced and feature distributions are better separated. By introducing novel surface convolution, pooling, and resampling layers, the proposed CNN extension applies to various kinds of clinical applications that involve learning from high-dimensional features living on non-Euclidean manifolds.

1.2.3 Deep Multi-Modal Multi-Task Learning for Robust Representation Learning

Motivation: Although deep learning models have recently reached near-human-level performance in a variety of tasks, the robustness of state-of-the-art deep classifiers is still an open question as it was shown that such models could be unstable to small perturbations in the input data (Elsayed et al., 2018). Deep learning models generalize well only when labeled data is plentiful (Bernardo et al., 2007); this limitation is critical factor in the context of learning from medical imaging datasets. Particularly, in developing robust deep learning models for early prediction of NDDs, collecting the required vast amounts of annotated training data is a significant and rarely achieved challenge. This leads to performance uncertainty, which imposes a key concern when deploying these models in clinical settings. Therefore, there is a need to move beyond generic models to personalized models that guarantee excellent performance on an individual level. To achieve that, deep learning models need to be designed to tackle different sources of heterogeneity, such as data source heterogeneity (i.e., different modalities) and task heterogeneity (i.e., learning multiple tasks). However, many problems exhibit the coexistence of both types of heterogeneity, which necessitate the development of frameworks that address modality and task heterogeneities simultaneously. Multi-modal deep learning aims to build models that can process and relate information from multiple modalities in a way that exploits the data complementarity and redundancy (Ngiam et al., 2011). Though combining different types of information for improving the performance and robustness of learning-based models, however, it is challenging to effectively combine modalities with varying level of noise and influence on the prediction output (Baltrušaitis et al., 2018). Using deep learning methods to deal with multi-modal medical images is more challenging because of the increased diversity of underlying mechanisms that generated the data. Nevertheless, deep neural networks, mainly using CNNs, are very actively explored in multi-modal medical imaging applications such as in infant brain image segmentation (Nie et al., 2018) and neurodegenerative diseases diagnosis (Sarraf et al., 2016). Though the imaging modalities selected and architectures implemented might vary across these methods, the common idea of combining the information in the joint hidden layer of a neural network remains the same. A drawback of current multi-modal CNNs is their inability to

incorporate non-imaging information in the network. In the field of medical imaging, a lot of relevant non-imaging information is often available; examples of that include patient information (e.g., demographic information, medical history, symptoms), scanner information (e.g., scanner model, field of strength), and acquisition parameters (field of view, image resolution, sequence parameters). The availability of such information to the network might improve the overall performance and robustness by forcing the network to learn parameters that would work in different settings.

Prior Work: Current deep learning models applied for diagnostic tasks are mainly focused on predicting a diagnostic categorical outcome. The next step beyond diagnostic categories is to predict continuous dimensions of disease-related features. For example, high-risk infants who develop ASD will have a range of outcomes along with multiple aspects such as language, cognitive function, and joint attention. The ability to predict later scores in these areas will facilitate the development and implementation of targeted, individualized, presymptomatic interventions. However, it is still challenging to achieve that task when applied in such low prevalence and highly heterogeneous setting. Multi-Task Learning (MTL) (Ruder, 2017) aims to learn multiple tasks simultaneously by exploiting the underlying similarities between correlated tasks. By allowing the network to learn common feature representation for multiple tasks, an improvement in the network generalization performance is gained. Additional MTL benefits include improvements in prediction accuracy and network scalability as well as a reduction in computational complexity (Ruder, 2017). This allowed utilization of MTL in various applications that involve learning from limited medical imaging datasets, including tissue segmentation (Moeskops et al., 2016b) and the prediction of neurodegenerative disorders (Thung et al., 2017). Challenges of MTL include selecting a paradigm for sharing network parameters among different tasks, i.e., hard or soft parameter sharing of hidden layers. In deep neural networks, MTL is typically done with either hard or soft parameter sharing of hidden layers. In hard parameter sharing, the hidden layers are shared between all tasks while several output layers are used for each task separately (Caruna, 1993). On the other hand, soft parameter sharing allow each task to has its model and parameters while regularization techniques (e.g., L_2 norm (Duong et al., 2015) or trace norm (Yang and Hospedales, 2016)) were employed to

keep the models' parameters similar. Despite the reduced model flexibility, a hard sharing strategy is more commonly used since it forces the network to rely on generalized representations which reduce the chance of overfitting (Baxter, 1997). Another issue to consider when applying MTL is the possibility of learning imbalances in training as some tasks may dominate others during training (Guo et al., 2018). Hence, choosing an appropriate loss or prioritization strategy becomes a necessity. Instead of using a fixed weighted loss function with weights determined through an expensive hyperparameters tuning, automatic methods are being proposed to weight each task loss dynamically. Criteria for dynamically weighting each task loss include using homoscedastic uncertainty (Kendall et al., 2018), task gradient magnitudes (Chen et al., 2017), or running averages of tasks losses (Liu et al., 2019). Alternative approaches include modeling the multi-task problem as a multi-objective problem (Sener and Koltun, 2018); however, such a strategy is still challenged by the need for defining task heuristics or require complex optimization techniques.

My Contributions: To address some of the limitations mentioned above, *I propose a novel multi-task multi-modal deep learning framework, namely **MM-Net**, which extends a DenseU-Net architecture (Huang et al., 2017) to the simultaneous learning of classification, regression, and segmentation tasks in multi-modal neuroimaging datasets.* The proposed architecture relies on novel multi-modal convolutional blocks to combine imaging and non-imaging information effectively at each level of the network. In addition to a convolutional layer, a multi-modal convolutional block also includes a normalization layer, a linear layer, and a nonlinearity to allow efficient information to flow between imaging and non-imaging data. The proposed framework is applied for predicting diagnostic outcomes as well as continuous scores in a disease-related domain from input MRI and associated non-imaging information (e.g., patient demographic data). Moreover, MRI segmentation is also defined as an auxiliary task to improve the performance of main prediction tasks.

1.2.4 Deep Generative Models for Anomaly Detection in Heterogeneous Datasets

Motivation: Discriminative deep learning models have shown notable success in many medical image analysis applications, including segmentation and disease prediction (Litjens et al., 2017).

However, on problems that involve learning from noisy datasets with highly imbalanced classes, their success is limited (Litjens et al., 2017). This problem is exacerbated when developing diagnostic or predictive approaches for infants with NDDs as prevalence rates are commonly low in NDDs (e.g. in ASD, $\leq 2\%$ for the general population and $\leq 20\%$ for high-risk population). It has been recognized that the class imbalance problem has a substantial negative impact on training deep learning models. With imbalanced datasets, deep learning models tend to focus on learning the classes with a large number of examples, leading to poor performance for the classes with a small number of examples. In a diagnostic setting based on medical image data, misclassification costs are typically unequal and classifying a diseased sample (minority class) as typical (majority class) has significant consequences that should be avoided. Currently there is no consensus on the effects of the class imbalance issues and how to mitigate this problem optimally; this limitation could affect the reproducibility and accuracy of medical imaging research.

In classical machine learning, the class imbalance problem is typically solved by adjusting the representation of the minority classes using data augmentation or oversampling techniques (Mazurowski et al., 2008; Maloof, 2003). However, the methods used to address the class imbalance in the case of traditional shallow models are not always applicable to deep learning applications on complex medical imaging datasets. Particularly, such strategies fail if the minority classes are also inherently heterogeneous with some classes missing from the training dataset. Thereby, the trained models can provide incorrect labels with very high confidence for underrepresented samples (Elsayed et al., 2018). Generative modeling that can exploit unlabeled data in addition to labeled data can provide a relief to resolve this problem. By only observing samples from the majority classes, these models can learn the underlying true data distribution, which can, in turn, be used to detect anomalous samples that are deemed by the network to be unlikely generated by the learned distribution. However, traditional generative models have shown an inability to scale up to high dimensional datasets (Bengio et al., 2014). Recent advances in parameterizing generative models via deep learning models, combined with advancement in stochastic optimization techniques, have allowed scalable modeling of complex, high-dimensional data.

Prior Work: Several medical imaging out-of-sample methods have been proposed that utilize deep generative architectures such as Variational Autoencoders (VAEs) or Generative Adversarial Networks (GANs). VAEs explicitly try to approximate the data distribution of the training data both in the latent and the original space by maximizing a variational lower bound (Kingma and Welling, 2013). In the original space, out-of-sample detection can be performed using the reconstruction score (Lu and Xu, 2018) or distance- or density-based approaches (Vasilev et al., 2018). Anomalies can also be detected in the learned latent space using VAE regularizer scores (Lu and Xu, 2018) or an offline one-class support vector machines (El Azami et al., 2016). Compared to VAEs, GANs avoid any strong distributional assumptions by implicitly specifying probabilistic models describing a stochastic procedure to generate higher quality images directly (Goodfellow et al., 2014). GAN-based anomaly detection for testing samples is performed based on post hoc estimated likelihood scores (Schlegl et al., 2017). Nonetheless, GANs are still challenging to train, and the post hoc score estimators are very slow and tend to produce inaccurate likelihood estimates. In 2018, architectures utilizing both VAE and GAN were introduced for pixel-wise anomaly detection. Chen and Konukoglu (Chen and Konukoglu, 2018) used an adversarial autoencoder (Makhzani et al., 2015) with latent space consistency constraints to identify anomalies based on created reconstruction error maps. Similarly, an architecture that combines a VAE and a GAN was also proposed (Baur et al., 2018) to detect within image anomalies, which showed improvements over the GAN-only approach in (Schlegl et al., 2017). However, these models were designed to detect only pixel-wise anomalies in limited 2D homogeneous datasets and failed to incorporate external information to organize the learned image manifolds. Hence, there is still a need for a new architecture to provide sample-wise anomaly scores in 3D high-dimensional heterogeneous data.

My Contributions: To tackle the limitations of prior work, *I propose a novel approach for identifying samples from untrained classes in imbalanced and heterogeneous datasets. The proposed semi-supervised **Anomaly-VAEGAN** architecture extends the deep generative model (Larsen et al., 2015) for robust anomaly detection.* The proposed framework relies on a high-level similarity metric and invariant representations learned by a semi-supervised discriminator to evaluate the

generated images. The encoded latent representations were constrained according to user-defined properties through a jointly trained predictor network. Anomalous samples are detected using learned similarity scores and/or scores from an online one-class neural network.

1.3 Thesis Statement

Thesis: *Applied to complex neuroimaging datasets, deep learning models can obtain an accurate and robust prediction of neurodevelopmental disorders and neurodegenerative diseases. However, key issues need to be considered for optimal performance, including (1) extracting relevant imaging and surface features based on prior domain knowledge, (2) determining the optimal parsimonious subset of cortical surface and imaging features needed for an accurate disease diagnosis, (3) learning from high-dimensional feature space on curved manifolds, (4) modeling modality and task heterogeneities for improved model robustness, and (5) identifying out-of-distribution samples to increase the confidence in the model decisions.*

The contributions of this dissertation include

[1] A *local measure of extra-axial cerebrospinal fluid:* a novel framework for the extraction of accurate and reliable local extra-axial cerebrospinal fluid measurements. The proposed software tool is the first to address the problem of quantifying extra-axial cerebrospinal fluid in a way that is suitable for localized surface-based analysis. The proposed tool allows neuroimaging labs to investigate the use of local extra-axial cerebrospinal fluid in early diagnostics of neurodevelopmental disorders and neurodegenerative diseases.

[2] A *framework for early prediction of autism spectrum disorder:* a novel deep learning-based approach that combines multiple measures of brain anatomy from a conventional structural magnetic resonance imaging scan at 6 months to improve the accuracy of predicting autism diagnosis at 24 months in high-risk infants. The proposed framework provides a clinically-

useful measure for the pre-symptomatic detection of autism, which will enable pre-symptomatic intervention, leading to earlier and more effective treatments for autism.

[3] A *convolutional neural networks extension to non-Euclidean cortical surfaces:* a novel generalization of convolutional neural networks on non-Euclidean manifolds such as cortical surfaces. A general definition of kernels on cortical surfaces is provided using a locally constructed geodesic grid. An accurate classification in a high-dimensional feature space without the need for a separate dimensionality reduction step is achieved using a learning framework that relies on novel surface convolution, pooling, and resampling layers. The data-driven approach allows the identification of disease phenotypes by providing interpretations of the trained models.

[4] A *multi-modal multi-task deep learning architecture for joint representation learning:* a novel deep learning architecture for addressing modality and task heterogeneities to improve the robustness of learned models. The proposed architecture relies on novel multi-modal convolutional blocks to combine imaging and non-imaging information effectively at each level of the network. In addition to a convolutional layer, a multi-modal convolutional block also includes a normalization layer, a linear layer, and a nonlinearity to allow information to flow between imaging and non-imaging data. The proposed multi-task learning framework allows for the simultaneous learning of classification, regression, and segmentation tasks defined as a primary or an auxiliary task. In a diagnostic setting the proposed framework allows for the simultaneous prediction of diagnostic outcomes as well as continuous scores in disease-related domains, thereby facilitating the development of targeted, individualized interventions.

[5] A *deep generative model for robust anomaly detection in imbalanced and heterogeneous datasets:* a novel semi-supervised out-of-sample detection framework based on a variational autoencoder-based generative adversarial network. The proposed framework can identify samples from untrained classes in a wide range of applications that involve learning in low prevalence and highly heterogeneous setting (e.g., autism). Anomalous samples are detected using high-level similarity scores and scores from a jointly trained one-class neural network. A

semi-supervised discriminator is proposed to stabilize network training and ensure a meaningful learned similarity metric. The encoded latent representations are also constrained using user-defined properties through a jointly trained predictor network.

1.4 Overview of Chapters

The remainder of this dissertation is organized as follows: Chapter 2 provides an overview of the necessary background in this dissertation, including cortical surface reconstruction, quantification of cortical shape measures, and deep learning architectures such as CNN, U-Net, DenseNet, VAE, and GANs. Chapter 3 presents ***Local-EACSF***, a processing pipeline for automatically computing local EA-CSF. Chapter 4 presents ***Deep-ASD***, a deep learning framework for predicting ASD diagnosis at 24 months using cortical measures of shape and CSF extracted from a 6-month sMRI scan. Chapter 5 presents ***Surface-CNN***, a data-driven approach for extending CNNs to non-Euclidean manifolds such as cortical surfaces. Chapter 6 presents ***MM-Net***, a deep learning architecture for the simultaneous learning of classification, regression, and segmentation tasks in multi-modal datasets that include imaging and non-imaging information. Chapter 7 presents ***Anomaly-VAEGAN***, a semi-supervised VAE-GAN model for robust anomaly detection in heterogeneous datasets. Chapter 8 concludes the dissertation with a discussion of its contributions and some potential future work.

CHAPTER 2: BACKGROUND

This chapter presents the background materials required for this dissertation. Section 2.1 highlights the current role of deep learning at each step of a typical neuroimaging processing and analysis pipeline, and discuss the current open challenges utilizing deep learning in infant MRI. Section 2.2 provides an overview of popular cortical surface reconstruction methods and describes different geometric and shape properties extracted from cortical surfaces. Finally, section 2.3 discusses discriminative and generative approaches to deep learning and summarizes the deep learning architectures that are utilized in this dissertation.

2.1 Overview of Deep Learning in Structural MRI Processing and Analysis

MRI Processing and Analysis Pipeline: MRI allows the non-invasive acquisition of 3D images at high spatial resolution. sMRI provides the so-called anatomical images on which one can observe the different tissues that constitute the brain (see Figure 2.1). In addition to T1w MRI to create precise anatomical images, other modalities of images can be acquired utilizing alternative properties of the MR signal (e.g., T2w and T2*w). Currently, T1w and T2w sequences form the

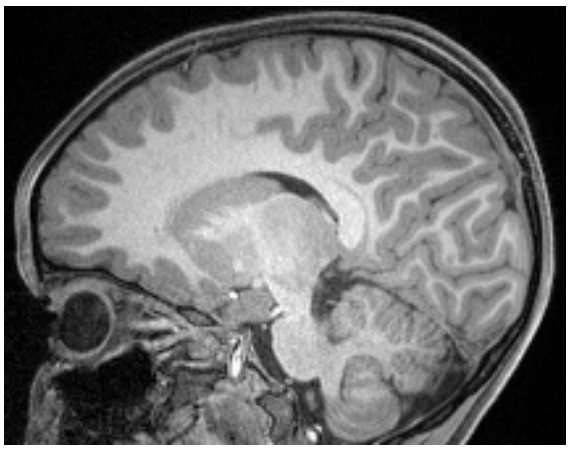


Figure 2.1: 3T structural (T1-weighted) MR brain image of a 20-month old infant. Gray matter (medium gray) is visible on the brain and cerebellum contours and in some part of the middle-brain. White matter is in light gray. Cerebrospinal fluid is in dark black between the brain and the skull. The shown high-resolution T1-weighted MR brain image (0.8mm) was obtained as part of the Baby Connectome Project (Howell et al., 2019).

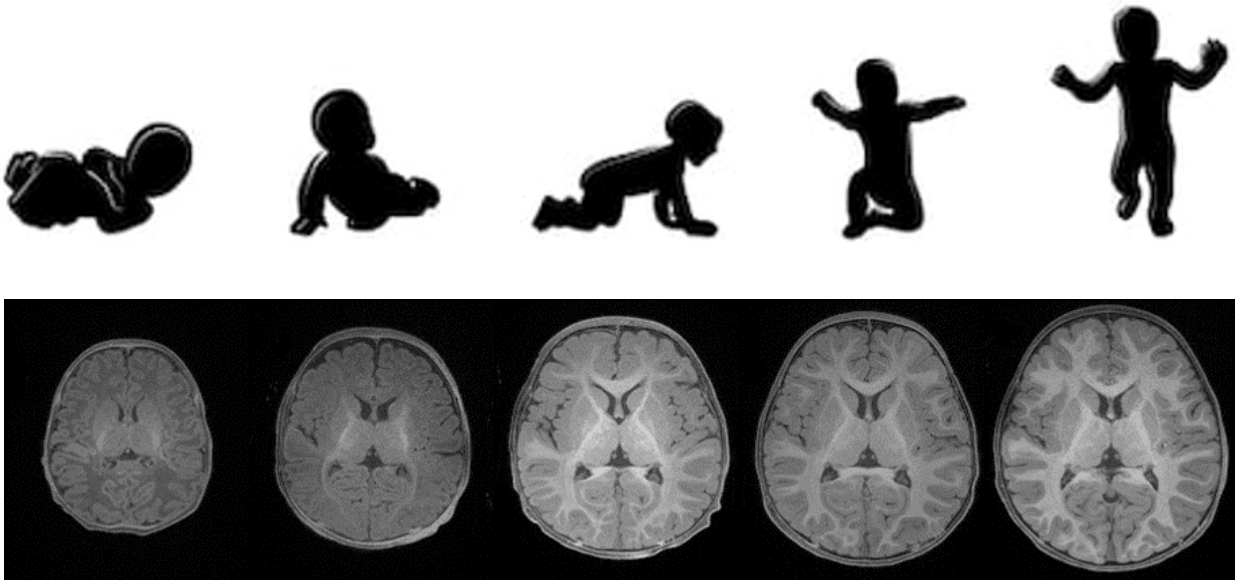


Figure 2.2: T1-weighted brain images of a typically-developing infant, scanned longitudinally at 0, 3, 6, 12, and 24 months of age. Longitudinal MRI scans allow researchers to study the first year of life, leading to critical insights into many neurodevelopmental and neuropsychiatric disorders.

center of nearly every clinical MRI protocol. Pathological processes are accordingly most often defined in terms of T1 and T2 signal behavior, and besides contrast enhancement, anatomical position, and morphological properties. Moreover, sMRI can be used in longitudinal studies, where the development of an individual's brain is observed over an extended period of time (up to several years) in multiple scans. Particularly, as shown in Figure 2.2, longitudinally acquired sMRI can provide unique opportunities to study early brain development, potentially aiding researchers in identifying and understanding the origins and abnormal developmental trajectories of NDDs such as ASD. Currently, relatively few processing and analysis pipelines have been developed for infant MRI as compared to adult MRI. This can be attributed to the numerous additional inherent challenges such as inhomogenous tissue appearance across the image, considerable image intensity variability across the first year of life, and a low signal to noise setting. As shown in Figure 2.3, a typical pipeline for early prediction of NDDs from infant sMRI consist of three main steps: *(1)* Image preprocessing, tissue segmentation, and regional labeling, and extraction of image-based features (Jenkinson et al., 2012; Dai et al., 2013; Penny et al., 2011). *(2)* Surface reconstruction, surface correspondence

establishment, surface parcellation, and extraction of surface-based features (Fischl, 2012; Ad-Dabbagh et al., 2006; Styner et al., 2006). *(3)* Feature preprocessing, feature extraction, machine learning model training, and prediction of unseen subjects (Kim et al., 2016b; Mostapha et al., 2018b; Kim et al., 2005; Hazlett et al., 2017).

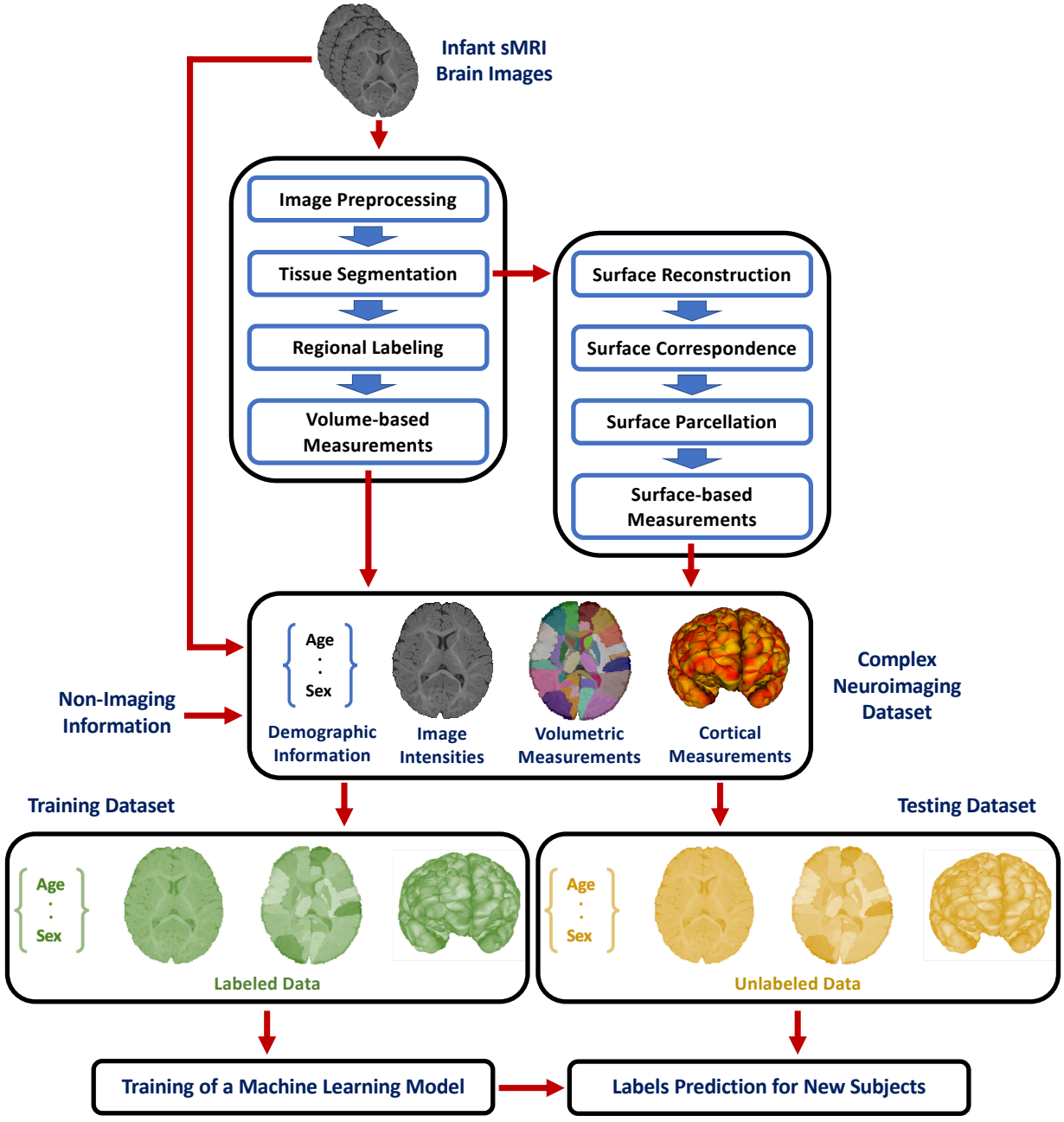


Figure 2.3: A typical pipeline for early prediction of neurodevelopmental disorders using infant structural MR brain images, which employs infant-specific processing and analysis steps.

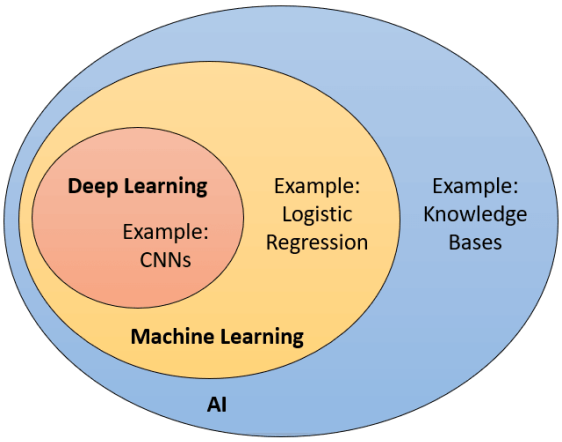


Figure 2.4: Venn diagram illustrating the relationship between deep learning, machine learning and artificial intelligence.

Machine Learning Role: In the realm of big data, AI and machine learning are often used interchangeably. However, AI has a broader notion than machine learning, it tackles the use of computers to mimic humans' cognitive functions. As illustrated in Figure 2.4, machine learning is a subset of AI that focuses on the design and development of algorithmic techniques that allow computer systems to solve problems by learning from experience. Given experience in the form of training data, machine learning models are tuned to provide accurate predictions for the training data through an optimization procedure. However, the main goal for these models is to generalize the learned experience to provide correct predictions for new, unseen data (including test data). Based on how training data are utilized, machine learning models can be loosely categorized into supervised learning, unsupervised learning, and reinforcement learning. In reinforcement learning, agents are created to learn from an environment through a trial and error procedure while optimizing some cost function (Silver et al., 2017). On the other hand, unsupervised learning aims to uncover patterns (i.e., clusters) in the data without the guidance of given labels. Most of current machine learning methods pertain to the category of supervised learning, where labeled or annotated data are provided with the task of producing correct labels on new datasets based on the rules learned from the labeled dataset. Although traditional machine learning methods have been successfully applied to a variety of applications in the medical imaging field (Wernick et al., 2010), they are still limited by their reliance on manually designed features extracted from the raw data. Hence, classical machine learning tends to not generalize well mainly in heterogeneous settings such as

learning from complex MRI datasets (LeCun et al., 2015). Alternatively, deep learning (LeCun et al., 2015; Schmidhuber, 2015) overcomes limitations of classical machine learning by directly learning useful representations and features in a data-driven manner (i.e., feature learning).

Deep Learning Role: Deep learning refers to advancement in machine learning that is based on the architecture of Artificial Neural Networks. When the number of hidden layers in a neural network is extended, it becomes a Deep Neural Network (DNN) (see Figure 2.5). DNNs allow for the extraction of a complex hierarchy of features from input data while simultaneously performing a task, leading to impressive performance and generalizability (Goodfellow et al., 2016).

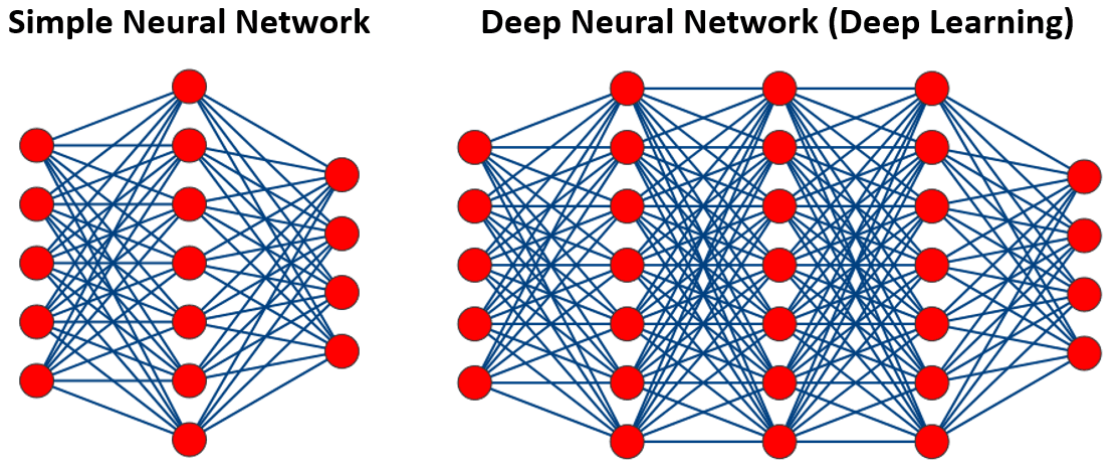


Figure 2.5: Architectures of two feed-forward fully-connected neural networks. A classical neural network containing one hidden layer and a deep neural network (deep learning) that has two or more hidden layers. Modern deep neural networks can be based on tens to hundreds of hidden layers.

In medical imaging, the success of deep learning is mostly triggered by the rapid progress in GPUs and development of CNNs (LeCun et al., 1998). In MRI processing and analysis, deep learning has been involved in each step of the entire workflow. In the lower levels of MRI pipelines, deep learning has been recently applied for image acquisition (Sun et al., 2016) and image reconstruction (Schlemper et al., 2017), estimation of quantitative tissue parameters (Rasmussen et al., 2018), MR fingerprinting (Cohen et al., 2018), image denoising (Bermudez et al., 2018), artifact detection (Küstner et al., 2018), image super-resolution (Zeng et al., 2018), and image synthesis (Wolterink et al., 2017). Moreover, deep learning techniques were heavily applied to higher-level,

down-stream applications including image registration (Balakrishnan et al., 2018; de Vos et al., 2019), image segmentation (Dora et al., 2017; Akkus et al., 2017), image retrieval (Qayyum et al., 2017; Pizarro et al., 2019), and disease prediction (Liu et al., 2018; Lu et al., 2018).

While deep learning-based methods have made significant strides in medical imaging applications, there are relatively few methods have been applied in infant MRI data (Mostapha and Styner, 2019). Deep learning methods, designed specifically for infant MRI, were applied in the segmentation of neonatal brain ($\leq$ 1 month) (Moeskops et al., 2016a; Rajchl et al., 2017), the segmentation of isointense infant brain ($\approx$ 6 months) (Zhang et al., 2015; Moeskops et al., 2016a; Nie et al., 2016; Moeskops and Pluim, 2017; Huang et al., 2017; Zeng and Zheng, 2018; Nie et al., 2018), and the early detection of NDDs such as ASD (Hazlett et al., 2017).

Current Open Challenges: With recent development in the field of deep learning, many innovative methods have been proposed to improve infant MRI brain image processing and analysis as presented above. The success of deep learning is attributed to its ability to discover general morphological and textural features in a data-driven way. Therefore, deep learning can handle different variabilities in infant MRI brain images that stem from complex brain anatomy and tissue appearance, imaging acquisition protocols, and pathological heterogeneity. However, there are still open issues to address, including the technical challenge of handling 3D medical images, with their added memory and computational consumption requirements. Current approaches to deal with this problem (treating 3D data as stacks of 2D images, patch-based training, and inference, or downsampling 3D images) are not ideal and come with the expense of performance degradation. In infant MRI, other critical challenges arise related to data size, class imbalance, and interpretability, as briefly discussed below.

[1] ***Data Size:*** In pediatric applications, available datasets are particularly small, as recruiting in such a vulnerable population is significantly more difficult than in adults or adolescents. This limits the power of deep learning methods, especially when compared to their success in computer vision applications that utilize large-scale datasets such as ImageNet (Deng et al., 2009). In infant MRI brain imaging, the lack of publicly available datasets and high-quality

labeled data imposes additional challenges that need to be addressed. Currently, there has been an effort to make data available through public research data repositories (e.g., NDAR for ASD research), which will allow deep learning models to discover more generalized features in NDDs. To avoid the challenge of training deep learning models from scratch in the context of limited datasets, an alternative is to fine-tune a deep learning model using a technique referred to as transfer learning (Tajbakhsh et al., 2016). For example, a deep learning model can be pretrained using a large dataset of labeled natural images (e.g., ImageNet (Russakovsky et al., 2015)). Nevertheless, it is still challenging to effectively utilize such 2D pre-trained networks for 3D infant MRI applications as such an approach would ignore the anatomical context in directions orthogonal to the 2D plane (Dolz et al., 2018; Xu et al., 2017). On the one hand, training data can be augmented by applying random transformations to the original data (Perez and Wang, 2017). Data augmentation aids in increasing the effective size of training samples and hence reduce overfitting by presenting random variations to the original data during training (Pereira et al., 2016; Akkus et al., 2016). However, augmenting infant MRI datasets is still challenging due to the regionally-heterogeneous image appearance and significant age-related intensity changes. In the presence of NDDs, additional sources of anatomical variabilities are introduced due to the heterogeneous nature of such disorders.

[2] ***Class Imbalance:*** Class imbalance refers to the number of training examples being skewed towards ”typical” as compared to ”atypical” or ”pathological” cases (Mazurowski et al., 2008). This problem is exacerbated when developing diagnostic or predictive approaches for infants with NDDs as prevalence rates are commonly small in NDDs (e.g. in ASD, $\leq 2\%$ for general population and $\leq 20\%$ for high-risk population). It has been recognized that the class imbalance problem has a substantial negative impact on training deep learning models. With imbalanced datasets, deep learning models tend to focus on learning the classes with a large number of examples, leading to poor performance for the classes with a small number of examples. In a diagnostic setting based on medical image data, misclassification costs are typically unequal; classifying a diseased sample (minority class) as typical (majority class)

implies significant consequences that should be avoided. Current methods for addressing class imbalance can be classified into those that operate on a training set by altering its class distribution, those that work on the classifier or algorithmic level while keeping the training dataset unchanged, and hybrid methods that that combine the two previously described categories. On the data level, methods which are commonly used in deep learning are based on oversampling, including techniques such as random minority oversampling (Janowczyk and Madabhushi, 2016), Synthetic Minority Over-sampling (SMOTE) (Chawla et al., 2002), and Local Synthetic Instances (Brown et al., 2015). Alternatively, undersampling can be used to solve the class imbalance problem by randomly removing samples from majority classes (Haixiang et al., 2017). In infant MRI, this motivated the development of undersampling methods that only remove redundant examples (Huang et al., 2016). One-class learning is an algorithmic strategy to achieve class balance by training deep learning models to recognize positive samples instead of discriminating between classes (Wang et al., 2015). Finally, hybrid approaches that combine multiple techniques from the previously described methods have been proposed to solve the class imbalance problem (Chawla et al., 2003; Havaei et al., 2017).

[3] ***Interpretation:*** In many medical image analysis tasks high-performance levels were achieved through highly flexible models with millions of weights. However, it is still challenging to compute an interpretation of how particular weights or inputs are contributing to the final model performance. Such interpretations are particularly crucial for successfully deploying deep learning models for early prediction of NDDs in a clinical setting. Recently, methods for the understanding of deep learning models have been increasingly proposed (Montavon et al., 2018; Lipton, 2016). These can be classified into two categories. The first approach is to find input images that maximize a class score at the output layer to visualize how the network represents a specific class (Erhan et al., 2009; Yosinski et al., 2015). One the other hand, methods have been proposed to visualize feature maps that explain the network classification or decision in response to a particular input image. The second approach is better suited for models trained on infant MRI datasets and in particular in NDDs, which are heterogeneous in

nature and hence require subject-specific interpretations. Several techniques were proposed to provide sample-specific interpretation, including the visualization of saliency maps for a given image (Simonyan et al., 2013), scores based on differences at the network output (Shrikumar et al., 2017), and pixel-wise importance estimated using sensitivity analysis (Zintgraf et al., 2017). Another class of methods try to understand individual decisions made by the classifier while assuming a black-box classifier or assuming a particular structure of how decisions are made (Ribeiro et al., 2016; Kumar et al., 2017). Similar methods were designed specifically for understanding CNNs through visualizing feature activity in intermediate layers (Zeiler et al., 2011; Shrikumar et al., 2017). The interpretation methods described above can produce valuable insights into what infant predictive deep learning models have learned; however, there is little agreement on how these methods should be evaluated for benchmarking. Interpretability can be evaluated in the context of a given application or using a proxy to provide a quantifiable evaluation (Doshi-Velez and Kim, 2017).

2.2 Cortical Surface Reconstruction and Shape Analysis

2.2.1 Cortical Surface Reconstruction

Motivation: Neuroanatomical abnormalities in the cerebral cortex are widely investigated by examining surface properties that can be measured directly from 3D volumetric images or the cortical surfaces reconstructed therefrom. Despite the simplicity of voxel-based approaches, they are limited since the spatial information of the cortical surface is intrinsically defined on a 2D manifold. Hence, voxel-based methods have limited ability to quantify cortical surface properties adequately. Surface-based methods, in which cortical surfaces are reconstructed from the volumetric 3D images, become necessary for appropriate quantification of surface properties. In early efforts to reconstruct a cortical surface model, a topologically correct model (e.g., a sphere) is deformed to match the Gray Matter (GM) and CSF boundary (MacDonald et al., 1994; MacDonald, 1997). Nevertheless, such an approach would lead to inaccurate representations, particularly at narrow and deep buried cortical folding. Such deficiencies can be attributed to the increasing difficulty

of designing general energy functions (without heuristics) that can attract the deformable model through the narrow opening of the sulci (Fischl, 2012).

Current Pipelines: Alternatively, current frameworks for cortical surface reconstruction, such as FreeSurfer (Dale et al., 1999) or CIVET (MacDonald et al., 2000) pipelines, identify the White Matter (WM) and GM boundary to aid in reconstructing the surface model. Both the FreeSurfer and CIVET pipelines share the similarity of guaranteeing a spherical topology for the reconstructed WM surface model. Furthermore, the GM surface is obtained by deforming the WM surface to fit the GM and CSF boundary, and consequently, a correspondence between the WM and GM surfaces is naturally established. In summary, the main steps involved in the overall surface reconstruction pipeline are as follows. *(1)* The intensities of the input MR images are first adjusted via bias field correction and intensity normalization. *(2)* The preprocessed images are then aligned to a common coordinate space (e.g., Talairach or MNI) for skull stripping and cortical structure segmentation. *(3)* A WM mask is generated by filling in the subcortical structures within the WM and then separating the results into left and right hemispheres. *(4)* An initial WM surface is obtained either using a rough triangulated tessellation (FreeSurfer and new versions of CIVET) or by deforming a topologically correct spherical model (old versions of CIVET). *(5)* The GM surface is reconstructed by deforming the constructed WM surface through an energy minimization procedure that involves using image intensity information and geometric constraints (e.g., curvature smoothing).

Current Challenges: In a low MR image resolution setting, cortical surface reconstruction becomes susceptible to Partial Volume Effects (PVE) in narrow sulcal fundi. In the presence of such PVE, surface models created using a tessellation technique from the image-based tissue segmentation will contain topological defects (i.e., holes and handles). In contrast, surfaces generated by starting from a spherical model would not suffer from such topological defects; however, defects in the WM and GM boundary can be observed due to imperfect energy function design. Hence, additional steps need to be employed for estimating PVE (Tohka et al., 2004) and correcting topological defects to ensure a correct topology (Fischl et al., 2001). Finally, although current pipelines were applied successfully to infant MRI datasets (see Figure 2.6), pipelines that

employ infant-specific steps are highly desirable. Recent efforts have been made toward that goal (Li et al., 2015; Kim et al., 2016a); however, publicly available tools are still missing.

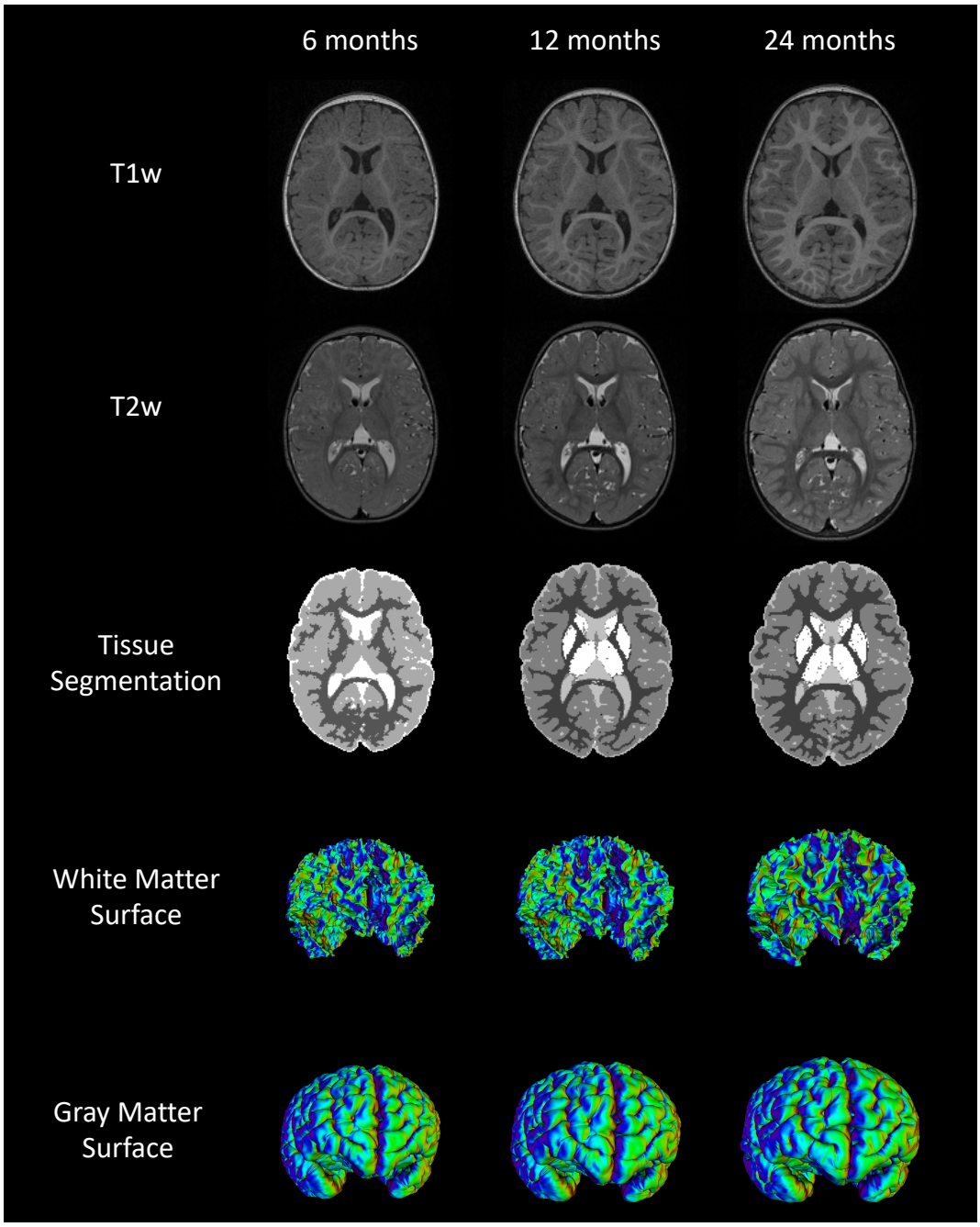


Figure 2.6: White matter and gray matter of a typically-developing infant, reconstructed from longitudinal structural MRI scans at 6, 12 and 24 months of age using the CIVET pipeline (Kim et al., 2005). Also shown: corresponding T1- and T2-weighted images and tissue segmentation results obtained using the AutoSeg pipeline (Wang et al., 2014).

2.2.2 Differential Geometry on Surfaces

Since this dissertation involves learning from surface properties extracted from 3D cortical surfaces, I briefly overview the basic concept of differential geometry on 2D manifolds.

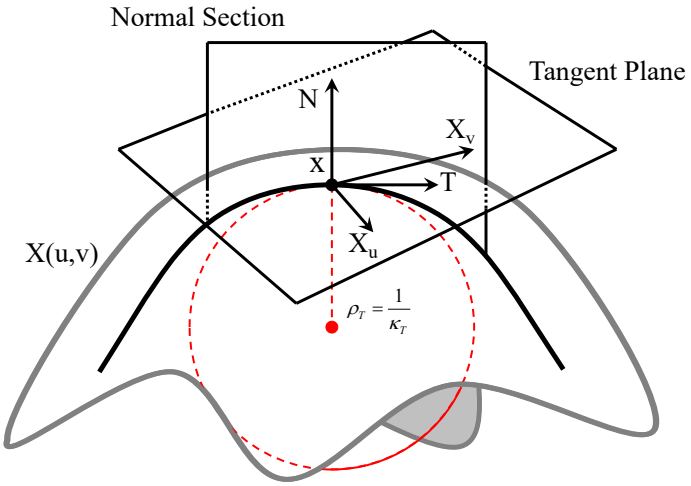


Figure 2.7: Illustration of the definition of the normal curvature on a smooth surface.

Curvature Metrics: Let $S \in \mathbb{R}^3$ be a smooth surface with parametrization $X(u, v) \in S$ such that the partial derivatives X_u and X_v are linearly independent, where $(u, v) \in \mathbb{R}^2$. Let T be the tangent direction at a given point x on S; the normal curvature κ_T at x is a real number that measures how S bends in the direction T as measured by the swing of the surface normal N at x. The normal curvature is easy to understand as it describes the curvature of the plane curve formed by the intersection of S with the plane passing through T, meeting S perpendicularly (Figure 2.7). Let D_T be the directional derivative in the T direction. The normal swing $D_T N$ at x along T can be decomposed into the normal curvature κ_T and its geodesic torsion τ_T according to

$$D_T N = \kappa_T \cdot T + \tau_T \cdot T^{\perp}, \tag{2.1}$$

where $T^{\perp}$ is the perpendicular direction to T in the tangent plane and

$$N = \frac{X_u \times X_v}{\|X_u - X_v\|}. \tag{2.2}$$

Similarly, walking along $T^{\perp}$, the normal swing at x in that direction can be expressed as

$$D_{T^{\perp}}N = \kappa_{T^{\perp}} \cdot T^{\perp} + \tau_{T^{\perp}} \cdot T. \tag{2.3}$$

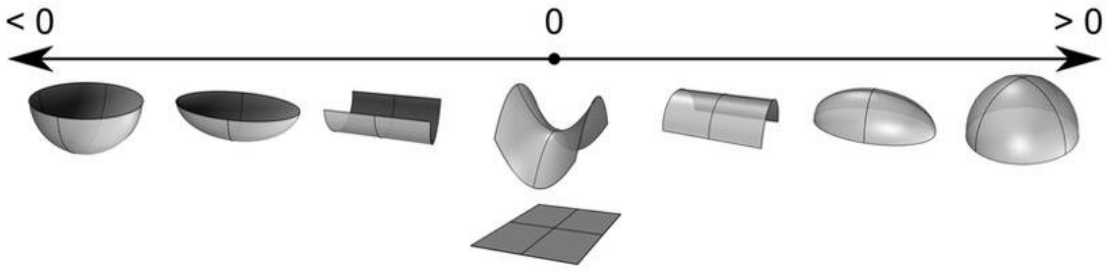
Equations 2.1 and 2.3 can be expressed in matrix form as follows

$$\text{II} = \begin{bmatrix} D_T N \\ D_{T^{\perp}} N \end{bmatrix} = \begin{bmatrix} \kappa_T & \tau_T \\ \kappa_{T^{\perp}} & \tau_{T^{\perp}} \end{bmatrix} \begin{bmatrix} T \\ T^{\perp} \end{bmatrix}. \tag{2.4}$$

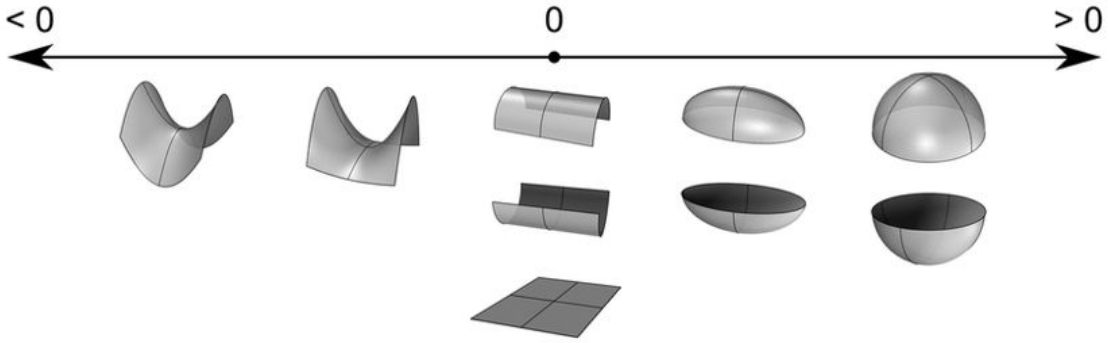
The symmetric 2×2 matrix shown in Equation 2.4 is called the second fundamental form of S and is denoted by the Roman numeral II. The eigenvalues of II provides what is called the principal curvatures κ_1 and κ_2, which captures the pure normal swing along the eigenvectors with no geodesic torsion (i.e., $\tau = 0$). The general shape morphology of a surface patch is determined by the different sign combinations of κ_1 and κ_2, while the magnitude of κ_1 and κ_2 describes how bent the surface is irrespective of shape morphology. Hence, each principal curvature separately does not provide a complete interpretation of local surface shape (Koenderink and Van Doorn, 1992). Therefore, to better describe shape proprieties, different geometric properties are described in terms of the principal curvatures ($\kappa_1 > \kappa_2$), namely mean curvature H, Gaussian curvature K, shape index S, and curvedness C (Koenderink and Van Doorn, 1992).

$$\begin{aligned} H &= \frac{1}{2}(\kappa_1 + \kappa_2), \quad K = \kappa_1 \cdot \kappa_2 \\ C &= -\frac{2}{\pi} \tan^{-1}\left(\frac{\kappa_1 + \kappa_2}{-\kappa_1 + \kappa_2}\right), \quad S = \frac{2}{\pi} \log\left(\sqrt{\frac{\kappa_1^2 + \kappa_2^2}{2}}\right) \end{aligned} \tag{2.5}$$

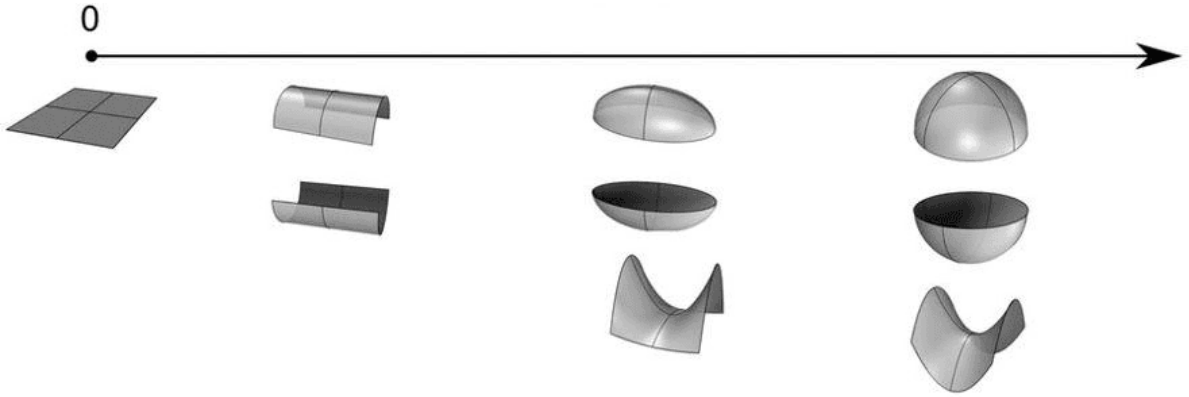
Figure 2.8 illustrates the characteristics of various curvature metrics, including their domains and how shapes are distinguished. H differentiates significantly areas of high and low curvature as well as convex and concave shapes, while K discriminates well between spherical and the saddle-like regions. On the other hand, S and C were proposed to describe the local shape more intuitively. C is less representative of a particular morphology but can measure how curved the surface is at each point. Finally, S is scale-independent and can differentiate between pure shape characteristics.



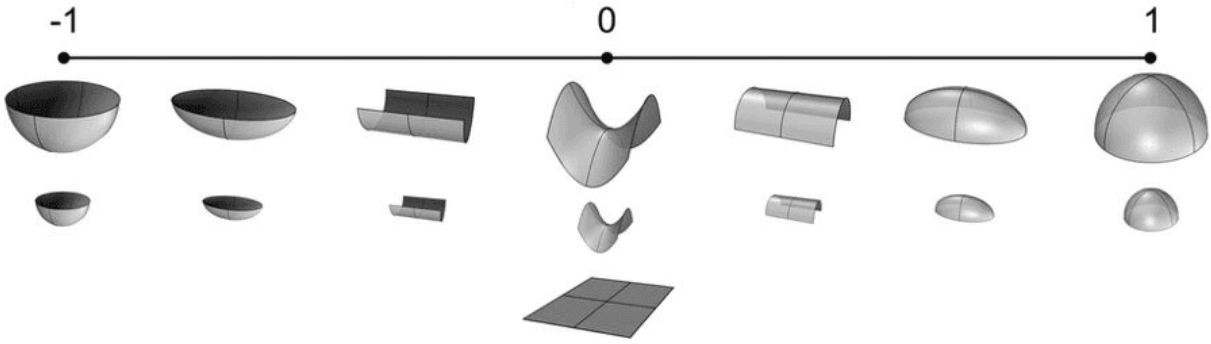
Mean Curvature (H)



Gaussian Curvature (K)



Curvedness (C)



Shape Index (S)

Figure 2.8: Shape characteristics of different curvature metrics. Each curvature metric highlights different aspect of the surface's underlying shape. Figure adapted from (Tsagkrasoulis et al., 2017).

Geodesics and Short Distance Path: Intrinsic distances on a surface S in 3D space is a critical geometric measure for understanding complex shapes. Such distances are computed along geodesics, which are a generalization of the concept of straight lines to a more general setting. Geodesics are also crucial because the shortest path between two points on S always follows a geodesic path, and the shortest distance between two points on S is called a geodesic distance.

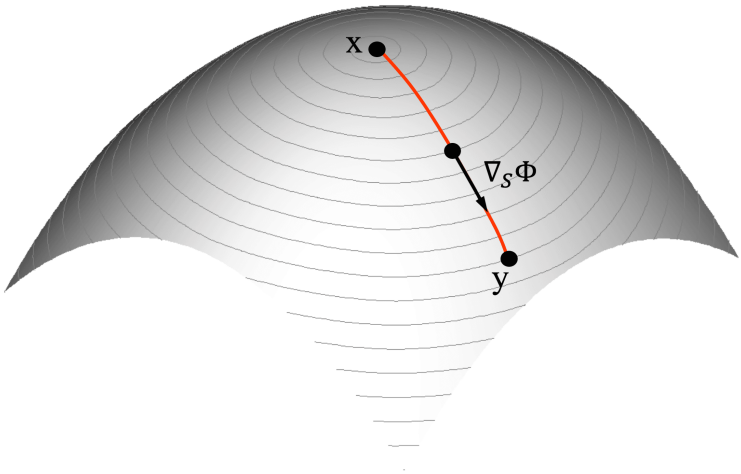


Figure 2.9: Illustration of geodesics and shortest path on a smooth surface.

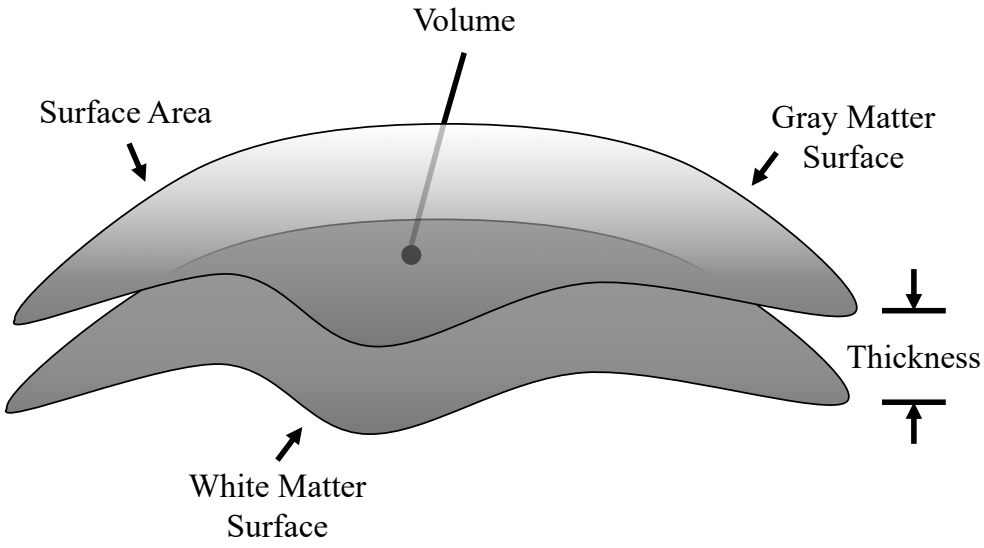
More formally, a geodesic on a smooth surface S is a differentiable curve that has zero geodesic curvature κ_g at each point on the curve. The geodesic curvature at a x is given by $\kappa_g = \kappa \sin\theta$, where κ is curve curvature at x, and θ is the angle between surface normal N and the normal to the curve n. For a surface and curve embedded in $\mathbb{R}^3$, this is equal to describing the geodesic curvature as the covariant derivative of the tangent vector to the curve (Balasubramanian et al., 2008). It can also be shown that the geodesic distance to a collection of seed points on S satisfies a non-linear differential equation, called the Eikonal equation (Kimmel and Sethian, 1998b), which is given by

$$\|\nabla_S \Phi\| = F. \tag{2.6}$$

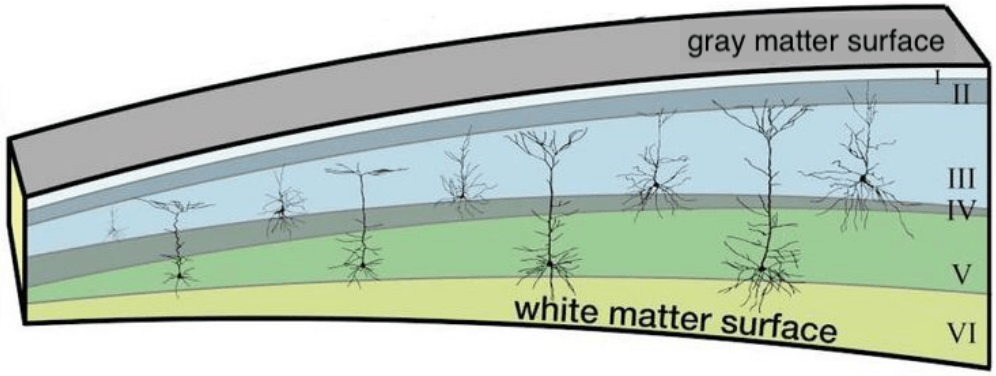
The solution $\phi(x, y)$ to the Eikonal is a weighted distance map from a set of seed points, where the value of Φ is provided, or assumed to be 0. The weights are given by the function F, which is a scalar positive function. When the function F is a constant, it can be defined as the distance function

to the set of seeds. Locally, geodesics are shortest paths since any perturbation of a geodesic curve will increase its length. Hence, the shortest distance between two points x, y on the surface are the minimal geodesics connecting those points. As shown in Figure 2.9, a geodesic path connecting x to y can be computed from the geodesic distance map Φ to one of the points by performing a gradient descent optimization on Φ.

2.2.3 Cortical Surface Proprieties



(a) Cortical Thickness, Surface Area, and Volume



(b) Cortical Layers

Figure 2.10: (a) The geometrical connection between cortical thickness, surface area, and volume. (b) The cortical layers of the cerebral cortex which, arranged parallel to the cortical surface.

Overview: The cerebral cortex refers to the outer layer of the gray matter, which is a highly folded structure with the degree of folding related to an evolutionary necessity to increase surface area without increasing the intracranial size (Magnotta et al., 1999). The cerebral cortex plays a critical role in cognitive development and decline, and alterations in cortical volume have been associated with NDDs (Ismail et al., 2016; Mensen et al., 2017) and neurodegenerative diseases (Dickerson et al., 2008). As shown in Figure 2.10, the components of cortical volume, namely Cortical Thickness (CT) and Surface Area (SA), have separable developmental trajectories and are related to distinct neurobiological processes. Though cortical volume measurements are affected by both CT and SA, volume was shown to be more closely linked to SA than CT (Winkler et al., 2010). Hence, in imaging studies conducting surface-based analysis, local measurements of SA and CT need to be examined separately and are preferred over gray matter volumes. Meanwhile, an improved understanding of the factors that dictate the folding of the cerebral cortex is required, as irregular folding patterns observed in individuals affected by NDDs emphasize the clinical importance of understanding the folding process.

Cortical Surface Area: Surface area-based analyses are influenced by the quality of the applied surface registration as well as the interpolation method used to resample the surface to a common resolution. The choice of interpolation method is particularly crucial as any involved surface resampling needs to preserve the amount of area at local, regional, and global scales (i.e., mass-conservative) (Winkler et al., 2012). Another consideration when analyzing SA is where to extract the local measurements; possible choices include the WM surface, the GM surface, and the middle surface (i.e., a surface that runs at the mid-distance between WM and GM). The WM surface is typically preferred over the GM surface as it represents a biologically meaningful cortical layer, leading to measurements that are directly related to brain morphology. Moreover, the SA extracted from a WM surface tends to be less sensitive to cortical thinning or thickening (Winkler et al., 2012). Even though the middle surface might not match any specific cortical layer, it avoids over or under-representing gyri or sulci, leading to reliable SA measurements (Van Essen, 2005).

Additionally, most studies choose to measure SA at the middle surface, as it captures a combination of both WM and GM surface properties, which is preferred to analyzing both of them separately.

Cortical Thickness: In the absence of a unique definition of a CT measure, several surface-based methods for measuring CT from MRI brain data have been proposed (Wagstyl and Lerch, 2018). Most of the existing methods define the WM surface as the inner surface and the GM as the outer surface. Hence, current methods for measuring CT are sensitive to the quality of the provided tissue segmentation (i.e., WM/GM and GM/CSF boundaries), leading to uncertainties in the obtained measurements due to loss of available information (Aganj et al., 2009). Following the construction of inner and outer surfaces, methods differ in their approach to computing the required thickness measure. Coupled surface methods (Fischl and Dale, 2000; MacDonald et al., 2000) employ the Euclidean distance between corresponding points on the inner and outer surfaces as the CT measure, which may lead to an overestimation of the thickness. To mitigate this, closest point methods (Miller et al., 2000) would first find for each point on the inner or outer surface, the closest point on the other surface and then would compute the Euclidean distance between them. To tackle the absence of symmetry in closest point-based CT measurements, methods that solve Laplace's equation in the GM region was proposed (Yezzi and Prince, 2003; Haidar and Soul, 2006). In these methods constant, but opposite-in-sign boundary condition potentials are set on each of the two surfaces. The CT is then defined as the length of the streamline obtained by tracing the gradient of the resultant vector field, which leads to a uniquely assigned measure at every point. Compared to coupled surface and closet point methods that utilize straight-line distances, streamline based CT leads to biologically meaningful measure as it captures the curved nature of cortical layers. Studies conducted to validate current CT methods on simulated data (LERCH et al., 2003) indicated significant differences among different cortical thickness metrics and emphasized the critical importance of an appropriate smoothing of obtained CT measures for optimal results.

Cortical Surface Folding: The ability to quantify the geometrical complexity and variability of the human cerebral cortex is essential in understanding brain structure and function (Fischl et al., 2007). Furthermore, a quantitative cortical folding descriptor can explain the underlying mechanism

of cortical gyrification (Toro and Burnod, 2005), which offers insights to understand abnormal cortical folding in brain disorders (Van Essen et al., 2006). Over the years different methods have been proposed to quantify cortical folding during early brain development. Methods utilizing local geometric features (e.g., curvature information) have been proposed to quantify cortical folding at a local scale or a more global scale by integrating such features across different brain regions or an entire hemisphere. Some of these methods provide measures that depend on the cortical surface area (Ajayi-Obe et al., 2000; Van Essen and Drury, 1997); this dependence limits their usefulness in developmental studies with brain size variations associated with gestational age and between different NDDs populations. Alternately, measures of global curvature that are independent of surface area have also been proposed to study cortical folding in early brain development (Rodriguez-Carranza et al., 2008; Awate et al., 2010). However, such measures still were unable to distinguish between normal and aberrant cortical development (Shimony et al., 2016).

On the other hand, non-curvature-based measures for analyzing cortical folding were proposed, including the Fractal Dimension (FD) and the Gyrification Index (GI). The global FD provides a numerical value to quantify the cortex complexity by assessing its fractal properties (i.e., the statistical similarity in shape) over a range of spatial scales (Free et al., 1996). Though FD can provide a folding measure without the need for a reference model, it was shown that FD is highly sensitive to the quality of the reconstructed surfaces (Free et al., 1996). Cortical folding can also be described using the global GI measurement, which is defined as the area ratio between the outer cortical surface (i.e., cerebral hull surface) and the GM surface (Armstrong et al., 1995). The main advantage of global GI is its ability to provide scale-invariant folding measures; however, they generally cannot yield localized measures of complexity. A window or a kernel size needs to be defined to perform a local analysis of GI (local GI). Different local measures of GI have been proposed using kernels that are defined either by Euclidean distances (Su et al., 2013) or a quasi-geodesic N-ring neighborhood (Li et al., 2014). However, for a reasonable local GI analysis, the designed kernels need to include at least one sulcal or gyral region. Hence, larger kernel sizes are typically chosen, running the risk of capturing regions across multiple sulci and gyri, i.e., combining

areas that may be functionally quite different. To tackle this issue, local GI measures that utilize spatially varying kernels that encode cortical folding patterns were proposed (Lyu et al., 2018).

Another fine-scale measurement of cortical surface complexity, referred to as the Shape Complexity Index (SCI) (Kim et al., 2016b), was proposed as an alternative to quantifying cortical folding. In contrast to the local GI, the SCI can detect whether a sulcal or gyral region experiences a widening or deepening process as well as capture the development of secondary and tertiary sulci. As shown in Figure 2.11, the shape index S is first calculated at each point on the surface, which yields a score ranging from -1 to 1. The S score is then dichotomized into predefined different geometric, topological situations. The SCI score is then assigned to reflect the S variability within a local region, measured from the S histogram within a given geodesic distance. Particularly, the discrete Earth Mover's Distance (EMD) is used to compute the difference between the observed S distribution and each of the defined basic topological settings; then the minimal EMD at each point is used as a measure of surface complexity. This results in an SCI score that ranges from 0 to 1 since the minimum EMD is maximally 1 over the selected geometric settings (the observed distribution is balanced and bi-modal at the S extrema of 1 and -1).

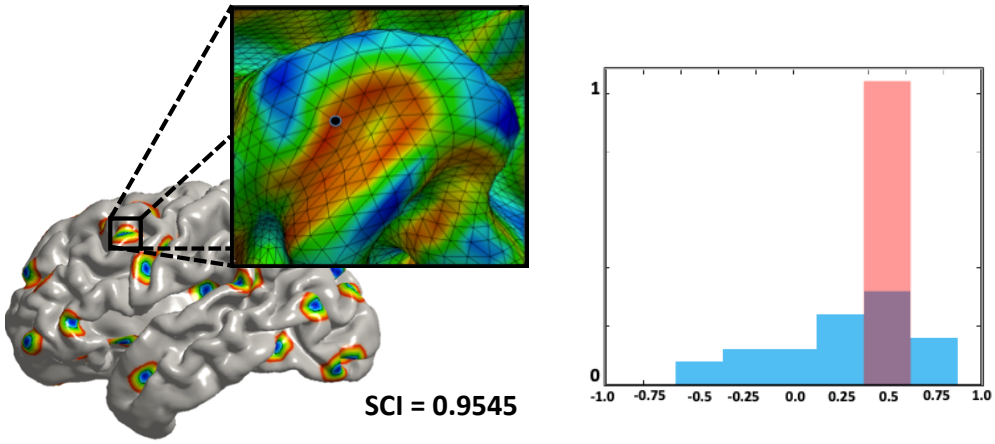


Figure 2.11: An example illustrating how the SCI captures local shape complexity within a given geodesic distance-based neighborhood. The local distribution of S (blue color histogram) is quantified within the defined neighborhood, and SCI is defined as the EMD to the best fitting basic model distribution (red color histogram). Figure adapted from (Kim et al., 2016b).

2.3 Common Deep Learning Architectures

In this section, I briefly overview discriminative and generative deep learning architectures that are being utilized in this dissertation to effectively learn from input complex neuroimaging datasets.

Discriminative and Generative Deep Learning: The key to the success of machine learning models rests in their ability to extract high-level and semantically meaningful features from the complex neuroimaging datasets. As shown in Figure 2.12, two main approaches have been proposed to achieve that goal, namely learning using discriminative models or generative models. Supervised learning using discriminative models attempts to directly learn the posterior, $p(y|x)$, of the inputs x and the label y, or learn a direct map from inputs x to the class labels y. The intution behind these models is that a good performance on the given supervised task must indicate that useful and meaningful high-level features have been learned.

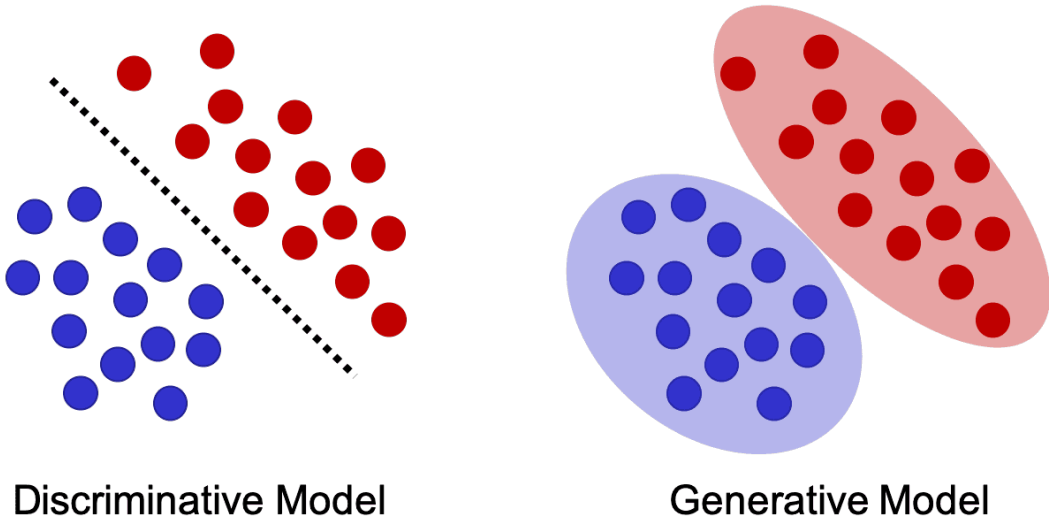


Figure 2.12: Illustration of generative vs. discriminative learning approaches. Discriminative models focus on classification boundaries between different classes, whereas generative models emphasize the data generation process in each particular class.

Training discriminative models is relatively simple as it typically involves optimizing for a simple final objective function (e.g., classification error). Yet, there are still open questions on the generalization performance of discriminative models; mainly, the availability of plentiful labeled data is critical for such models to generalize well (Bernardo et al., 2007). In contrast, unsupervised

generative learning seeks to model the input data, $p(x)$, without expecting the availability of training labels. In a supervised setting, generative models can also be trained to learn a model of the joint probability, $p(x, y)$, where predictions can be made by first computing $p(y|x)$ using Bayes rule and by then picking the most likely label y (Ng and Jordan, 2002). With generative learning, the intuition is that if the model can efficiently and faithfully generate the data, then a good latent representation must have been learned. Generative models are promising in providing solutions for problems associated with learning from complex neuroimaging datasets; however, such models are challenged by high-dimensional data with complicated data manifolds. Below, deep discriminative (e.g., MLP, CNN, U-Net, DenseNet) and generative (e.g., VAE, GAN) architectures are reviewed.

Multilayer Perceptron: A neural network that has at least one hidden layer and all layers are fully-connected is referred to as a Multilayer Perceptron (MLP) (see Figure 2.5). MLPs is a fundamental tool that provides the foundation of many advanced techniques in deep learning. Mathematically, an MLP is a composite function of alternating affine and non-linear transformations, which can be expressed as

$$f(x) = g(W_l \phi_{l-1}(W_{l-1}(\cdots \phi_1(W_1 x + b_1) \cdots + b_{l-1}) + b_l), \tag{2.7}$$

where W_i, b_i, ϕ_i are the i^{th} layer's weights, biases, and activation function respectively and g is the activation function at the output layer. According to the universal approximation theorem (Hornik, 1991), an MLP with one or more hidden layers and non-linearities (e.g., sigmoid) with a finite number of units can approximate any continuous function with a desired non-zero error given appropriate parameters. Such a network's parameters are learned by iteratively minimizing the loss function with parameters being updated (e.g., using stochastic gradient descent – SGD) via back-propagation until convergence is reached. However, in practice, finding such optimal configuration is not guaranteed, which may lead to poor approximation quality. Moreover, MLPs still suffer from the need to train a large number of parameters which can become unfeasible when dealing with high-dimensional data if overfitting to be avoided.

Convolutional Neural Networks for Classification Tasks: Unlike MLPs, where inputs are always in vector form, CNNs maintain and utilize the structural and spatial information among neighboring pixels or voxels in the input 2D or 3D images. Also, as illustrated in Figure 2.13, CNNs limit the degrees of freedom of the deep learning model through exploiting a local receptive field, weights sharing, and sub-sampling techniques. Hence, this will result in models that can learn from limited datasets usually present in medical applications in a way that is less prone to overfitting.

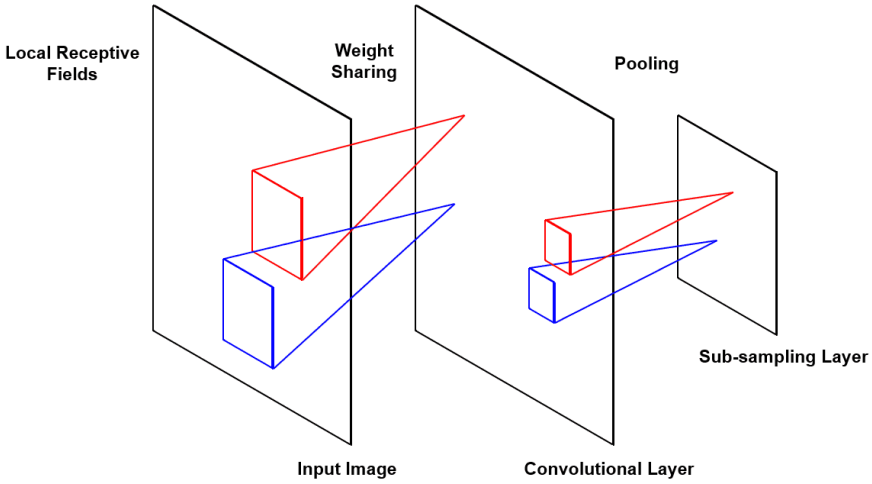


Figure 2.13: Illustration of how convolutional neural networks control the model complexity through three fundamental mechanisms, namely, local receptive field, weight sharing, and subsampling.

Although CNNs were first introduced in 1989 (LeCun et al., 1989), they did not gain interest until the introduction of deep CNN architectures, such as AlexNet (Krizhevsky et al., 2012), which accomplished remarkable results in the ImageNet (Russakovsky et al., 2015) competition in 2012. As shown in Figure 2.14, AlexNet represents a currently typical CNN architecture that consists of a sequence of layers of convolution, pooling, activation, and fully connected operations. In the image classification task, AlexNet almost halved the error rates of the formerly best-performing methods (Russakovsky et al., 2015). Since then, CNN architectures are increasingly deeper, which resulted in improved error rates, often achieving human performance levels (He et al., 2016; Hu et al., 2017). Below, the main building blocks of CNNs are briefly discussed.

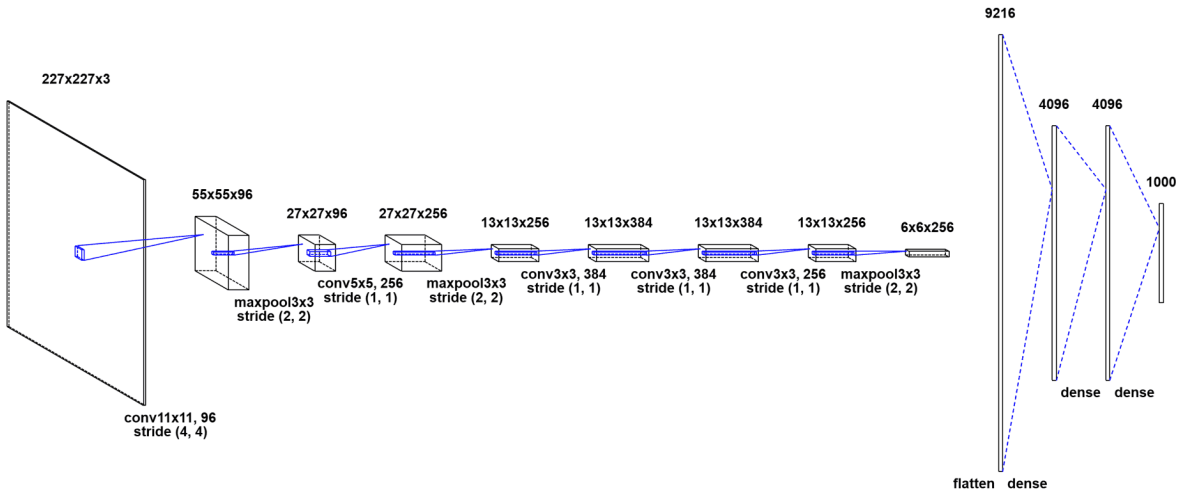


Figure 2.14: AlexNet, a convolutional neural network designed by Alex Krizhevsky (Krizhevsky et al., 2012). it achieves a 10.8% better error performance than the runner up in the 2012 ImageNet competition that used a shallow machine learning method. AlexNet has eight layers; the first five were convolutional layers, max-pooling layers follow some of these, and the last three layers were fully connected. AlexNet was written with CUDA to be trained using GPUs, which made the computationally expensive training of such a complex network feasible.

1. *Convolutional Layer:* A convolutional layer is designed to detect localized features extracted at different locations in the input feature maps using learnable kernels or filters. Notably, the input I to the convolutional layer is convolved with a set of small parameterized filters, also called kernels K. In practice, most deep learning libraries utilize cross-correlation is instead of the convolution operation (cross-correlation does not "flip" the kernel). Hence, the output at i, j, k voxel coordinates produced by a 3D convolution can be written as

$$(I * K)(i, j, k) = \sum_{1}^{m} \sum_{1}^{n} \sum_{1}^{p} I(i + m, j + n, k + p) K(m, n, p), \tag{2.8}$$

where m, n, p describe the kernel size that is typically small and isotropic. Note that by having each filter share the same weights over the entire input domain, i.e., translational equivariance at each layer, one achieves a significant reduction in the number of trainable weights. In practice, the behavior of convolutional layers is controlled by *(1)* the kernel size, which reflects locality assumptions of the low-level features, *(2)* the number of kernels, which controls width the CNN

model, *(3)* padding type, which controls how convolution is performed on the borders of the input, and *(4)* stride value, which controls the step of the convolution filter.

2. *Activation Layer:* The output of convolutional layers is passed through non-linearities introduced by an activation function. This makes it possible for the entire neural network to approximate almost any nonlinear function similar to an MLP. The Rectified Linear Unit (ReLU), defined as $\text{ReLU}(x) = \max(0, x)$, and its alterations, such as Leaky ReLU or parametric ReLUs, are among the most commonly used activation functions (Clevert et al., 2015). These functions enabled the training of deeper models in a way that avoids vanishing gradient problems typically found when using the more traditional tanh and sigmoid activation functions.

3. *Pooling Layer:* Feature maps produced by feeding the data through convolutional and activation layers are then typically downsampled or pooled using a pooling layer. Pooling operations are applied to small grid regions to produce a single number summary for each region. This summary measure can be computed by using the max function (max-pooling) or the average function (average pooling). As a small shift of the input image would result in insignificant alterations to the activation maps, pooling layers provide CNNs with translational invariance. Recently, strided convolutions are more commonly used as an alternative to achieve the required downsampling. Such approach simplifies the network by removing pooling layers while maintaining similar performance levels (Springenberg et al., 2014).

4. *Dropout Layer:* In addition to the commonly used ℓ_1 and ℓ_2 losses as regularization techniques in DNNs, dropout layers were introduced to provide additional regulation as well as a boost in the overall performance. Dropout (Srivastava et al., 2014) is an averaging technique based on a stochastic sampling of DNNs that allow for a performance boost through ensembling sampled models. Particularly, during the training phase, network units are randomly removed (controlled by dropout rate $p \in [0, 1]$) along with all associated connections. During the testing phase, all nodes are inserted back to use the full power of the model; however, activation maps need to be scaled by p or invert the dropout during the training phase.

5. *Batch Normalization Layer:* These layers are placed after activation layers, providing normalized activation maps by first subtracting the mean μ and dividing by the standard deviation σ for each training batch, followed by a learnable scale γ and shift β (Ioffe and Szegedy, 2015):

$$\begin{aligned} \hat{x} &= \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \\ y &= \gamma \cdot \hat{x} + \beta \end{aligned} \tag{2.9}$$

   By forcing the network to periodically standardize activation maps as the training batch runs through batch normalization layers, the network becomes better regularized, and the training gets simpler leading to faster convergence (Ioffe and Szegedy, 2015). Note that in modern DNNs, in addition to batch normalization, other normalization methods have been proposed such as weight normalization (Salimans and Kingma, 2016), instance normalization(Ulyanov et al., 2016), layer normalization (Ba et al., 2016), and group normalization (Wu and He, 2018).

Convolutional Neural Networks for Semantic Segmentation Tasks: In classification tasks, since the main goal is to predict the class of the whole image, there is more emphasis on learning abstracted representations. In contrast, for semantic segmentation tasks, the focus should be made on both abstract information as well as spatial information. As the information flows forward through pooling layers, the spatial information will decrease, and meanwhile, the feature map becomes higher level and more abstract. Hence, for accurate voxel-wise classification, a balance between spatial information and abstract features needs to be established. Current semantic segmentation models utilize the so-called encoder-decoder architecture based on a Fully Convolutional Neural Network (FCN). An FCN architecture can be trained end-to-end to produce dense pixel output with fewer network's parameters as it avoids the use of fully connected layers (see Figure 2.15). Thus, FCNs can simplify and accelerate both the learning and the inference of deep segmentation neural networks. In particular, in this architecture type, a classification is given for each voxel of the input image. It include an encoder network that extracts compact high-level features using convolutional

and pooling layers and include a decoder network that upsamples the higher-level features extracted by the encoder network using deconvolution layers. This produces pixel-wise class probabilities that can be used to classify pixels and produce the required segmentation map.

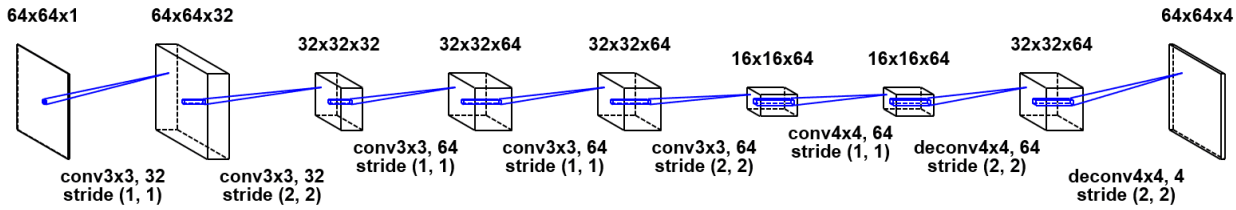


Figure 2.15: An example of encoder-decoder architecture based on a fully convolutional network that was used to for single modality infant MRI tissue segmentation (Nie et al., 2016).

The U-Net (Ronneberger et al., 2015), which was originally designed for biomedical image segmentation, is similar to FCN in terms of both down-sample up-sample structure. As shown in Figure 2.16, the U-Net modifies the original FCN architecture by utilizing both long and short skip connections similar to the ones introduced in residual networks (He et al., 2016), which enables the network to learn from fewer training images and produce more precise segmentation. Another CNN architecture, called DenseNet (Huang et al., 2017), which uses a combination of dense blocks and pooling layers, was also introduced. As shown in Figure 2.17, each dense block is an iterative concatenation of previous feature maps, which can be seen as an extension to the residual blocks (He et al., 2016) that instead performs iterative summation of previous feature maps. However, DenseNets allow for a thinner and more compact networks, i.e., smaller number of channels. Additional advantages of DenseNets include avoiding vanishing-gradient problems, encouraging feature reuse, and improving feature propagation in both forward and backward directions. Such DenseNets qualities make them a particularly great fit for semantic segmentation tasks where skip connections are naturally induced and multi-scale supervision is performed. Recently, FC-DenseNet architecture (Jégou et al., 2017) was introduced to extend the DenseNet architecture to work as FCN for semantic segmentation, while avoiding the feature map explosion when building an upsampling path. By only upsampling the feature maps created by the preceding dense block, the number of dense blocks at each resolution of the upsampling path

becomes independent of the number of pooling layers. Similar to a U-Net, information is passed by means of a standard skip connection between the downsampling and the upsampling paths.

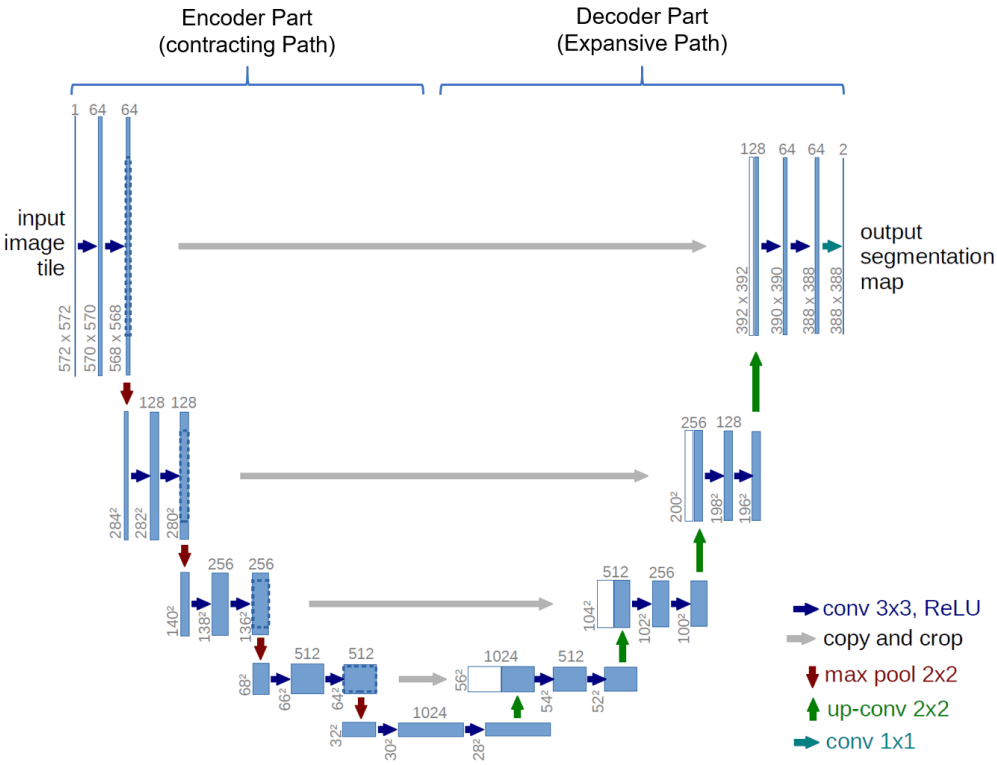


Figure 2.16: The U-Net architecture proposed for biomedical image segmentation tasks. The U-Net architecture formed of a contracting path to capture context through compact representations and an expanding path that allows accurate localization, which enables learning from few examples and yields superior segmentation results. Image adapted from (Ronneberger et al., 2015)

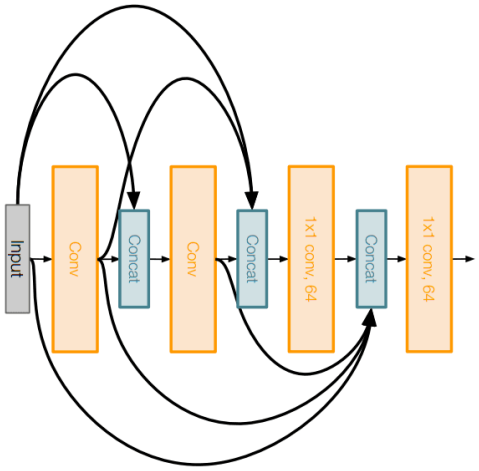
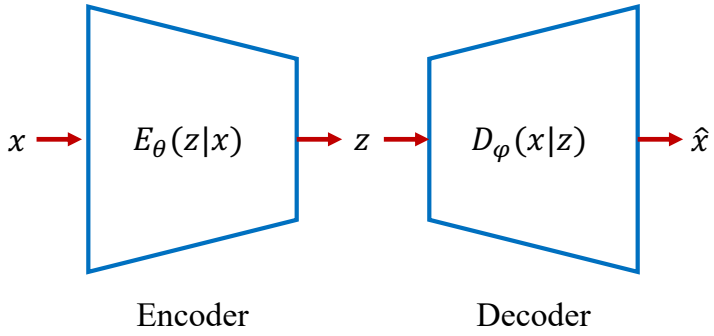
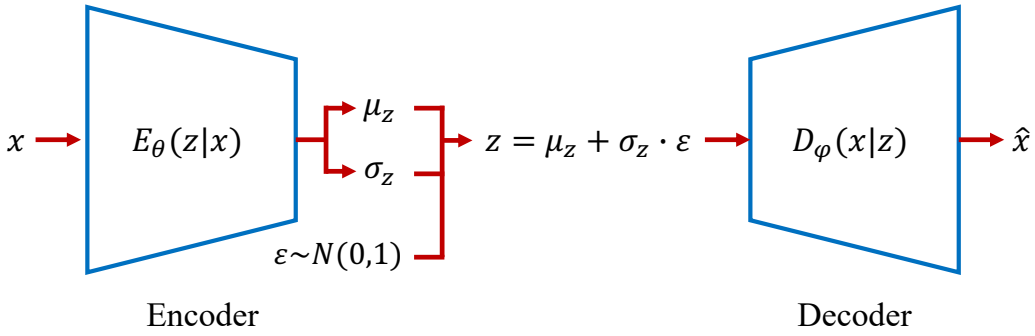


Figure 2.17: Illustration of a 4-layer dense block (Huang et al., 2017). Given a growth rate of k, the first layer is applied to the input to create k feature maps. Then, the next layer is applied to create new k feature maps, which is also concatenated to the previously created feature maps. This process is repeated 4 times, and the block output is a concatenation of outputs from the 4 layers (i.e., $4*k$ feature maps at the dense block output).

Variational Autoencoders: As shown in Figure 2.18a, Autoencoders (AEs) are neural networks used for unsupervised representation learning. An AE is composed of two networks, an encoder network E which is parametrized by θ and a decoder network D with parameters ϕ; the first network encodes the input x into a low-dimensional encoding (latent) space z, while the second network takes z as its input with the purpose of reconstructing the input $\hat{x}$. By restricting the dimensionality of the encoding z, the AE is forced to learn efficient representations of the inputs. Such a model is trained by minimizing a reconstruction loss ℓ (typically ℓ_1 or ℓ_2) between the input x and its reconstruction $\hat{x}$. On the other hand, as shown in Figure 2.18b, Variational Autoencoders (VAEs) (Kingma and Welling, 2013) present an alternative formulation of the traditional AEs to provide a probabilistic interpretation of the learned latent space.



(a) Autoencoder



(b) Variational Autoencoder

Figure 2.18: The general architecture of (a) an autoencoder and (b) a variational autoencoder. Autoencoders can extract features whose dimensionality is much fewer than the samples, while variational autoencoders is a type of the variational Bayesian inference which has encoder and decoder structure as with autoencoders (Kingma and Welling, 2013).

In VAEs, the encoder network E now outputs two parameters μ_z and σ_z used in sampling a point in the Gaussian distribution $N(\mu_z, \sigma_z)$ modeled by E. To find the optimal network parameters θ and ϕ, a variational lower bound loss ℓ_{VAE} need to be minimized. This involve minimizing the reconstruction error between x and $\hat{x}$ and the KL divergence term ℓ_{KL} of the distribution modeled by E and some prior distribution $p(z)$ such as the standard Gaussian distribution $N(0, I)$.

$$\ell_{VAE} = \ell_{rec} + \ell_{KL} = \ell(x, \hat{x}) + KL(N(\mu_z, \sigma_z)||p(z)). \tag{2.10}$$

Directly sampling z from $N(\mu_z, \sigma_z)$ would result in noncomputable gradients which would prevent backpropagating the error during training. The so-called "reparameterization trick" is used to solve this problem by moving the sampling operation to the input layer. In particular, $\varepsilon \sim N(0, 1)$ is added to the model to make the sampling operation differentiable by multiplying it to σ_z.

Advantages of using VAEs include that there is a clear and recognized way to evaluate the quality of the model using the estimated log-likelihood (variational lower bound). However, due to the strong assumptions and approximations involved, they may lead to suboptimal models and the generated images tend to be more blurred than those coming from competing generative methods.

Generative Adversarial Networks: In contrast to VAEs that explicitly try to approximate the true data distribution, Generative Adversarial Networks (GANs) implicitly specify probabilistic models that describe a stochastic procedure to directly generate data. Such a framework for constructing generative models can provide samples that are sharp and compelling without having to specify a likelihood function. As shown in Figure 2.19, GANs contain two simultaneously trained networks, namely a generator network G and a discriminator network D with parameters θ and α respectively. GAN training is based on a game-theoretic scenario, where two players are competing in a zero-sum game. Particularly, the generator G task, conditioned on a sampled noise variable $z \sim p(z)$, is to generate samples that follow the real distribution of the data $p(x)$. On the other hand, the discriminator D tries to distinguish between real data x and the fake generated data $\hat{x}$ by outputting a probability score $D_\phi(x) \in [0, 1]$. In this adversarial arrangement, training is done simultaneously on both G and D by alternating updates of G with updates of D. Specifically, D is

trained to maximize the probability of predicting the correct class label associated with both real data and fake data, while G is trained to maximize D uncertainty by minimizing $\log(1 - D(G(z))$. This results in the following objective function $V(G, D)$ defined by the following minimax game:

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p(x)}[\log D(x)] + \mathbb{E}_{z \sim p(z)}[\log(1 - D(G(z))]. \tag{2.11}$$

The training process of GAN can be viewed as finding a Nash equilibrium, where the discriminator D is incapable of recognizing real data from fake data, which means that the generator G has successfully approximated the real data distribution $p(x)$. An advantage of this GAN setting is that both generator and discriminator can be trained with backpropagation, without the need for unwieldy inference. Though GANs show such essential advantages over other generative models, training GANs is difficult as it may lead to oscillatory behavior and suffer from a problem referred to as mode collapse in which all latent space inputs are mapped to the same data point (Dosovitskiy and Brox, 2016). Moreover, reaching Nash equilibrium does not mean that the generator has matched the real data distribution, but only that the discriminator has reached its highest classification abilities given the current setup. Hence, more advanced models have been proposed that use alternative formulations to solve this problem (Arjovsky et al., 2017; Gulrajani et al., 2017).

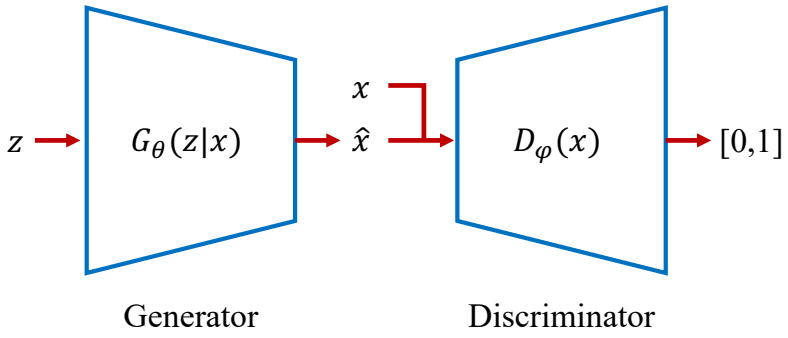


Figure 2.19: Architecture of a generative adversarial network. The generator and discriminator play a game against each other, where the objective of the Generator is to produce samples that would look like a real one, while the goal of the discriminator is to be able to tell the difference between generated and real samples (Goodfellow et al., 2014).

CHAPTER 3: EXTRACTION OF LOCAL EXTRA-AXIAL CEREBROSPINAL FLUID

3.1 Overview

This chapter describes a novel approach for extracting a local measure of CSF from structural MRI. Before presenting the details of the proposed extraction framework, I first outline the importance of CSF circulation in regulating healthy brain development and function and highlight the current limitations in quantifying MRI-based CSF biomarkers.

General Information on CSF: Produced in the brain, CSF is a clear, colorless fluid that circulates in the brain, acting as a cushion or buffer for the brain, providing necessary mechanical and immunological protection to the brain inside the skull. In addition to its protective purpose, recent findings have shown that CSF circulation plays a crucial role in brain development and function prenatally and during the lifespan (Shen, 2018). Figure 3.1 illustrates CSF circulation in the subarachnoid space around the brain, spinal cord, and in the ventricles of the brain. Following CSF production by the choroid plexus in the ventricles, it circulates from the lateral, third, and fourth ventricles to the cisterns of the brain. CSF flow continues to the subarachnoid space, where it covers the cortical convexities of the brain. CSF then flows into the parenchyma, where it interacts with the interstitial fluid in the perivascular space, alongside blood vessels in the brain. Finally, CSF returns to the subarachnoid space where it is absorbed through meningeal lymphatic vessels and arachnoid granulations. It is now realized that a healthy CSF circulation serves two essential functions to the brain. The first is a regulatory function through the delivery of growth factors and signaling molecules critical to brain development (Lun et al., 2015). CSF circulation also provides a cleaning or filtration mechanism of the brain by cleaning of the brain by removing neurotoxins and metabolic waste byproducts of neuronal function that would otherwise accumulate (Iliff et al., 2012). Therefore, disrupted CSF circulation has been shown to be involved in neurodegenerative

conditions such as Alzheimer's Disease, ischemic and traumatic brain injury, and neuroinflammatory conditions such as multiple sclerosis (Simon and Iliff, 2016). More recently, abnormalities in CSF circulation were also linked to NDDs, particularly to atypical brain development associated with ASD (Shen et al., 2013, 2017, 2018).

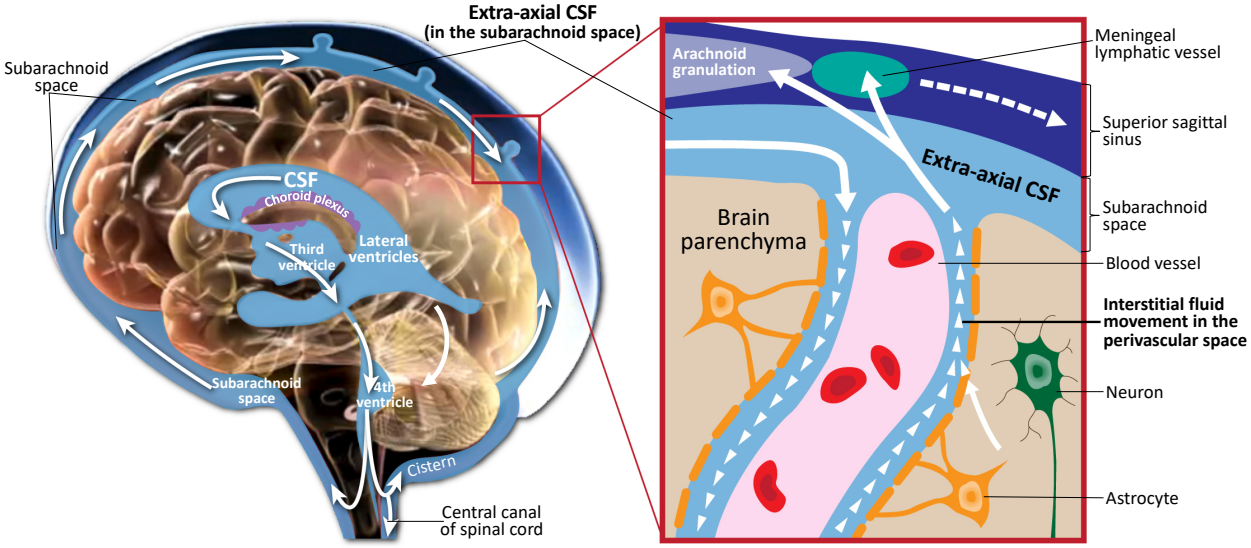
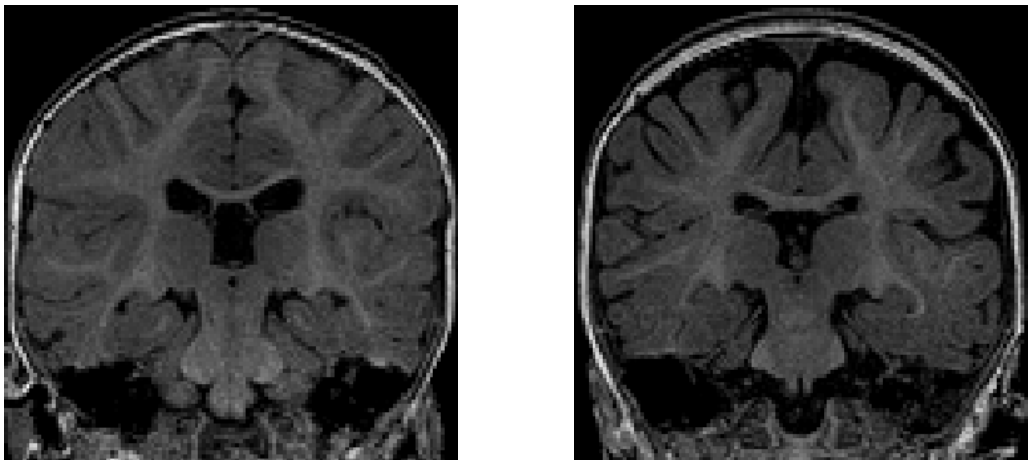


Figure 3.1: Illustration of cerebrospinal fluid circulation in the ventricles, subarachnoid space surrounding the brain, interstitial space within the brain parenchyma, and draining into the meningeal lyphatic system. Figure adapted from (Shen, 2018).

MRI-based CSF Biomarkers: The volume of different CSF chambers can be accurately measured from sMRI in-vivo scans, which can serve as indirect markers of abnormal CSF production and absorption. Current findings in lateral ventricles Lateral Ventricles volume related to ASD have indicated no significant group differences in children (Vidal et al., 2008) or adults (McAlonan et al., 2002). In contrast, there is evidence for an increased volume of CSF associated with ASD located outside the ventricles (Hallahan et al., 2009), as well as an increased volume of global CSF across the entire brain (McAlonan et al., 2004). Notably, studies of infants at high familial risk for ASD have indicated brain anomalies related to the global volume of CSF in the subarachnoid space, particularly over the frontal lobes (Extra-Axial CSF or EA-CSF). An increase in the volume of EA-CSF at 6 months (see example in Figure 3.2) of age was observed in infants who later diagnosed

with ASD (Shen et al., 2013). Further, EA-CSF remained abnormally elevated at 12 and 24 months of age (Shen et al., 2013). Moreover, a greater EA-CSF volume at 6 months was also associated with more severe autism symptoms at the time of diagnosis at 36 months of age (Shen et al., 2013). Such EA-CSF findings were later confirmed through replication in a larger, independent cohort of infants (Shen et al., 2017). Such studies relied on a novel method in infant MRIs to quantify the volume of EA-CSF in the dorsal subarachnoid space above the horizontal plane of the anterior-posterior commissure, thereby avoiding ventral regions that contain cisterns, sinuses, and vasculature that should not be classified as EA-CSF.



(a) (b)

Figure 3.2: (a) T1w image of an infant with a normal MRI at 6 months of age. (b) Similar T1w image of an infant with enlarged EA-CSF at 6 months, who was diagnosed with ASD at 2 years of age. The dark regions between the brain folds and skull indicate increased amounts of EA-CSF.

EA-CSF Quantification: The earlier results indicate that the quantification of global EA-CSF could be important in characterizing the nature of the pathology and its relation to ASD symptoms, particularly its more severe forms. However, the global EA-CSF measure does not provide an anatomical localization of the effect. A more sophisticated, localized EA-CSF extraction would provide measurements suitable for localized group analysis or localized discriminative analysis. Moreover, the ability to obtain anatomically precise measures would allow for an appropriate examination of CSF circulation, leading to better insight on the underlying processes as well as better ability to detect abnormalities in NDDs. One way to extract local measurements of

EA-CSF is through Voxel-based Morphometry (VBM), which utilizes a statistical approach of parametric mapping (Ashburner and Friston, 2000). VBM methods are usually computationally efficient as they do not involve surface reconstruction of complex cortical surfaces. However, the accuracy and precision of localized VBM EA-CSF measurements are limited by the voxel resolution and are sensitive to volumetric registration errors, which are known to be abundantly present in most cortical areas due to the inherent cortical folding variability. Moreover, such voxel-based EA-CSF measurements cannot be easily correlated with other cortical surface-based measurements (e.g., CT and SA) that has been shown to provide early biomarkers for NDDs such as ASD (Hazlett et al., 2017). Hence, the ability to extract high-dimensional surface-based local EA-CSF measurements would allow for a better understanding of how such biomarkers interact, leading to optimal combinations for accurate early prediction of NDDs using deep learning.

Proposed Local EA-CSF: To overcome the limitations mentioned above, as shown in Figure 3.3, this chapter presents a novel framework[1] for extracting surface-based local EA-CSF measurements from sMRI. The proposed framework first computes a probabilistic tissue segmentation of WM, GM, an CSF. A hard segmentation is obtained from the tissue probability maps, which are then used to reconstruct polyhedral models of the outer CSF hull surface as well as WM and GM surfaces. A Laplacian partial differential equation (PDE) is solved between a defined inner surface and the CSF hull surfaces to generate a vector field that is used to create streamlines connecting the surfaces. Along these streamlines, the CSF space is sampled, and CSF probability values are integrated to generate local EA-CSF measures at each point cortical surface. To the best of my knowledge, the proposed framework is the first to address the problem of extracting local EA-CSF measurements in a way that is suitable for localized surface-based analysis. Details of the proposed local EA-CSF extraction pipeline are discussed in the next section.

[1]The work is based on the previously published paper (Mostapha et al., 2018b). This chapter partially adapts text descriptions and figures from the published paper.

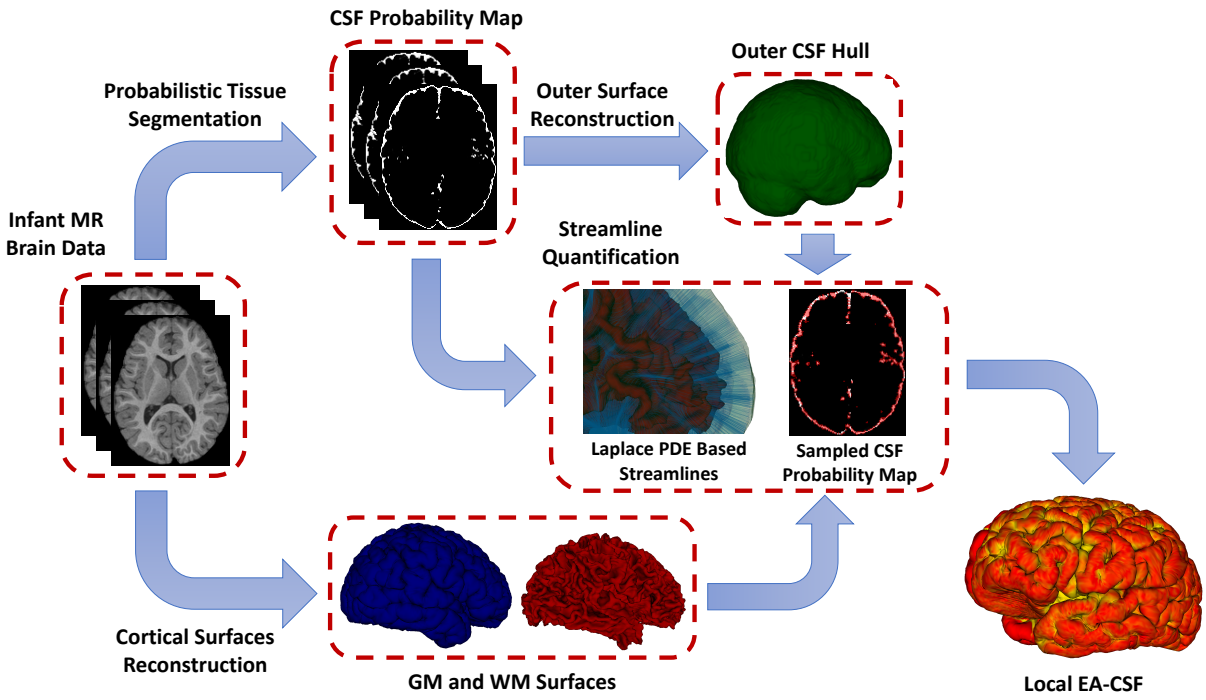


Figure 3.3: The proposed framework for the local EA-CSF from sMRI. The proposed processing pipeline combines probabilistic brain tissue segmentation, cortical surface reconstruction as well as streamline-based quantification to produce accurate and reliable local EA-CSF measurements.

3.2 Methods

3.2.1 Materials

IBIS Dataset: I analyzed 153 structural MR brain images from low-risk infants with no family history of ASD, intellectual disability, or major psychiatric disorder and who had an older sibling with typical development. These infants were scanned longitudinally at 6, 12, and 24 months of age as part of a National Institutes of Health-funded, multi-site, Autism Centers of Excellence (ACE) Network study: the Infant Brain Imaging Study (IBIS). The MRI scans were acquired at 4 different sites (University of North Carolina at Chapel Hill, University of Washington at Seattle, Washington University at Saint Louis and Childrens Hospital of Philadelphia), each equipped with 3T Siemens Tim Trio scanners (Wolff et al., 2012). The scan sessions included T1w (160 sagittal slices with TR=2400ms, TE=3.16ms, flip angle=8°, field of view 224 × 256) and T2w (160 sagittal slices

with TR=3200ms, TE=499ms, flip angle=120°, field of view 256 × 256) MRI scans. All datasets possess the same spatial resolution of $1 \times 1 \times 1$ mm^3. Subjects with scans from all time points were included to enable an accurate longitudinal study of extracted local EA-CSF trajectories. The cohort included 57 scans from female subjects (14.75 ± 7.42 months) and 96 scans from male subjects (14.67 ± 7.49 months); see Table 3.1 for complete demographic information.

	All	Male	Female
Number of sMRI Scans	153	96	57
6 Months (Mean age ± STD)	51 (6.81 ± 0.80)	32 (6.77 ± 0.81)	19 (6.89 ± 0.80)
12 Months (Mean age ± STD)	51 (12.72 ± 0.72)	32 (12.65 ± 0.66)	19 (12.84 ± 0.81)
24 Months (Mean age ± STD)	51 (24.57 ± 0.55)	32 (24.60 ± 0.58)	19 (24.51 ± 0.50)

Table 3.1: Demographic information of typically developing infants included from the IBIS dataset.

Multiple procedures for quality control were employed to assess scanner stability and reliability across sites, times, and procedures. A Lego (Lego Group, Billund, Denmark) brick-based phantom (Fonov et al., 2011) was scanned monthly at each location and analyzed to assess image quality and quantitatively address site-specific local distortions. Also, two adult subjects were scanned once per year at each scanner and after any significant scanner update. The data for these phantoms were evaluated for scanner stability across sites and time (Gouttard et al., 2008) and are also used here to assess stability for the proposed local EA-CSF measure.

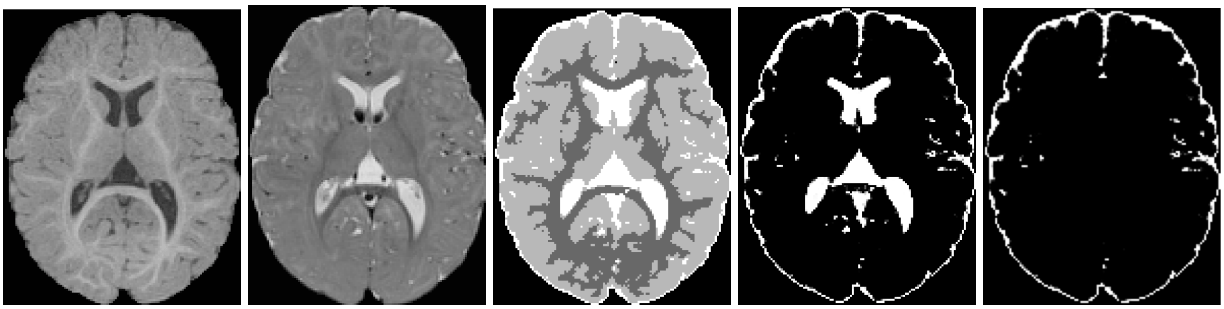
3.2.2 Image Processing and Surface Generation

Initial Preprocessing: The raw T1w and T2w brain images were corrected for intensity non-uniformity using the N4 algorithm (Tustison et al., 2010) (Figure 3.4a,b). Correction of geometric distortions was also applied for the optimal processing of multi-site longitudinal data (Fonov et al., 2010). T1w and T2w images were rigidly transformed to a prior pediatric 1-year-old atlas in

stereotaxic space. A prior intensity growth map was applied to the 12-month T1w and T2w scans to improve the poor contrast of the WM/GM boundary from under-myelination (Kim et al., 2013).

Skull Stripping: The brain mask necessary to perform skull stripping was performed using a multi-atlas approach that combines multiple candidate brain masks obtained via deformable registration of a prior set of atlases (each consisting of a T1w, T2w and brain mask label image). Five brain masks were utilized, namely FSL-BET (Smith, 2002), two in-house prior atlases, and two atlases of the CIVET pipeline (Kim et al., 2005). The deformable registration was computed via the ANTs registration toolkit (Avants et al., 2011) using both T1w and T2w data. The fusion of the candidate brain masks was performed via a straightforward majority vote.

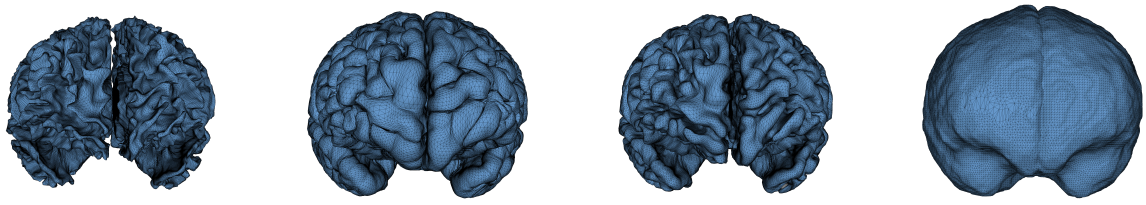
Tissue segmentation: Different brain tissues were then segmented using a framework of atlas-moderated expectation-maximization implemented in the AutoSeg toolkit (Wang et al., 2014). Particularly, a deformable registration was applied to propagate a prior template and prior tissue probability maps for WM, GM, and CSF from MNI space into individual T1w data (Fonov et al., 2011). Then, an Expectation Maximization (EM) based tissue segmentation was performed (Van Leemput et al., 1999) to obtain a label map with segmentations for WM, GM, and CSF (Figure 3.4c). Ventricular CSF space (lateral ventricles, third and fourth ventricles) was then removed from the resulted CSF posterior (Figure 3.4d) by deformably co-registering a single prior template with an existing ventricular area mask and using the registered mask to remove the ventricle (Figure 3.4e).



(a) (b) (c) (d) (e)

Figure 3.4: An example of the probabilistic tissue segmentation obtained for an input infant sMRI scan. (a) T1-weighted scan, (b) T2-weighted scan, (c) WM, GM, and CSF label map, (d) CSF probability map, (e) CSF probability map with ventricular CSF space removed.

Surface Reconstruction: Cortical surfaces were reconstructed with an adapted version of the CIVET workflow (Kim et al., 2005). The cortical surface model consisted of high-resolution triangle meshes (81,920 triangles and 40,962 vertices) in each hemisphere, and cortical surface correspondence among subjects was established via spherical registration to an average surface template (Robbins et al., 2003). CIVET was applied as described in Chapter 2, Section 2.2 to 12 months and 24 months of data following tissue segmentation with AutoSeg to construct WM and GM surfaces (Figure 3.5a,b). However, tissue segmentation for 6-month-old subjects did not yield reliable WM vs. GM segmentation because the WM and GM have almost the same intensity level in both T1w and T2w scans of isointense-phase infants (around 6-8 months of age). Cortical surfaces at 6 months were determined longitudinally for subjects with MRI data at both 6 months and 12 months visit to solve this problem. Using ANTs (Avants et al., 2011) deformable registration with normalized cross-correlation (metric radius 2mm, Gaussian smoothing of 3 mm of the deformation map) of joint T1w and T2w data (both image sources were equally weighted), the pre-processed, brain masked MRI data of 12-month-old subjects was registered to data from the same subject at age 6 months. This registration was applied to the cortical surfaces of the 12-month-old subjects to propagate them into the 6 months space (see Figure 3.6). In addition to the WM and GM surfaces, a smoothed middle surface (Figure 3.5c) was obtained by averaging WM and GM surfaces and then two iterations of averaging based surface smoothing. Moreover, the outer CSF hull surface (Figure 3.5d) was generated by first dilating the intracranial mask, followed by a surface reconstruction using standard marching cubes algorithm (Lorensen and Cline, 1987) and a subsequent Laplacian surface smoothing. Finally, all the reconstructed surfaces were visually QC'ed with a surface cut overlay on the MR images.



(a) WM Surface (b) GM Surface (c) Middle Surface (d) CSF Hull Surface

Figure 3.5: An example of cortical surfaces reconstructed from input infant sMRI.

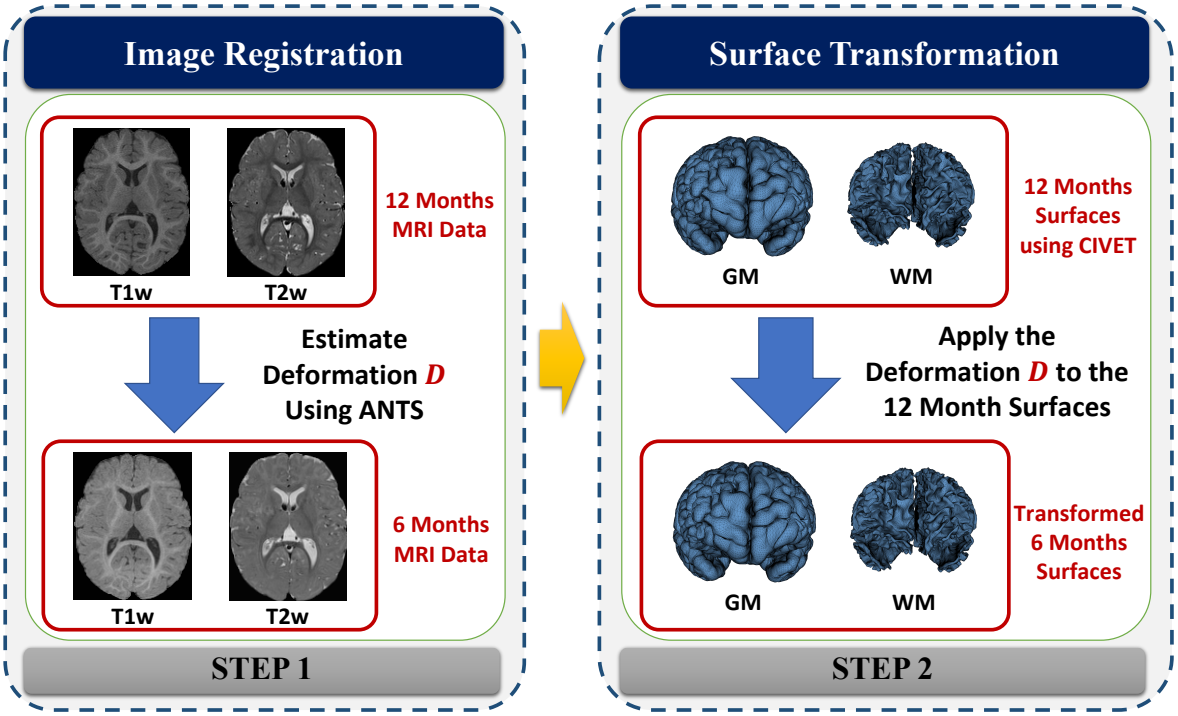


Figure 3.6: The process of generating cortical surfaces for the 6-month-old sMRI scans.

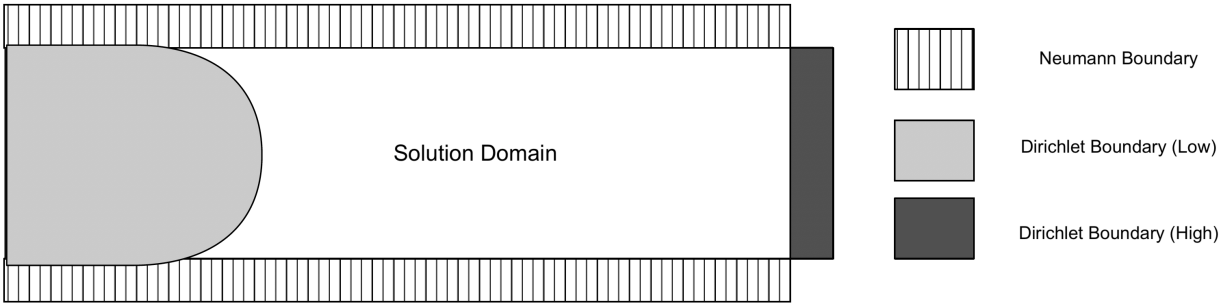
3.2.3 Extraction of Local Extra-Axial Cerebrospinal Fluid

Solving Laplace PDE: Following the reconstruction of the cortical surfaces, the next step is to solve a Laplace's equation between the inner surface (S_{inner}) and a corresponding outer surface (S_{outer}). While the CSF hull surface (Figure 3.5d) is defined as S_{outer}, in this work, I utilize the middle surface (Figure 3.5c) as S_{inner} instead of the GM surface to accommodate for potential surface reconstruction errors. Both S_{inner} and S_{outer} are assumed to have spherical topology, i.e., can be stretched and warped without breaking to form a sphere surface. Laplace's equation is a second-order Partial Differential Equation (PDE) solved for a scalar field $u(x)$ that is enclosed between boundaries S_{inner} and S_{outer}. The Laplace PDE takes the following form

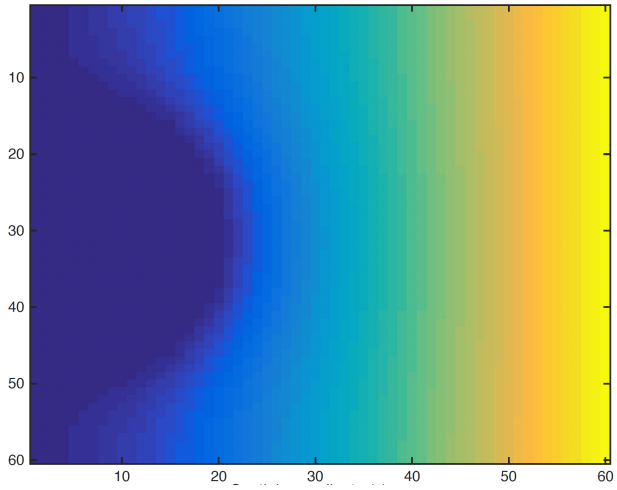
$$\triangle u = \nabla^2 u(x) = 0, \tag{3.1}$$

where $u(x) = u_L$ for $x \in S_{inner}$ and $u(x) = u_H$ for $x \in S_{outer}$. To correctly measure local EA-CSF, a boundary condition map that defines the Laplace PDE boundary condition must be defined in an

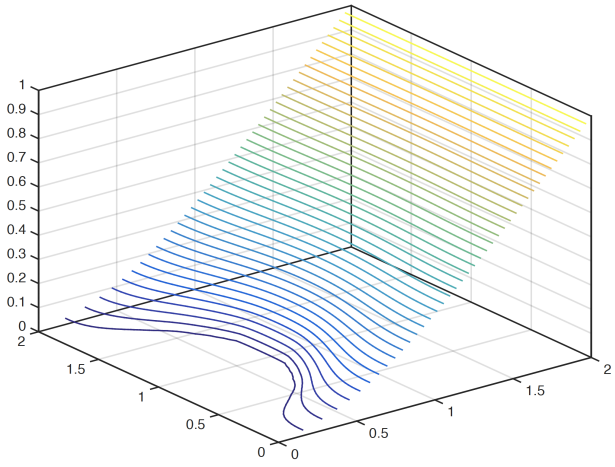
anatomically consistent manner. In the solution of the Laplace PDE in the proposed framework, the solution domain is bounded by the Dirichlet condition and the Neumann condition. The Dirichlet condition specifies values of the solution itself on the boundary, while the Neumann boundary condition defines values for the first-order derivative of the solution. The interface with the Dirichlet condition defines S_{inner} and S_{outer} where streamlines start and arrive, and the Neumann condition defines an open boundary that is parallel to the streamlines (see Figure 3.7). A consistent boundary map generation is ensured using surface-based pre-processing steps that were applied to create a boundary label map in the image domain (Lee et al., 2016). The Laplace PDE is iteratively solved in the created image voxel grid using the Jacobi method (Causon and Mingham, 2010).



(a) Initial Dirichlet and Neumann Boundary Conditions



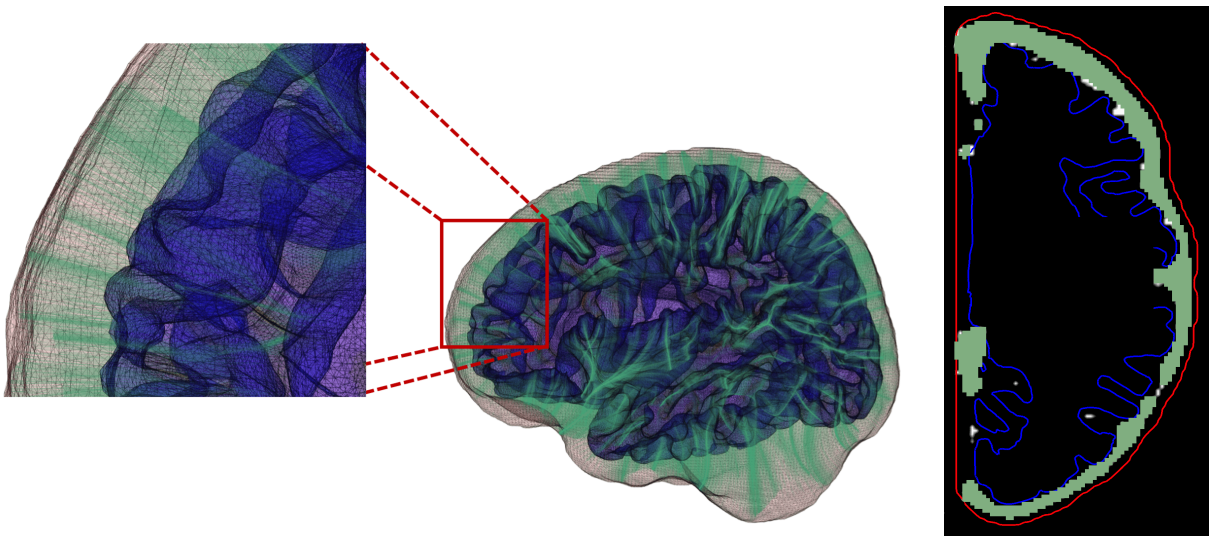
(b) Laplace PDE Solution Color Map



(c) Laplace PDE Solution Contour Image

Figure 3.7: A solution of the (a) Laplace equation with two different boundary conditions. As observed in the color map (b) and the contour image (c), the solution isolines are parallel to the Dirichlet boundary and perpendicular to the Neumann boundary. Figure from (Lee et al., 2016).

Streamline-Based local EA-CSF: After obtaining the Laplace PDE solution, the next step is the computation of the local EA-CSF between S_{inner} and S_{outer}, which I define as CSF accumulated along the lines connecting the two surfaces. Such lines need to be orthogonal to the PDE solution isolines at each point to obtain a biologically plausible path. To achieve that, streamlines that are tangent to the normalized gradient field of the PDE solution are utilized to provide an analogy to cortical columns and to establish a one-to-one correspondence between S_{inner} and S_{outer}. Such streamlines are then constructed explicitly by the integration of the Lagrangian vector field. A fourth-order Runge-Kutta (RK4) integration method (Yaakub and Evans, 1999) is used in generating the streamlines to minimize local truncation error and provide faster convergence.



(a) Computed Streamlines Between Inner and Outer Surfaces (b) CSF Visitation Map

Figure 3.8: (a) Streamlines generated using a fourth-order Runge-Kutta (RK4) integration method. (b) Local EA-CSF measure is computed by accumulating CSF along the generated streamlines.

To achieve sub-voxel accuracy, I process the starting and ending segments of the streamlines to fit them perfectly within the boundary of the defined inner and outer surfaces. Finally, a local EA-CSF measure is computed for each vertex v by accumulating CSF probability P at each point k on the streamline l_v associated with v. A linear approximation is utilized to account for the streamlines non-uniformity

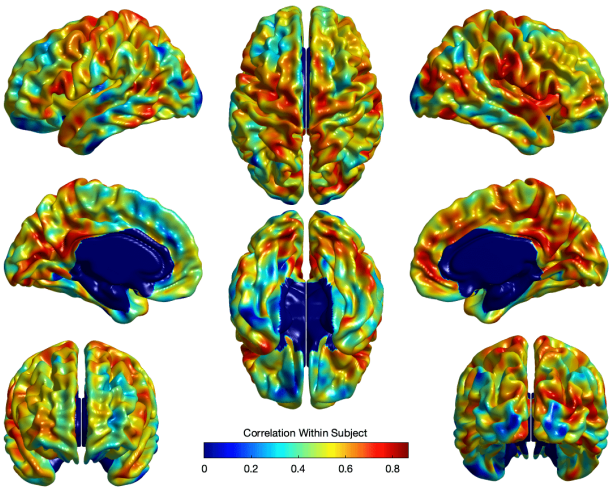
$$\text{EA-CSF}_v = \sum_{k \in l_v} \left(\frac{P(k) + P(k+1)}{2} \right) \times \Delta_k, \tag{3.2}$$

where Δ_k representing the Euclidean distance between point k and the successive point $k+1$. Figure 3.8 provides an example for the computed streamlines and sampled CSF probability map.

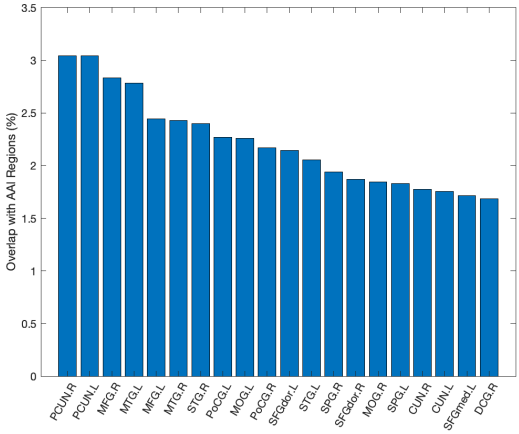
Statistical Analysis: The extracted raw local EA-CSF maps were first mapped to the common MNI surface template space (Lyttelton et al., 2007) for additional processing and analysis. As an initial standard processing step in the study of cortical surface measurements, a geodesic heat kernel-based smoothing (FWHM of 20 mm) was applied to the EA-CSF maps (Chung et al., 2005). The effect of sex and age on local EA-CSF was tested using a longitudinal mixed-effects model in SurfStat, which is a toolbox for statistical analysis of cortical surface measurements applying random field theory for statistical inference (Chung et al., 2010). The longitudinal linear mixed model included a subject-specific random intercept to induce equal correlations between observations on the same subject. Slope were also added to model the fixed effects of sex, age as well as sex and age interactions. In particular, with the local EA-CSF was used as a dependent variable Y, the following linear mixed model was fitted for each subject i:

$$Y_i = \beta_0 + \beta_1 \text{Sex}_i + \beta_2 \text{Age}_i + \beta_3 \text{Sex}_i \text{Age}_i + U_i + \varepsilon_i, \tag{3.3}$$

where U_i capturing estimates for the subject-specific random effect and ε_i is the independent noise term in every observation. The standard False Discovery Rate (FDR) (Benjamini and Hochberg, 1995) correction was applied to correct for the multiple comparisons for the model in Equation 3.3. Figure 3.9 illustrates the within-subject correlation of local EA-CSF across age, as revealed by the linear mixed model. High correlations are shown across most of the brain regions, particularly in the frontal, parietal, and temporal lobes. In the presence of highly correlated areas, subject-specific random effects need to be incorporated in the linear mixed model.



(a) Correlation within Subject



(b) Top 20 Overlapped AAL Regions

Figure 3.9: (a) The correlation of local EA-CSF within-subject across age; high correlation in the frontal, parietal, and temporal lobes was observed. (b) AAL regions (Achard et al., 2006) overlapped with brain regions showing high correlations, ordered by their percentage overlap.

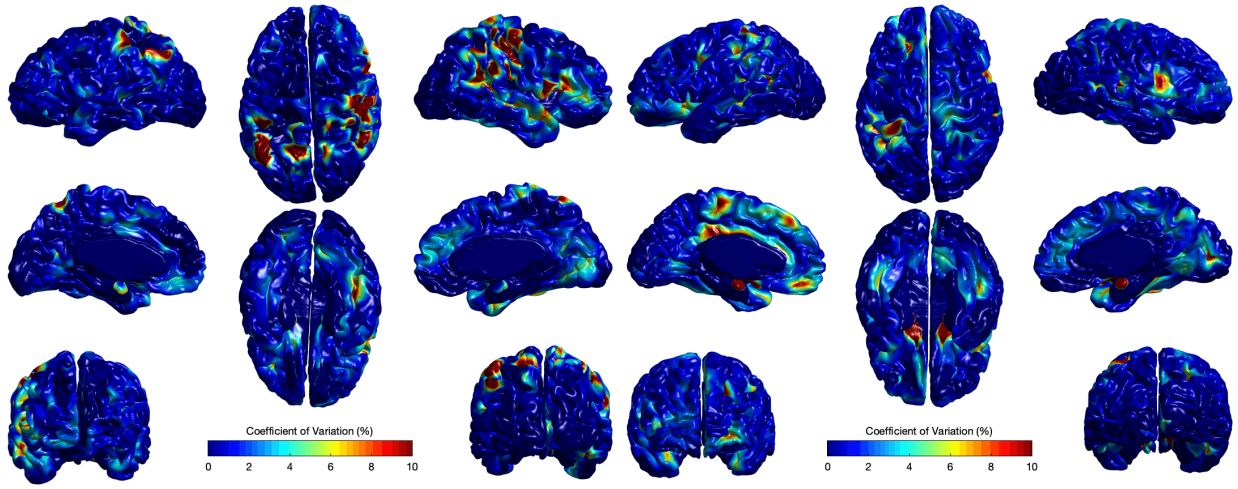
3.3 Experimental Results

Local EA-CSF Reproducibility: The stability and reliability of the proposed local EA-CSF measure were tested using a dataset with a large set of scan/rescan MRIs. Two human phantoms (young male subjects, age 26, and 27 to assess scan consistency), were scanned with the same pulse sequence at four different sites using Siemens 3T Tim Trio scanners, at irregular intervals over 2.5 years. This resulted in 35 MRI scans for subject I and 31 MRI scans for subject II. The tissue segmentation, brain surface reconstruction, and computation of local EA-CSF maps were performed independently. Local EA-CSF maps were then analyzed using the local Coefficient of Variation (CV) as a measure of stability. The CV for a vertex v is defined as the ratio between the standard deviation and mean of the extracted local EA-CSF across diffident scans of the same subject:

$$CV_v = \frac{\sigma_v}{\mu_v} \times 100. \tag{3.4}$$

The CV analysis showed excellent stability with mean across-site CV of 1.15% and 1.56% for all cortical regions in Case I and Case II respectively (Figure 3.10). Higher CVs were observed in few

regions, including left supramarginal gyrus, left postcentral gyrus, left gyrus rectus, right postcentral gyrus, right superior temporal gyrus, and right precentral gyrus. Local EA-CSF variability in these regions was mainly linked to imperfect CSF tissue segmentation. It is worthwhile to mention that the global Intra-cranial Volume (ICV) measures showed CV values around 1% (Bryson et al., 2008; Hazlett et al., 2012) in the same human datasets. Hence, the proposed local EA-CSF extraction framework can provide a stable local measure as compared to global ICV, at least in adult brains.



(a) Case I Coefficient of Variation (b) Case II Coefficient of Variation

Figure 3.10: Coefficients of variation $(\sigma_v/\mu_v \times 100)$ for local EA-CSF maps of two sets of adult brains. Mean coefficient of variation of 1.15% and 1.56% were observed for (a) Case I and (b) Case II respectively. Regions with high coefficient of variation were linked to CSF segmentation issues.

Local EA-CSF over Age: Figure 3.11 shows the mean and standard deviation of local EA-CSF maps on a template middle surface at the 3 ages of 6 months, 12 months, and 24 months. Overall local patterns of mean EA-CSF values at 6, 12, and 24 months are highly correlated ($\rho > 0.98$). Consistent with our previous global EA-CSF report (Shen et al., 2018), we observe local EA-CSF decrease across time, mainly in the frontal lobe, with a larger reduction in local EA-CSF from 6 to 12 months as compared to from 12 to 24 months. Significant inter-subject variability was observed as evident from the mean coefficients of variation of 44.6%, 42.2%, and 42.9% computed at the 3 ages of 6 months, 12 months, and 24 months, respectively. At 6 months, local EA-CSF is more abundant in the central and precentral sulci, since the arachnoid granulations are located just above

these cortical regions, so local EA-CSF may remain transiently elevated there before draining. This observation is consistent with a visual inspection of hundreds of 6 months MR infant brain images. Figure 3.12 shows the linear local EA-CSF change rates per month between 6 and 24 months of age. In regions that showed statistically significant differences, a negative change rate was observed with a mean rate of -0.011 per month. Though a substantial negative slope of EA-CSF was found mostly in the frontal areas, the maximum decrease rate of -0.046 per month was seen at the left superior temporal gyrus. While statistically not significant, the left parahippocampal gyrus showed increasing local EA-CSF at a change rate of 0.053 per month. Figure 3.13 shows the raw T scores for the age effect resulted from fitting the longitudinal mixed-effects model as well as the corrected T scores (FDR threshold $Q < 0.01$). Local EA-CSF showed a highly negative correlation with age in most of the frontal lobe areas, including bilateral middle frontal gyrus and bilateral superior frontal gyrus. Negative correlation with age was also observed in some temporal regions, such as the bilateral superior temporal gyrus. No cortical areas showed a significantly positive association of age and the local EA-CSF.

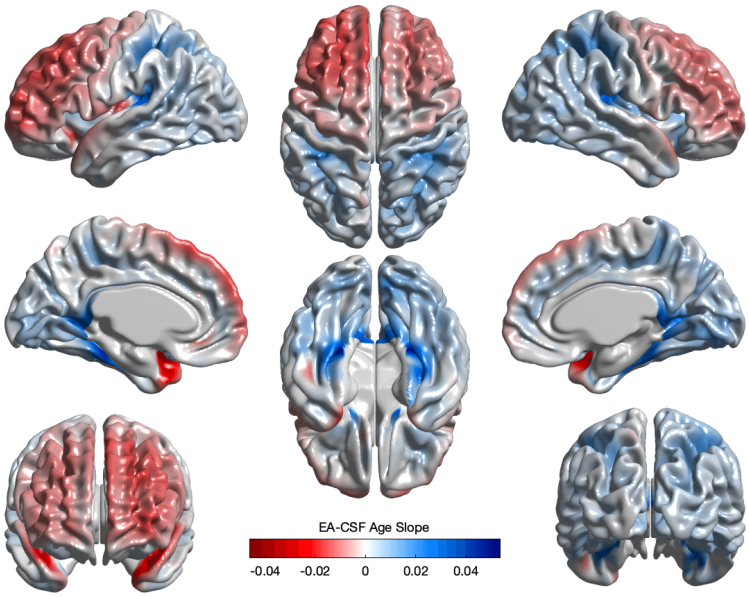


Figure 3.12: The local EA-CSF change rates per month between 6 and 24 months old. Frontal areas showed decreasing local EA-CSF with mean rate -0.011, maximum decrease rate of -0.046 per month was observed at the left superior temporal gyrus, maximum increase rate of 0.053 per month was found at the left parahippocampal gyrus.

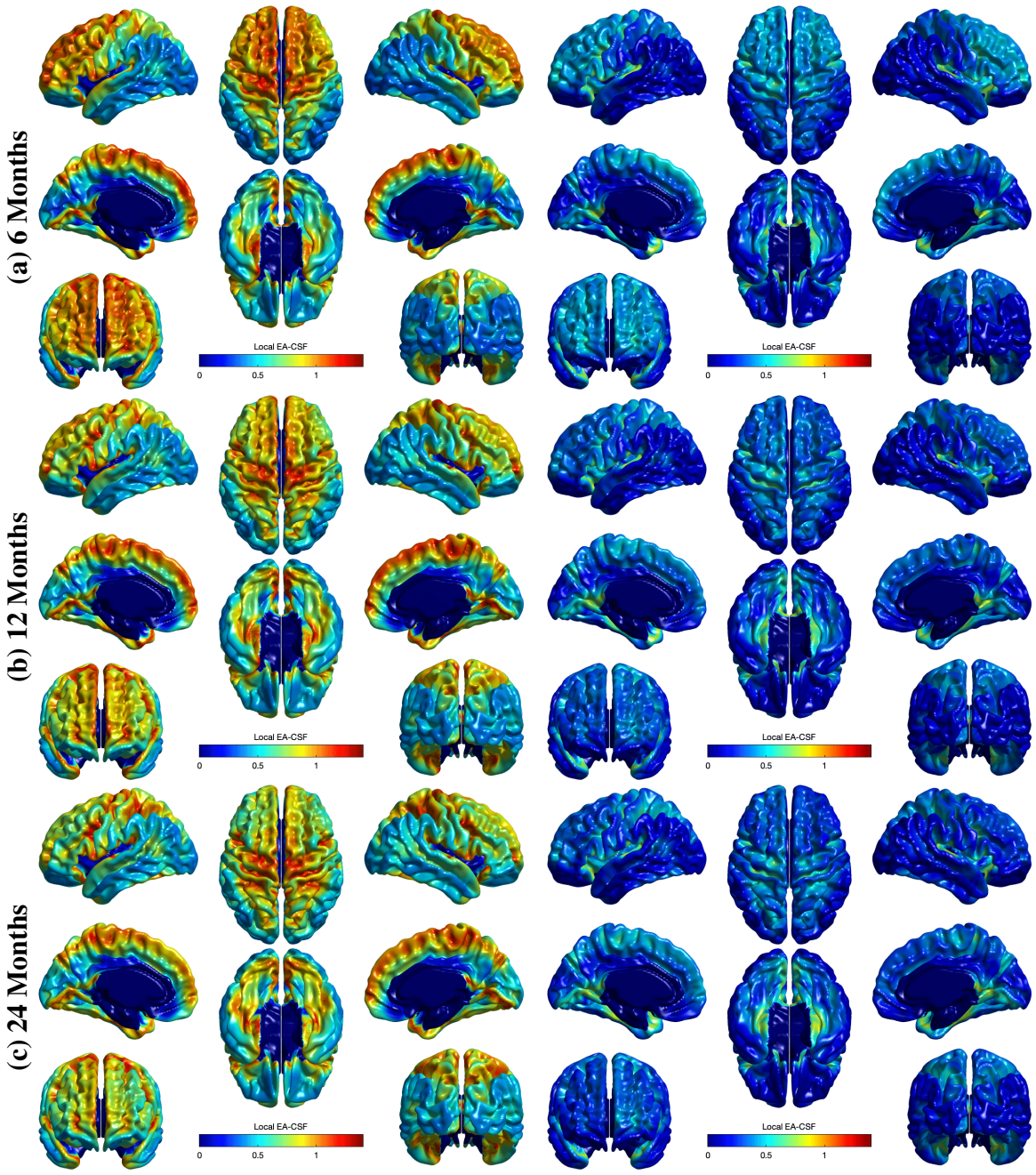
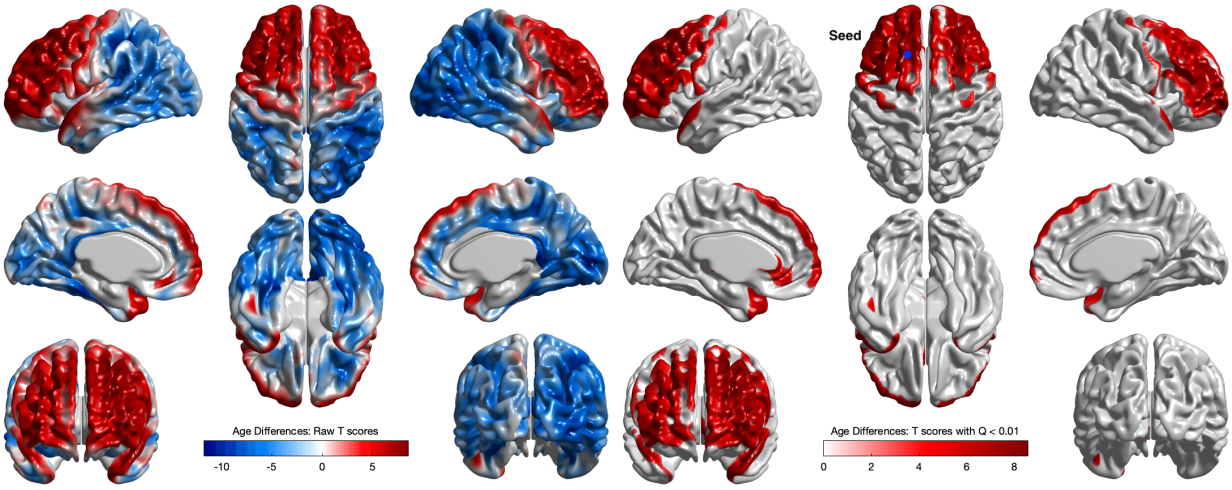
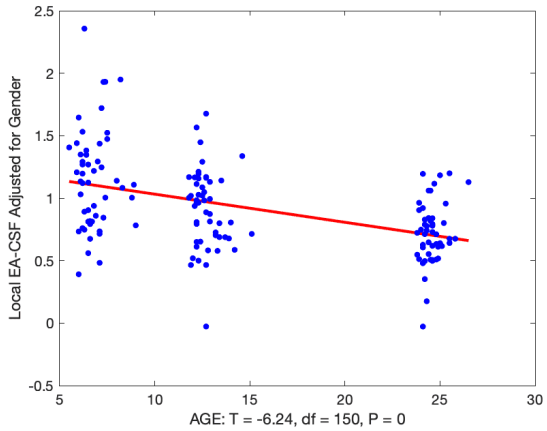


Figure 3.11: The mean and standard deviation of local EA-CSF measured at (a) 6 months, (b) 12 months, and (c) 24 months. Overall the patterns at the 3 ages look quite similar; however, significant inter-subject variability is observed (mean coefficients of variation of 44.6%, 42.2%, and 42.9% computed at 6 months, 12 months, and 24 months, respectively). The highest levels of extracted local EA-CSF are found in the medial and ventral temporal areas.

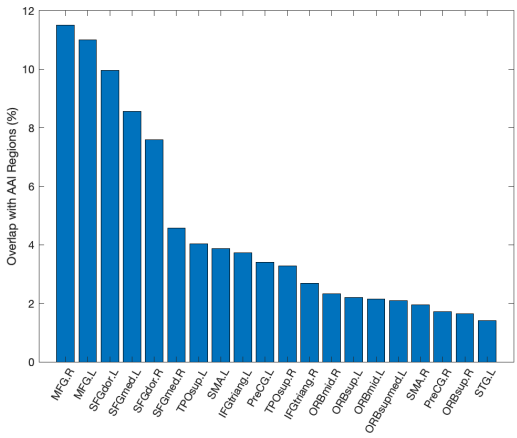


(a) Raw T scores

(b) T scores with FDR $Q < 0.01$

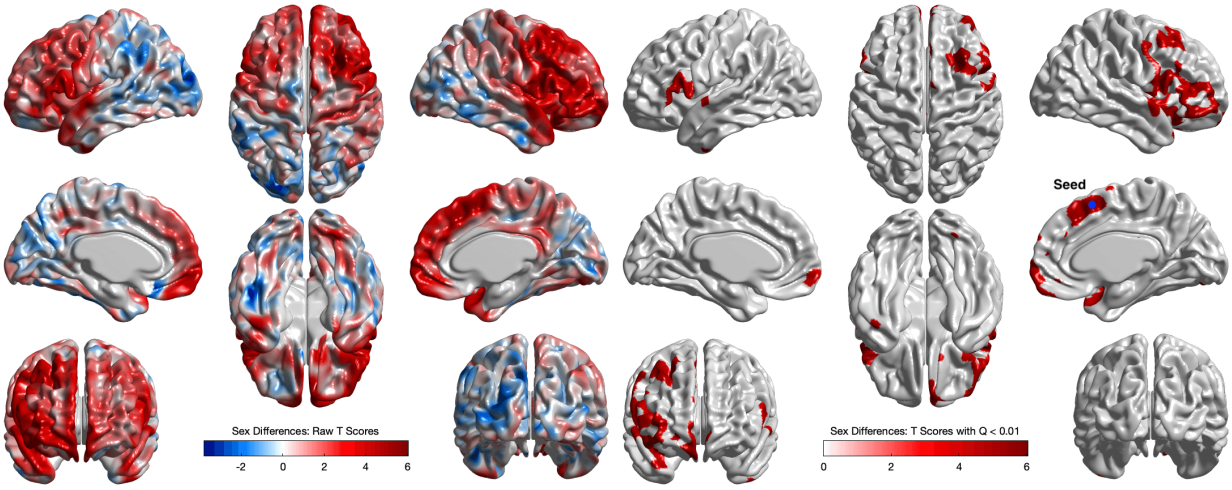


(c) Age Differences at Seed Vertex



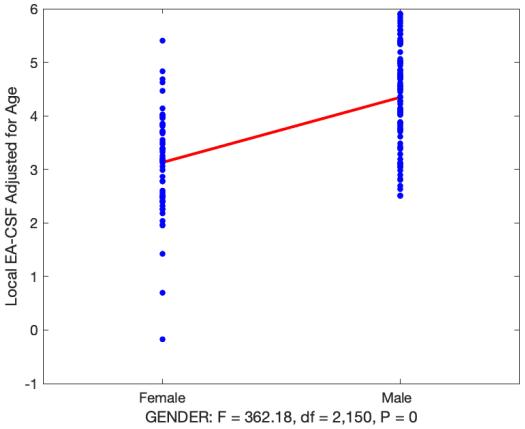
(d) Top 20 Overlapped AAL Regions

Figure 3.13: The longitudinal result of age effect on local EA-CSF. (a) Raw T-scores before multiple comparison correction (red: negative age effect; blue: positive age effect). (b) T-scores for brain regions with statistically significant (FDR $Q < 0.01$) age effect; 16.9% of the brain showed a decrease in local EA-CSF in the first two years of infant's life. (c) Age differences at a seed vertex that correspond to the maximum T score, ($T_{max} = 8.57$). (d) AAL regions overlapped with brain regions showing statistically significant age effects, ordered by their percentage overlap.

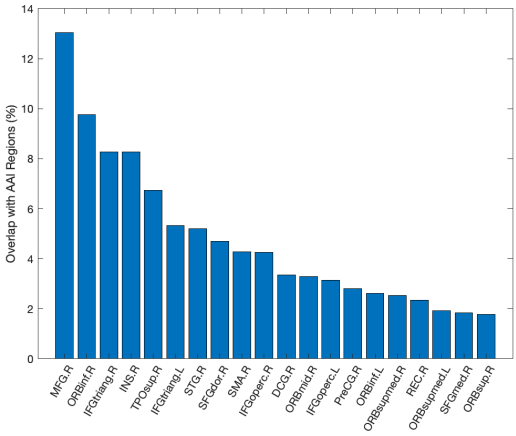


(a) Raw T scores

(b) T scores with FDR $Q < 0.01$



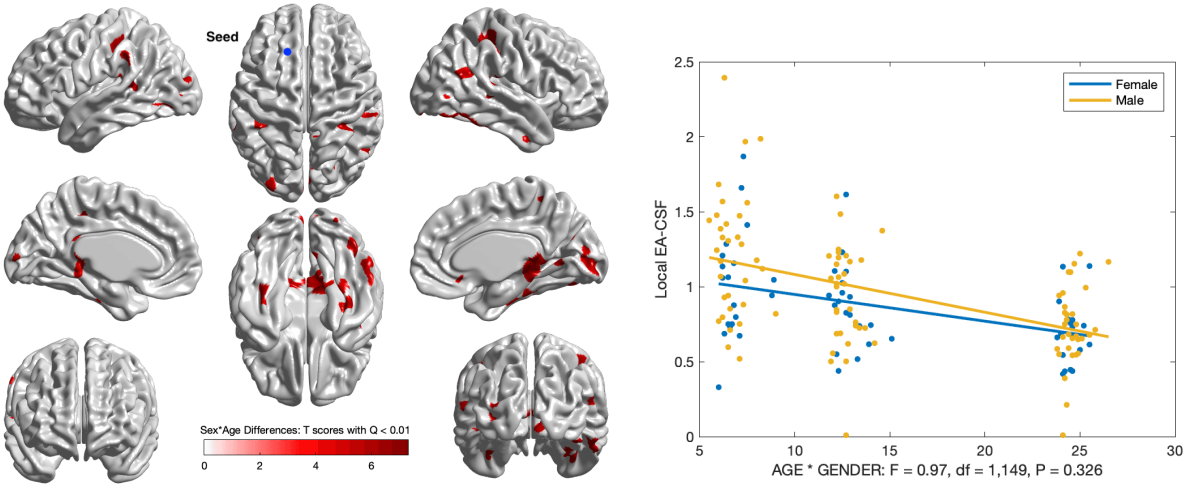
(c) Sex Differences at Seed Vertex



(d) Top 20 Overlapped AAL Regions

Figure 3.14: The longitudinal result of the sexual dimorphism effect on local EA-CSF. (a) Raw T-scores before applying multiple comparison correction (red: male > female; blue: male < female). (b) T-scores for brain regions with statistically significant (FDR $Q < 0.01$) sex differences; 6.2% of the brain showed more local EA-CSF in males compared to females. (c) Sex differences at a seed vertex that correspond to the maximum T score, ($T_{max} = 6.03$). (d) AAL regions overlapped with brain regions showing statistically significant sex differences, ordered by their percentage overlap.

Sexual dimorphism of Local EA-CSF: Cross-sectional analysis conducted at each of the 3 time points separately indicated no cortical regions with statistically significant (FDR threshold $Q < 0.01$) sex differences. On the other hand, the longitudinal analysis of sex (see Figure 3.14), revealed several brain regions with statistically significant (FDR threshold $Q < 0.01$) higher local EA-CSF in males compared to females. Right middle frontal gyrus, bilateral inferior frontal gyrus, right insula, and right superior temporal gyrus are examples of such regions with significant longitudinal sex differences. Figure 3.15 shows the results of the longitudinal analysis of sex and age interaction, where only a few regions showed statistically significant differences. The longitudinal negative change of local EA-CSF in male subjects is higher than females in areas such as bilateral calcarine fissure and bilateral parahippocampal gyrus. In the regions that show significant (FDR threshold $Q < 0.01$) age-related local EA-CSF differences, I observed a similar age slope for both males and females.



(a) T scores with FDR $Q < 0.01$

(b) Sex-Age Interaction at Seed Vertex

Figure 3.15: The longitudinal result of sex and age interaction effect on local EA-CSF. (a) Few regions showed statistically significant differences (FDR $Q < 0.01$). (b) Local EA-CSF profile at vertex showing peak age effect; little difference in slope between males and females was observed.

3.4 Summary

In this chapter I proposed a novel framework for extracting surface-based local EA-CSF measurements from MR brain images. The proposed framework is the first to address the problem

of obtaining local EA-CSF measurements in a way that is suitable for localized surface-based analysis. The proposed processing relies on a probabilistic tissue segmentation approach to generate a CSF probability map that is used to reconstruct the outer CSF hull surface. A Laplacian partial differential equation is solved between the inner cortical surface and the CSF hull surfaces to generate a vector field that is used to create streamlines connecting the surfaces at sub-voxel accuracy via a fourth-order Runge-Kutta approach. The starting and ending segments of the streamlines are then processed to fit them correctly within the boundary of the defined inner and outer surfaces. Along these streamlines, the CSF probability values are accumulated to quantify EA-CSF measures at each vertex on the cortical surface mesh. The proposed local EA-CSF extraction tool was used to study the early postnatal brain development in typically developing infants from the IBIS dataset. Due to a high within-subject correlation, a longitudinal linear mixed model was proposed to incorporate fixed effects of age and sex as well as sex and age interactions.

The experimental results obtained from the scan/rescan human dataset have shown that the proposed local EA-CSF measure is reliable in a scan/rescan setting and produces reasonably stable results. The stability of the proposed method was further confirmed by the consistency in local EA-CSF patterns across the 3-time points used in studying local EA-CSF trajectories in the first two years of infancy. The experimental results revealed several findings through my proposed processing and analysis pipeline. First, local EA-CSF in several cortical regions, mainly in the frontal lobe areas, showed a statistically significant negative correlation with age. The longitudinal analysis also revealed several cortical regions with statistically significant higher local EA-CSF in males compared to females. Most of these regions also showed a more substantial decrease in local EA-CSF across age. However, few cortical areas showed higher negative local EA-CSF change rate in male subjects compared to females. Such localized findings confirmed that the proposed local EA-CSF extraction pipeline could reveal specific regions of significant change that would not be possible to be observed using the previous global EA-CSF approach. Moreover, in Chapter 4, I demonstrate that a combination of local EA-CSF and cortical shape measures, extracted from a 6-month-old sMRI scan, can improve the accuracy of predicting ASD diagnosis at 24 months.

CHAPTER 4: EARLY PREDICTION OF AUTISM SPECTRUM DISORDER

4.1 Introduction

Motivation: Despite tremendous efforts by parents, researchers, clinicians, and educators, ASD continues to be a lifelong condition that imposes challenges to affected individuals and their families. Behavioral, educational, and pharmacologic interventions have achieved only moderate success, at best, in ameliorating associated and core features of this disorder (Rogers et al., 2014; Green et al., 2015; Jones et al., 2017). Treatment does not begin until after diagnosis, which does not typically occur until two years of age and beyond (Baio, 2012), as behavioral features under 24 months of age have not proven sufficiently accurate for clinically-useful diagnostic prediction (Rogers, 2009; Ozonoff et al., 2010; Landa et al., 2013; Zwaigenbaum et al., 2005). Research on degenerative neurological disorders (e.g., Parkinson disease) and neurodevelopmental disorders (e.g., ASD) has demonstrated that brain changes precede clinical manifestations (Aylward et al., 2012; Kandiah et al., 2014; Paulsen et al., 2010). In addition to the general understanding in medicine that earlier (and particularly presymptomatic) treatment is better than later treatment, recent studies confirm that for ASD (Vivanti et al., 2016). However, the absence of an accurate presymptomatic biomarker for autism has precluded efficient studies from examining the efficacy of the presymptomatic intervention. Therefore, there is a strong rationale for identifying early, presymptomatic markers of ASD to accurately identify those children at highest risk who may benefit most from the presymptomatic intervention for ASD. Next, I continue the discussion in Chapter 1 on the current findings on the early prediction of ASD from neuroimaging MRI studies.

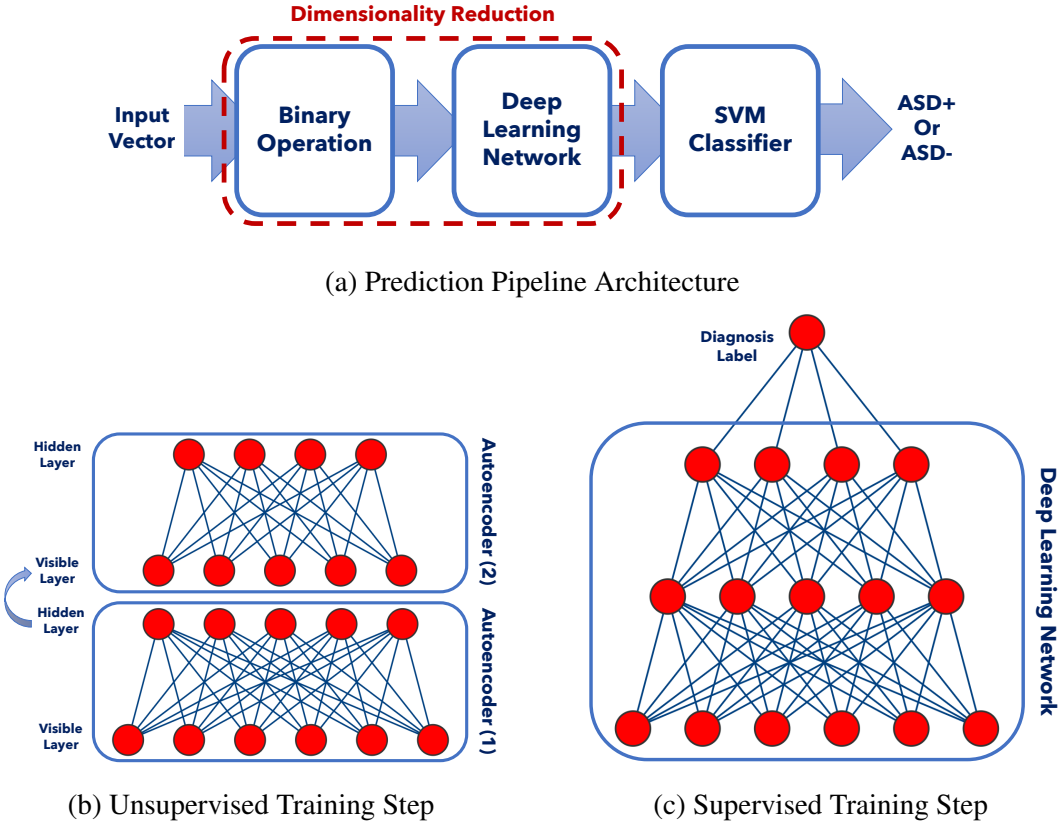
Group Correlation in ASD Early Detection: Efforts to identify early markers of ASD have benefited from the prospective study of younger siblings of children with ASD, who are at 15- to 20-fold High-familial Risk (HR) of developing ASD than the general population (Ozonoff et al.,

2011). It was demonstrated that infants who develop ASD show a variety of age-specific brain differences between 6-24 months, and demonstrating early brain changes that correlate with 24 month ASD outcome and related behaviors (Wolff et al., 2012; Elison et al., 2013; Lewis et al., 2014; Wolff et al., 2015). Although significant, these group-level differences and brain-behavior correlations do not allow for individual-level outcome prediction critical for application to clinical practice. Many studies have used machine learning methods with MRI data to classify individuals with autism (Mostapha et al., 2015; Chen et al., 2016; Yahata et al., 2016) or to understand their patterns of behavior (Lombardo et al., 2015; Plitt et al., 2015). However, few studies prospectively classify infants before the onset of the core features of the disorder.

Individual Prediction in ASD Early Detection: The IBIS Network recently reported that sMRI (predominantly based on SA measurements), at 6 and 12 months of age, accurately predicted ASD diagnosis at 24 months (PPVs of $\sim 80\%$) in HR infants (Hazlett et al., 2017). Since the work in (Hazlett et al., 2017) provides a comparable baseline to the current work (a deep learning-based model using cortical measures extracted from sMRI), I provide next a summary of the utilized methods. In such a prediction framework, the dimensionality of the cortical measurements was first summarized using the AAL parcellation atlas (Achard et al., 2006). A deep learning model was then used as a second dimensionality reduction (Figure 4.1a). The input feature vector was first binarized using a trained binary masking operation to allow performing an initial, unsupervised autoencoder training procedure on individual two-layer greedy networks (Lee et al., 2009) (Figure 4.1b). A supervised training step was then applied to the autoencoder via a single classification layer with a single output representing the binary diagnosis label (Figure 4.1c). Upon training completion, the last layer is removed, and the number of nodes in the last hidden layer represents the final dimension of the reduced dimension output. Finally, the generated reduced representations, along with the binary training labels, are then used to train a two-class Support Vector Machine (SVM) classifier.

Subsequently it was also reported that functional connectivity MRI at 6 months could predict ASD diagnosis in a small sample of eleven HR infants who developed ASD (Emerson et al., 2017). Finally, in two separate samples, an excessive total volume of EA-CSF at 6 months of age was found

in HR infants who developed ASD two years later (Shen et al., 2013, 2017). Collectively, these previous studies demonstrated that presymptomatic brain markers of ASD are present in the first year of life before the full manifestation of the disorder, suggesting that a combination of cortical surface anatomy and CSF volume may be useful in predicting later ASD diagnosis.



(a) Prediction Pipeline Architecture

(b) Unsupervised Training Step

(c) Supervised Training Step

Figure 4.1: The prediction pipeline proposed in (Hazlett et al., 2017) for early prediction of ASD.

Current Limitations: The ASD prediction approaches taken to date suffer from several notable limitations. In particular, the previous report using sMRI (Hazlett et al., 2017) required predictors at both 6 and 12 months of age. Moreover, its multi-stage design (i.e., not end-to-end learning) may suffer from group-contamination in the cross-validation analysis, leading to optimistic prediction results. On the other hand, ASD prediction based on EA-CSF was only performed using a global measure that lacked sufficient accuracy (Sensitivity (SEN) of 66% and Specificity (SPC) of 68%) to serve as a single predictor of later diagnosis (Shen et al., 2017). Finally, the prediction based on functional connectivity MRI was conducted on a small sample and required acquisition and

processing methods that may be too complex to implement for widespread clinical applications. Therefore, the objective of this work is to combine multiple measures of brain anatomy from a conventional sMRI scan at 6 months to improve the accuracy of predicting later ASD diagnosis.

Proposed Deep-ASD Framework: I propose a novel framework to generate multiple measures of infant brain morphology across the brain, intending to find the best combination of brain measures at 6 months that optimized prediction of later ASD diagnosis at 24 months (see Figure 4.2).

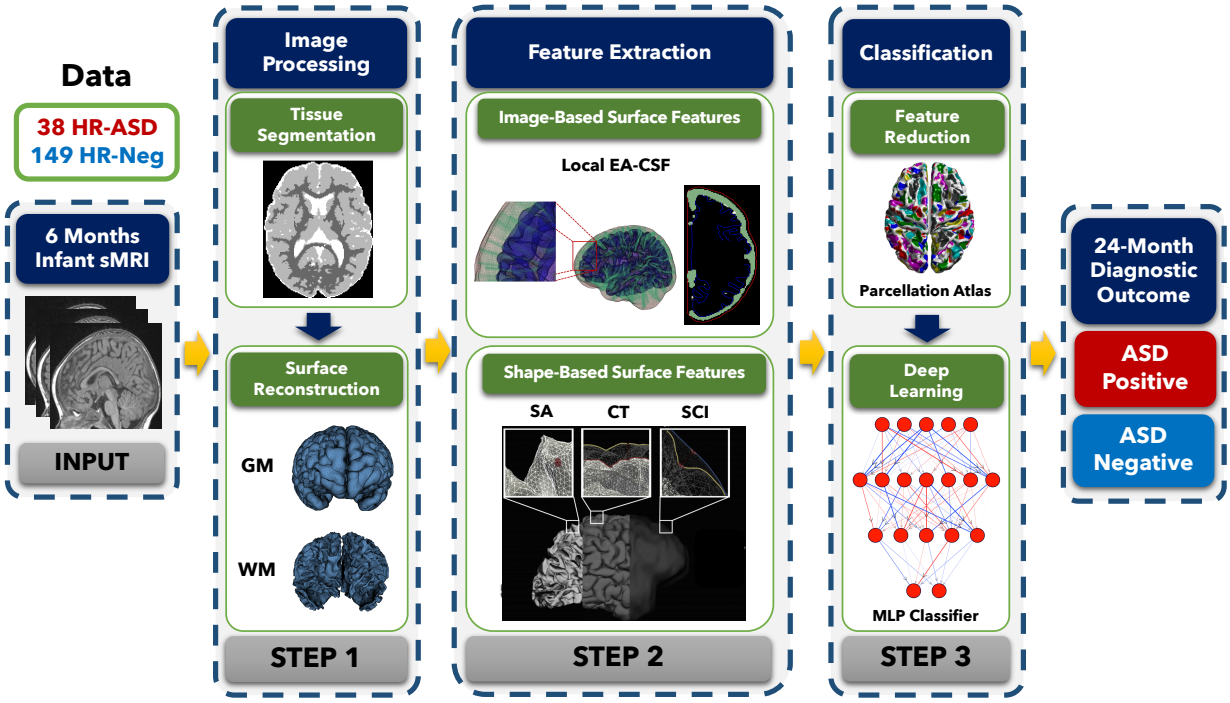


Figure 4.2: Block diagram of the proposed prediction system for early diagnosis of ASD.

The input to the proposed framework is 3D T1w and T2w sMRI scans of the 6 month-year-old infant's brain. The proposed prediction system consists of three main steps. The first step of the proposed prediction system is to accurately segment different brain tissues, namely WM, GM, and CSF from the input infant brain data. These segmented images will be used to reconstruct 3D models of the brain WM and GM surfaces. Following brain reconstruction, the next step is to extract discriminatory features, which are numerical values that are corresponding to the attributes of the extracted cortical surfaces. In addition to the CT and SA that have been previously investigated in the

early prediction of ASD (Hazlett et al., 2017), two additional features were extracted. In particular, the local cortical shape is quantified via the local SCI (see Chapter 2, Section 2.2) and local EA-CSF volume is measured using the novel quantitative approach presented in Chapter 3. Several anatomically and functionally based cortical parcellation atlases are investigated in summarizing the extracted high-dimensional cortical features. Finally, the diagnostic capabilities of the reduced features are examined separately and combined using deep learning techniques with particular attention paid to solving the class-imbalance problem during the training phase.

4.2 Methods

4.2.1 Materials

Participants: The current study was conducted as part of IBIS Network, a multi-site longitudinal study of infants at high-familial risk for ASD. Infants who have an older sibling with ASD are at 15-20 times greater risk for developing ASD themselves, compared to the general population (Ozonoff et al., 2011). Thus, the prospective longitudinal study of these 'high-risk' infants from 6 months of age enables the possibility of identifying early risk markers that precede the diagnosis of ASD (Ozonoff et al., 2011). In the current study, infants at high and low familial risk for ASD (HR and LR, respectively) were enrolled at 4 clinical sites in the United States: University of North Carolina at Chapel Hill, University of Washington, Washington University in St. Louis, and Children's Hospital of Philadelphia (Wolff et al., 2012). HR infants had an older sibling with a clinical diagnosis of ASD, corroborated by the Autism Diagnostic Interview-Revised (ADI-R) (Lord et al., 1994). LR infants had a typically developing older sibling and no first- or second-degree relatives with intellectual or developmental disabilities (Estes et al., 2015). Details on recruitment and exclusionary criteria are reported in a previous publication on this sample (Hazlett et al., 2012). Data collection sites had study protocols approval from their Institutional Review Boards, and all enrolled subjects had informed consent provided by their parent/guardian.

Behavioral and Diagnostic Assessments: Infants underwent an MRI scan at 6 months (see details below) and at 24 months were classified as either positive or negative for ASD, based

on expert clinical judgment based on all available clinical data that included: (a) DSM-IV-TR criteria (Association, 2000); (b) Autism Diagnostic Observation Schedule-G (ADOS) (Lord et al., 2000); and (c) ADI-R (Lord et al., 1994). None of the LR infants in the current study turned out to be positive of ASD based on the same outcome criteria used for HR infants. Further details on the assessment and diagnostic procedures can be found in a previous publication (Estes et al., 2015). This study design yielded three diagnostic outcome groups: *(1)* HR infants who were positive for ASD at 24 months [HR-ASD]; *(2)* HR infants who were negative for ASD at 24 months [HR-Neg]; and *(3)* LR infants who were negative for ASD at 24 months [LR-Neg] (see Table 4.1 for complete demographic information). Importantly, the HR cohort in the current study was a naturalistic and representative sample of what is expected in the HR population, exhibiting the same ratio of ASD positive vs. negative outcomes (4:1) that have been consistently reported in the literature.

	HR-ASD	HR-Neg	LR-Neg
Number of Participants	38	149	102
Sex Information	33 M; 5 F	79 M; 70 F	61 M; 41 F
Age at MRI (Months)	6.7 ± 0.8	6.6 ± 0.7	6.7 ± 0.6
Age at Diagnosis (Months)	25 ± 1.9	24.8 ± 1.2	24.7 ± 1.1
Mullen Early Learning Composite score at Diagnosis	78.1 ± 16	103 ± 15.6	111.5 ± 17
ADOS severity score at Diagnosis	6 ± 1.9	1.6 ± 0.9	1.3 ± 0.7

Table 4.1: Demographic information of LR and HR infants included from the IBIS dataset.

MRI Acquisition: Infants were scanned during natural sleep at 6 months of age on a 3T Siemens Tim Trio with a 12-channel head coil. The scan sessions included T1w (160 sagittal slices with TR=2400ms, TE=3.16ms, flip angle=8°, field of view 224×256) and T2w (160 sagittal slices with TR=3200ms, TE=499ms, flip angle=120°, field of view 256×256) MRI scans. All datasets possess the same spatial resolution of $1 \times 1 \times 1$ mm^3. Description of the MRI acquisition parameters, image quality control, and cross-site reliability are detailed in (Hazlett et al., 2012).

4.2.2 Image Processing and Surface Generation

Initial Preprocessing Steps: The following steps were applied: *(1)* Rigid transformation of both T1w and T2w image to a prior pediatric 1-year-old atlas in stereotaxic space (Kim et al., 2013). *(3)* Correction of intensity inhomogeneity via the N4 algorithm (Sled et al., 1998). *(3)* Correction of geometric distortions for optimal processing of multi-site longitudinal data (Fonov et al., 2010).

Skull Stripping: The brain mask necessary to perform skull stripping was performed using a multi-atlas approach that combines multiple candidate brain masks obtained via deformable registration of a prior set of atlases (each consisting of a T1w, T2w and brain mask label image). For stability, each of the five brain masks results was fused, namely, FSL-BET (Smith, 2002), two in-house prior atlases, and two atlases of the CIVET pipeline (Kim et al., 2005). The deformable registration was computed via the ANTs registration toolkit (Avants et al., 2011) using both T1w and T2w data. The fusion of the candidate brain masks was performed via a majority vote.

Tissue Segmentation: Because the MRI contrast between WM and GM is significantly lower at 6 months (isointense-phase) than at older ages, typical segmentation pipelines (Wang et al., 2014) would not yield reliable WM and GM segmentations. Such a problem was previously solved by leveraging the availability of an individual's 12-month scan; applying its tissue segmentation to their 6-month scan (Hazlett et al., 2017). While this approach allowed performing WM and GM segmentation of 6-month neuroanatomy accurately, there is an obvious limitation for clinical use in that it required a scan at 12 months of age. The current work avoids such a limitation by computing tissue segmentation for WM, GM, and CSF using a deep learning approach (Zeng and Zheng, 2018) that relied on a context guided, multi-stream, two-stage 3D FCN architecture as shown in Figure 4.3. The first stage FCN-1 is used to learn tissue-specific probability maps from the input T1w and T2w images. After obtaining an initial segmentation, a distance map is computed for each brain tissue to model the spatial context information. The second stage, FCN-2, is used to obtain the final segmentation based on the computed spatial contact information and the original multi-modality MR brain images. All 6-months segmentation maps were then visually QC'ed by an MR expert while overlaid on the MR images (Hazlett et al., 2017). The ICVs were then computed by summing the

volumes of the segmented WM, GM, and CSF brain tissues. In addition to the generated CSF hard segmentation, a CSF probability map was generated by propagating a prior template and prior CSF probability map from MNI space into individual T1w data using a deformable registration (Fonov et al., 2011) implemented in the AutoSeg toolkit (Wang et al., 2014). The ventricular CSF space was then removed by deformably registering a single prior template with an existing ventricular mask and using the registered mask to remove the ventricles from the CSF probability map.

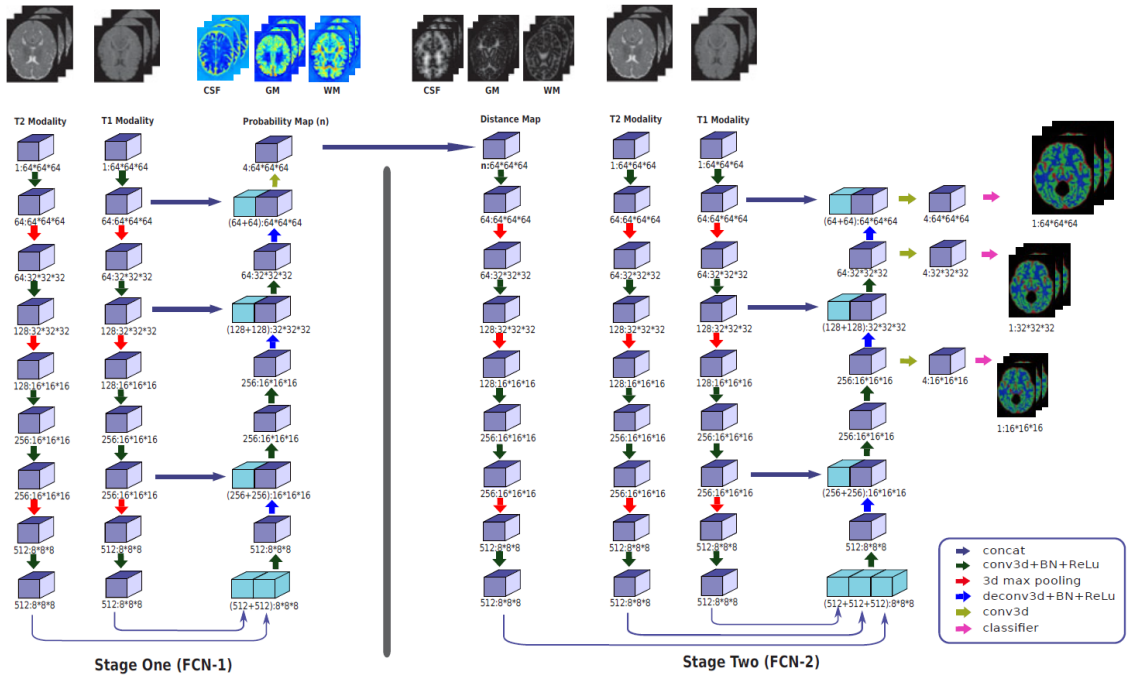


Figure 4.3: Illustration of the two-stage 3D FCN architecture that was adapted for segmenting 6-month-old structural MRI scans in the IBIS dataset. The first stage FCN will produce an initial segmentation that in turn is used to model the spatial context map for each brain tissue using distance maps. The final segmentation is obtained in the second stage using both the spatial contact information and the original T1w and T2w images. Figure adapted from (Zeng and Zheng, 2018).

Surface Reconstruction: Cortical surfaces were reconstructed with an adapted version of the CIVET workflow (Kim et al., 2005). The cortical surface model consisted of high-resolution triangle meshes (163,840 triangles and 81,924 vertices) in each hemisphere, and cortical surface correspondence among subjects was established via spherical registration to an average surface template (Robbins et al., 2003). CIVET was applied as described in Chapter 2, Section 2.2 to 6-months data following the tissue segmentation step described above to construct WM and GM

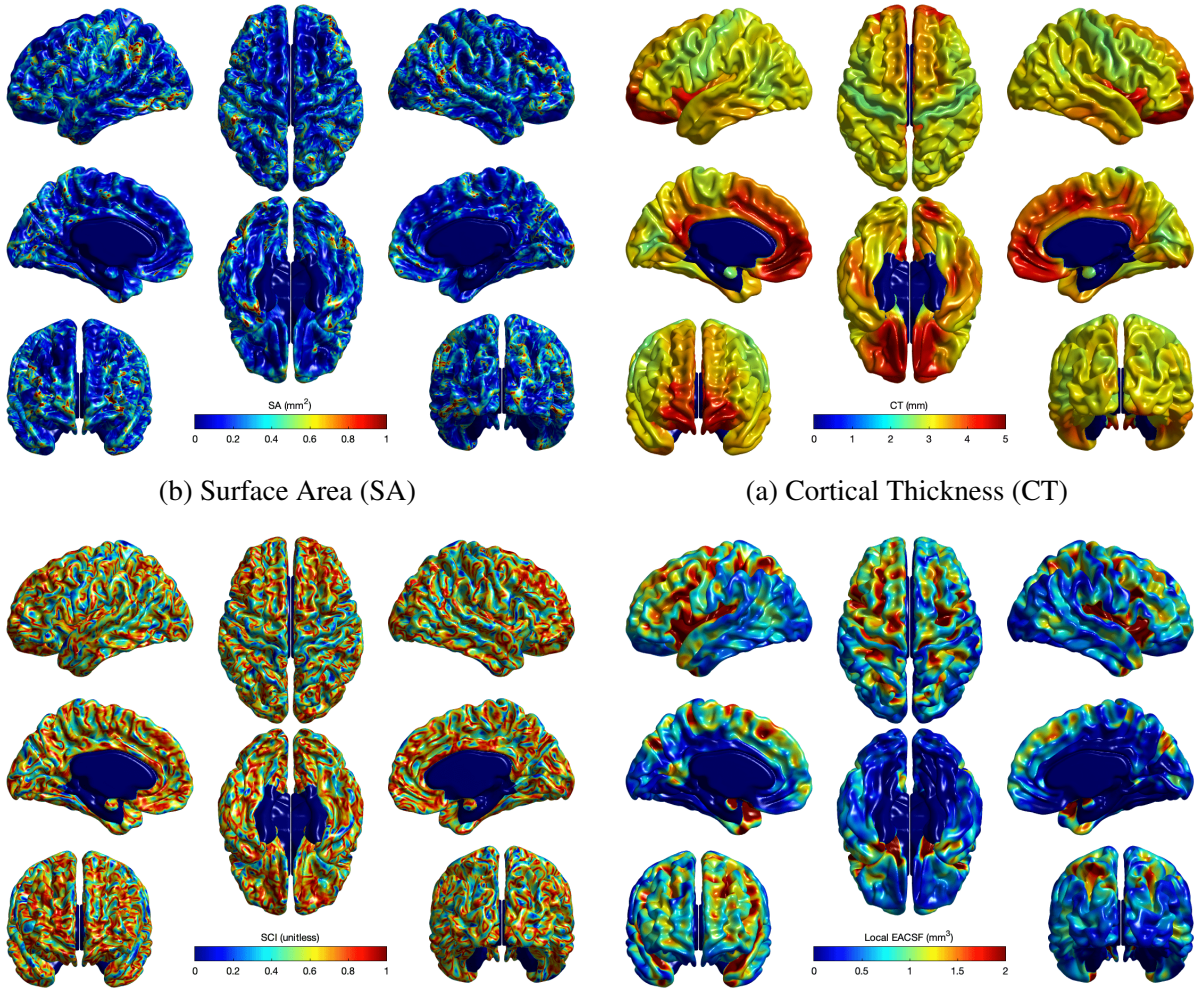
surfaces. In addition to the WM and GM surfaces, a smoothed middle surface was obtained by averaging WM and GM surfaces followed by two iterations of averaging based surface smoothing. Moreover, the outer CSF hull surface was generated by first dilating the intracranial mask followed by a surface reconstruction using standard marching cubes algorithm (Lorensen and Cline, 1987) and a subsequent Laplacian surface smoothing. Finally, all the reconstructed surfaces were visually QC'ed by an MR expert with a surface-cut overlay on the MR images (Hazlett et al., 2017).

4.2.3 Feature Extraction

Generating Cortical Measurements: Predictive anatomical measures were generated from the 6-month scan. Based on the previously published results (Hazlett et al., 2017), SA and CT measures were extracted first from the 6-month surfaces using the CIVET workflow (Shaw et al., 2008, 2012) (see Chapter 2, Section 2.2). SA was computed at the middle surface, i.e., the surface that runs at the mid-distance between WM and outer GM surfaces to avoid over or under-representation of gyri or sulci (Van Essen, 2005). CT was calculated in the native space of each subject by solving a Laplace's equation in the GM region, where the CT was defined as the length of the streamline obtained by tracing the gradient of the resultant vector field (Kim et al., 2005). The CT maps were then smoothed using a 20 mm bandwidth geodesic heat kernel (Chung et al., 2005).

In addition to SA and CT measures, I also quantified the local cortical shape using the SCI (see Chapter 2, Section 2.2). SCI is a measure of the local shape variance that was computed as the difference between the observed distributions of local surface shape index and its best-fit basic shape index model within a limited neighborhood (radius 3 mm) (Kim et al., 2016b). Such a distribution difference was estimated via the Earth Mover Distance. Furthermore, I generated local measures of EA-CSF via the novel framework presented in Chapter 3, which combines probabilistic brain tissue segmentation, cortical surface reconstruction, and streamline-based regional EA-CSF quantification. For streamline computation, I employed the vector field generated by solving a Laplacian PDE between the middle cortical surface and the outer CSF hull. To achieve sub-voxel accuracy while minimizing numerical errors, I utilized an RK4 integration to generate the streamlines (Lee et al.,

2016). The local EA-CSF was computed by integrating the CSF probability along these generated streamlines. Similar to CT measurements, a geodesic heat kernel with a 20 mm bandwidth was used to smooth the resulting local EA-CSF maps (Chung et al., 2005). Figure 4.4 provides an example of the SA, CT, SCI and local EA-CSF features extracted along 6-month-old infant's cortical surface.



(b) Surface Area (SA)

(a) Cortical Thickness (CT)

(c) Local Shape Complexity Index (SCI)

(d) Local Extra-Axial CSF (EA-CSF)

Figure 4.4: The features extracted from the cortical surface of a 6-month-old infant scan. The CT and local EA-CSF measures shown here were smoothed using a 20 mm geodesic heat kernel.

The extracted measures reflect changes in biological processes; however, a degree of dependence exists among measures. While SA and CT are independent of each other, SA and local EA-CSF are sensitive to similar anatomical changes. For example: *(1)* widening of a sulcus would not change

CT, but it would increase SA, increase local EA-CSF, and decrease SCI; *(2)* widening of a gyral crown would not change CT and local EA-CSF, but it would increase SA, and decrease SCI.

Dimensionality Reduction: The previously described cortical measures were extracted from highly-sampled 3D cortical surfaces, forming a high-dimensional feature list (160,000 features or more). Based on the extracted features, machine learning classifiers can be used to predict diagnostic outcomes. Such classifiers undergo a learning process in which they learn the correct category labels for each set of features. However, it is hard to train stable classifiers using such a high-dimensional feature space without overfitting, particularly in the context of limited MRI datasets. One possible solution to overcome such an overfitting problem is to employ a supervised or unsupervised dimensionality reduction technique (Van Der Maaten et al., 2009). Unsupervised methods run the risk of losing relevant information while supervised methods tend to be more biased and, therefore, harder to generalize. Alternatively, dimensionality can be reduced by first summarizing the features via prior anatomical knowledge, such as summarizing vertex-wise cortical features via a cortical subdivision that divides the cortex into a mosaic of anatomically and/or functionally distinct, spatially adjacent areas using prior parcellation atlases (Glasser et al., 2016).

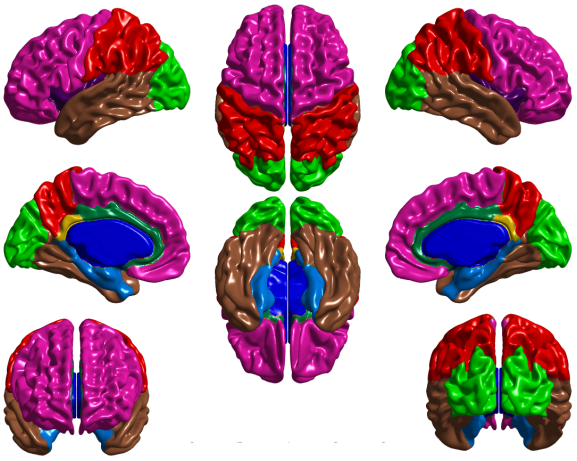
Due to the scarcity of infant cortical surface parcellations, I explored the use of different adult parcellations to reduce the dimensionality of the extracted high dimensional cortical features (see Figure 4.5). In particular, I investigated utilizing 4 anatomically-based cortical parcellations, namely, Lobar (8 ROIs), Desikan-Killiany (Desikan et al., 2006) (62 ROIs), AAL (Achard et al., 2006) (78 ROIs) and Destrieux (Destrieux et al., 2010) (148 ROIs) atlases. I also investigated using functionally based labeling (333 ROIs) of the brain to provide the required regional summaries (Gordon et al., 2014). As illustrated in Figure 4.5c, a Brodmann-based anatomic labeling atlas such as AAL separates the cortical regions using sulcal valleys, whereas many sulcal valley areas are part of the same functional parcel (Figure 4.5e). This suggests that such functional parcellations could be more useful when analyzing NDDs, such as ASD. SA, CT, SCI, and local EA-CSF cortical measurements were summarized by averaging their values across each of the ROIs defined by the

parcellation atlases mentioned above. The investigated parcellation atlases vary in the number of ROIs defined, allowing for a multi-resolution analysis of the extracted cortical measurements.

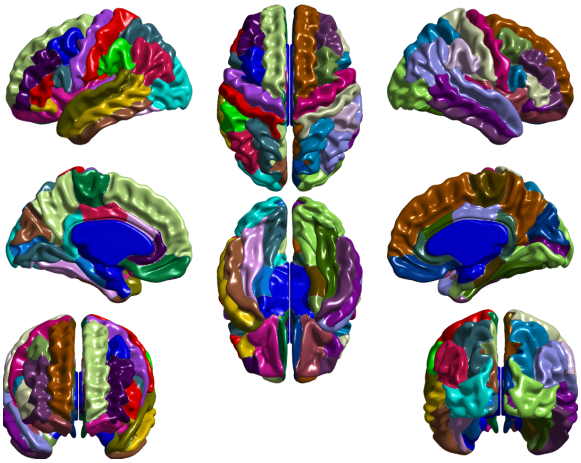
4.2.4 Deep Learning Prediction

The objective of this work is to find the optimal parsimonious subset of MRI brain features at 6 months that would result in the most accurate prediction of ASD at 24 months of age. A machine learning classification algorithm based on a deep learning network was used to achieve that goal. I investigated the use of SA, CT, SCI, local EA-CSF, and ICV along with binary sex information and age at MRI scan in predicting HR-ASD diagnosis at 24 months of age. As shown in Table 4.1, there is a significant sex imbalance between HR groups. Hence, the sex information predictive ability (e.g., when classifying all males as HR-ASD: 55% Accuracy (ACC); 86% SEN; 46% SPC; 29% PPV; 93% Negative Predictive Value (NPV)) was exploited in this work by feeding such information directly to the deep learning network as an additional feature. Finally, all features were rescaled or standardized using Z-score normalization that rescales the features to have a zero mean and unit variance. Standardizing the features is not only important if we are comparing measurements that have different units, but it is also a general requirement for many machine learning algorithms.

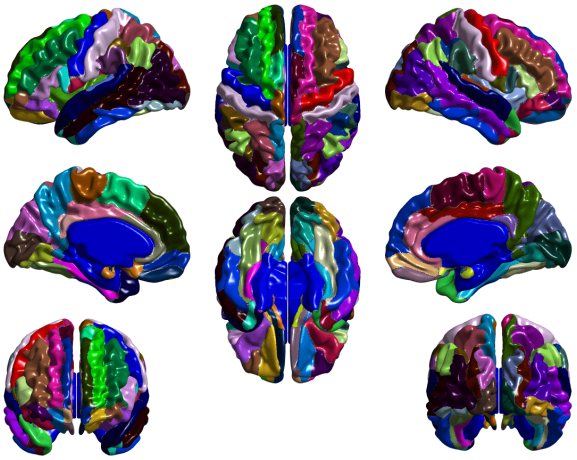
Solving Classes Imbalance Problem: It has been recognized that the class imbalance (HR-ASD: HR-Neg ratio = 4:1) problem present in the IBIS dataset (which is consistent with the HR population) has a substantial negative impact on training deep learning models (see Chapter 2, Section 2.1 for more information). To handle the group class imbalance in the given dataset, I applied SMOTE (Chawla et al., 2002), which instead of creating copies, works by creating synthetic samples from the minority class (HR-ASD) to match the number of samples in the majority class (HR-Neg). The algorithm selects two or more similar instances (using a distance measure) and perturbing an instance one attribute at a time by a random amount within the difference from the neighboring samples. These additional synthetic samples were used only for training the proposed classifier and were not used when applying the trained classification.



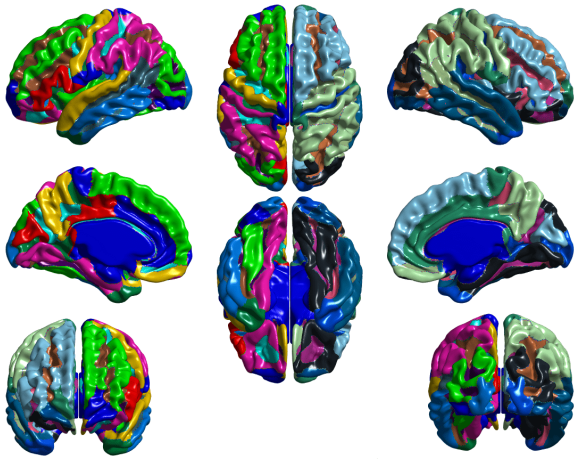
(a) Lobar (8 ROIs)



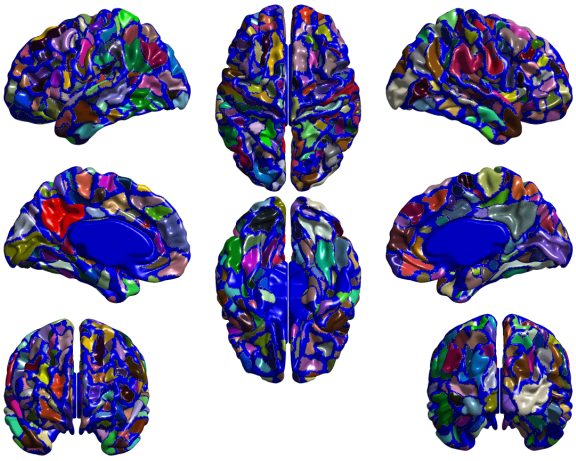
(b) Desikan-Killiany (62 ROIs)



(c) AAL (78 ROIs)



(d) Destrieux (148 ROIs)



(e) Functional (333 ROIs)

Figure 4.5: Different (a-d) anatomically and (e) functionally based parcellation atlases utilized for reducing the dimensionality of cortical measures by averaging feature values within a given ROI.

Rationale for Non-Linear Classification: The ASD brain exhibits an abnormal growth trajectory, including a period of early overgrowth that occurs within the first year of life (Hazlett et al., 2011). Hence, the brain of a child with ASD may show more considerable morphological abnormalities. Linear models (Hazlett et al., 2011; Fishbaugh et al., 2012) have been applied to estimate these growth trajectories or growth patterns. However, these are primarily concerned with changes in total volume and not specific regions in the brain, and on the other hand, how the growth patterns in particular brain regions may differentiate typically developing children with autism. Also, linear models may not accurately represent complex (i.e., non-linear) growth patterns in the developing brain (Giedd et al., 1999). As a result, classification methods that use linear models to recognize these intricate brain growth patterns may not be the most suitable choice. Recently, deep learning (Hinton and Salakhutdinov, 2006; Shin et al., 2013; Cireşan et al., 2013) has been used to estimate low dimension codes that are capable of encoding possibly non-linear relationships in high dimension data. Unlike linear models, the general concept of deep learning is to learn highly compact hierarchical feature representations by inferring simple ones first and then progressively building up more complex ones from the previous level. Deep learning may help in understanding such brain growth patterns because of its ability to infer complex non-linear relationships.

Multi-Layer Perceptron Classifier: To test the diagnostic capabilities of the selected features, an MLP classifier was used (Figure 4.6) (Chapter 2, Section 2.3 for more information). An MLP can be viewed as a logistic regression classifier where the input is first transformed using a learned non-linear transformation. This transformation projects the input data into a space where it becomes linearly separable. This intermediate layer is referred to as a hidden layer. A single hidden layer is sufficient to make MLPs universal approximators. However, there are substantial benefits to using many such hidden layers (deep learning) (Gardner and Dorling, 1998). As shown in Figure 4.6, the proposed network include two hidden layers with non-linear activation functions (e.g., ReLU) and an output layer with a single neuron with a sigmoid activation function. Dropout layers (Srivastava et al., 2014) were used after each hidden layer to overcome the over-fitting problem associated with deep neural nets with many parameters. The fundamental idea is to randomly drop units and their

connections from the neural network during training, which prevents units from co-adapting too much. During training, dropout samples from an exponential number of different thinned networks. At test time, it is easy to approximate the effect of averaging the predictions of all these thinned networks by merely using a single unthinned network that has smaller weights. The use of dropout techniques leads to significant improvements compared to other regularization techniques. Weight decay (ℓ_2) regularization was also used to constrain further the trained network, which is useful when dealing with limited medical imaging datasets.

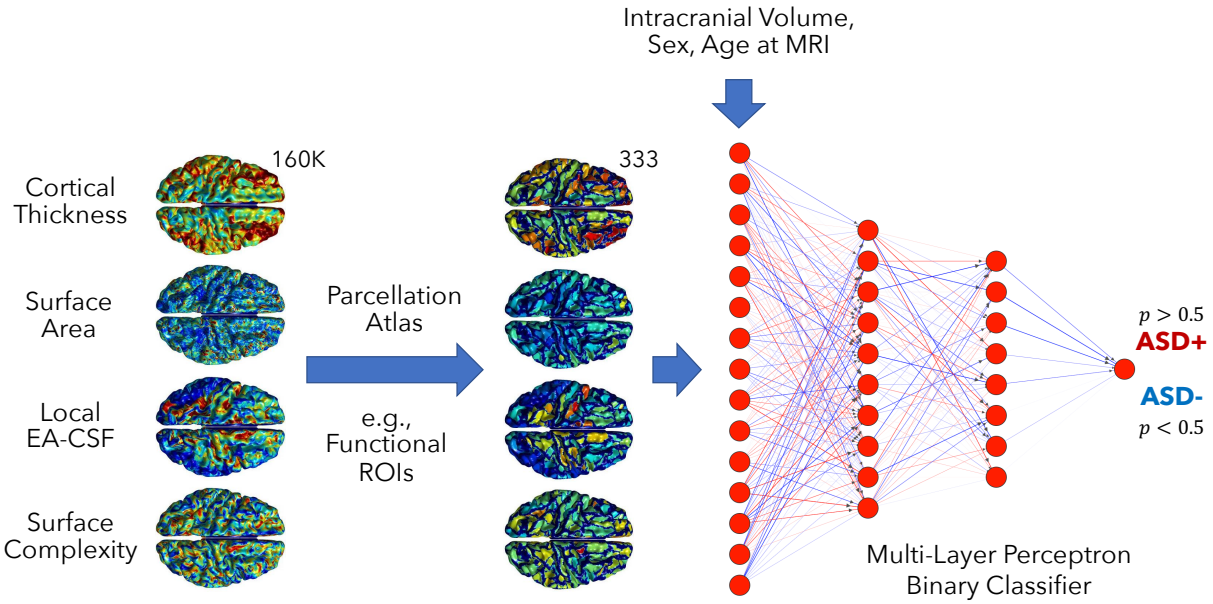


Figure 4.6: Illustration of the MLP classifier used to test the diagnostic capabilities of the summarized cortical measurements when combined with other relevant demographic information.

4.3 Experimental Results

In this section, I first illustrate the usefulness of the proposed local EA-CSF measure in highlighting group differences at 6 months of age in HR infants of developing ASD. Then, I investigate the ASD predictive ability of the different extracted cortical measures (SA, CT, SCI, and local EA-CSF) using the proposed Deep-ASD prediction framework. I finally explore finding the optimal parsimonious combination of such cortical measures at 6 months that would result in the most accurate prediction of ASD at 24 months of age in the HR population.

Local EA-CSF Group Analysis: Figure 4.7 shows the mean and standard deviation of local EA-CSF maps on a template middle surface for the HR-ASD and HR-Neg subgroups. Consistent with previous reports (Shen et al., 2013, 2017), infants who developed ASD visually had a higher EA-CSF volume at 6 months compared to infants without ASD, which was mainly observed over the frontal lobe areas. Moreover, significant inter-subject variability was observed in both HR-ASD and HR-Neg subgroups (mean coefficients of variation of 31.58% and 32.61%), confirming the heterogeneous nature local EA-CSF patterns in the investigated IBIS dataset.

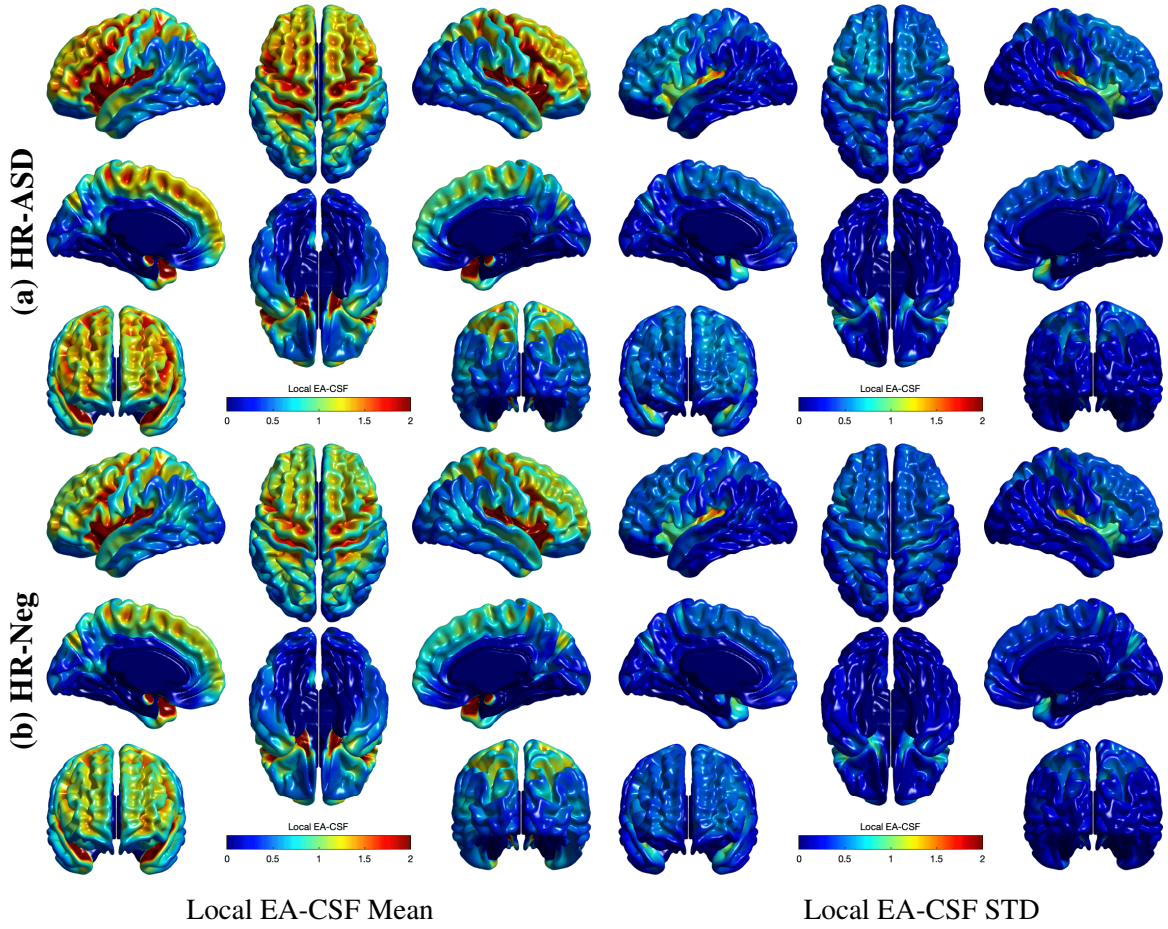
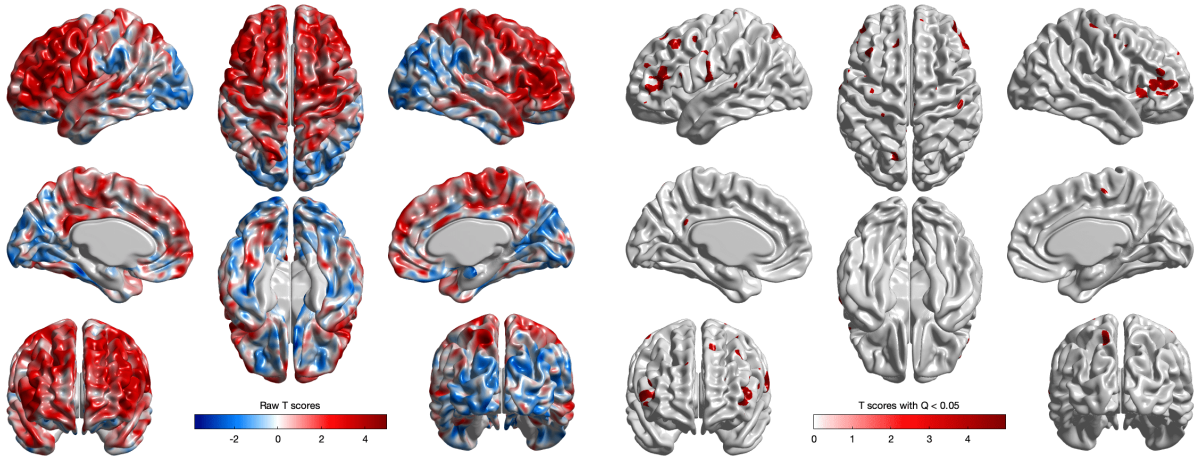


Figure 4.7: The mean and standard deviation of vertex-wise local EA-CSF extracted from 6-month-old (a) HR-ASD and (b) HR-Neg infants. Overall the patterns at the 2 groups look similar but with higher mean local EA-CSF observed for the HR-ASD group, particularly in the frontal lobe.

I designed a preliminary linear model to investigate the 6 months local EA-CSF group differences in infants with HR of developing ASD. The local EA-CSF was used as a dependent variable Y, and the fixed effects were composed of 3 covariates: age at scan, sex, and the diagnostic group:

$$Y_i = \beta_0 + \beta_1 \text{Sex}_i + \beta_2 \text{Age}_i + \beta_3 \text{Group}_i. \tag{4.1}$$

Vertex-Wise Analysis Results: The cross-sectional linear model was implemented and fitted in SurfStat, which is a toolbox for statistical analysis of cortical surface measurements applying random field theory for statistical inference (Chung et al., 2010). As expected, statistical analysis revealed that local EA-CSF was higher in most brain regions in HR-ASD infants, as demonstrated in Figure 4.8a. However, after correcting for multiple comparisons using standard FDR (Benjamini and Hochberg, 1995) correction ($Q < 0.05$), only few brain regions showed statistically significant group differences (Figure 4.8b). Each surface model is composed of a large number of vertices (= 163,842) compared to the sample size (= 187) in the HR population, which could explain why subtle group differences did not survive the employed FDR multiple comparisons correction.



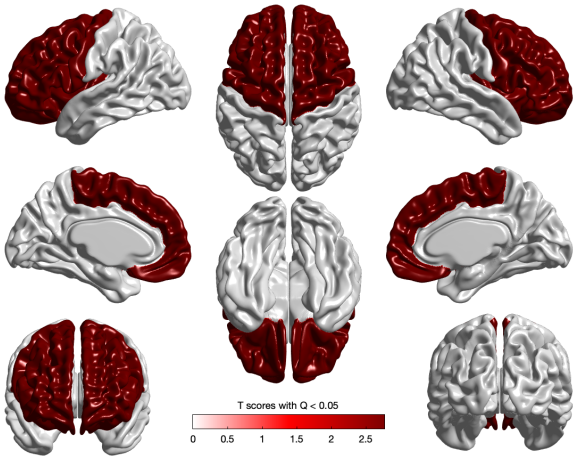
(a) Raw T Scores (b) T scores with FDR $Q < 0.05$

Figure 4.8: The group effect (HR-ASD > HR-Neg) on local EA-CSF at 6 months in the HR population. As indicated from the (a) raw T scores, most cortical regions showed higher local EA-CSF in the HR-ASD group; however, only a few cortical areas showed statistically significant differences after (b) correction for multiple comparisons using FDR correction ($Q < 0.05$).

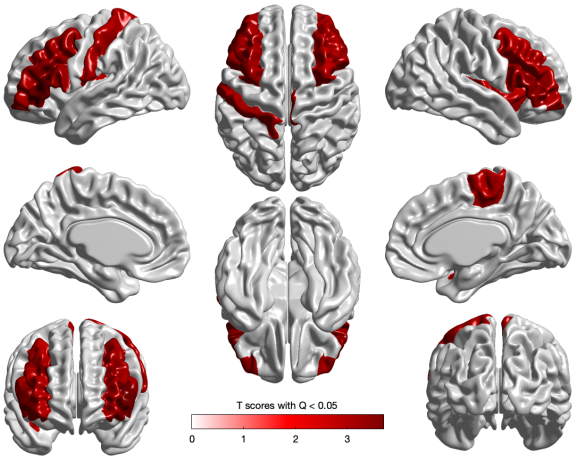
ROI Analysis Results: I also employed an ROI-based statistical analysis using parcellation atlases (Lobar, Desikan-Killiany, AAL, Destrieux, Functional; see Figure 4.5). In particular, local EA-CSF cortical measurements were summarized by averaging their values across each of the defined ROIs before fitting the cross-sectional linear model described in Equation 4.1 and the

subsequent correction for multiple comparisons (FDR $Q < 0.05$). First, as shown in Figure 4.9a, robust group differences were observed in the frontal lobe, consistent with the mean local EA-CSF maps shown in Figure 4.7. As shown in Figure 4.9, brain regions showing statistically significant group differences vary widely by parcellation scheme, which suggests that such an unsupervised dimensionality reduction step may result in the loss of relevant predictive information. However, among the investigated parcellation atlases, an ROI-based analysis functionally based labeling of the brain (Gordon et al., 2014) was able to highlight more brain areas with significant group differences (FDR $Q < 0.05$) as compared to the ROI-based analysis performed using other anatomically-based parcellation atlases. Hence, a prediction framework utilizing such functional parcellations for dimensionality reduction would yield optimal prediction of later ASD diagnosis when applied to the 6 months local EA-CSF information. To confirm such observation, the extracted cortical measures (including local EA-CSF) have been evaluated individually as single predictors of 24-month ASD diagnosis with different parcellation atlases employed for dimensionality reduction.

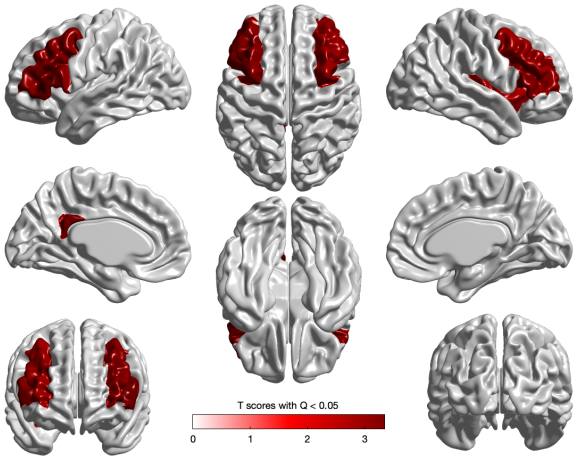
Evaluating the Predictive Ability of the Extracted Cortical Measurements: A 6-month structural MRI scan was used as input into the Deep-ASD computational neuroscience framework (Figure 4.2) to generate the SA, CT, SCI, local EA-CSF cortical measures describing the infant brain anatomy (Figure 4.4). Each of these measures was generated at 163,848 points across the cortical surface. Then, for feature reduction purposes and anatomical localization, the points were parcellated into regions-of-interest defined by the utilized anatomically and functionally based cortical parcellation atlases (Figure 4.5), thereby generating precise, anatomically localized regions for each of the 4 measures. The 4 measures were then entered independently as a single measure into a fully cross-validated 10-fold deep learning prediction algorithm to predict ASD diagnosis in the HR population (i.e., classifying HR-ASD from HR-Neg infants). The age at scan, sex, and ICV information was also included in all prediction analyses as such information was found to be of critical importance to parse the heterogeneity of ASD datasets (Hazlett et al., 2017).



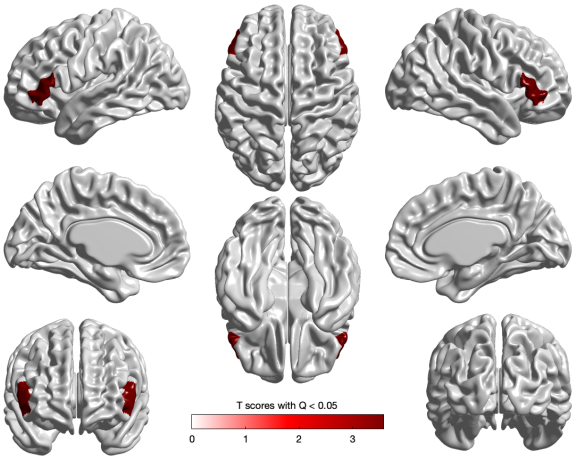
(a) Lobar (8 ROIs)



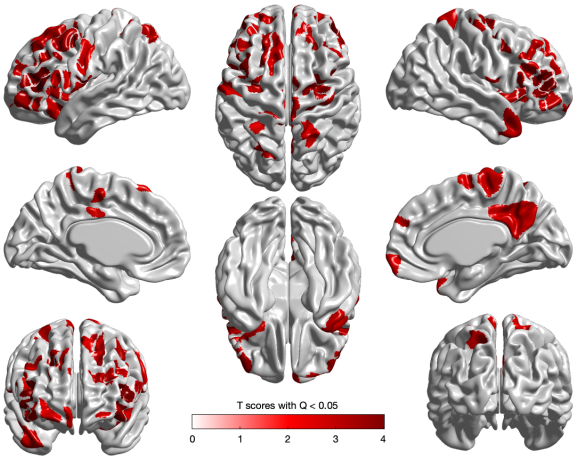
(b) Desikan-Killiany (62 ROIs)



(c) AAL (78 ROIs)



(d) Destrieux (148 ROIs)



(e) Functional (333 ROIs)

Figure 4.9: The group effect (HR-ASD > HR-Neg; FDR $Q < 0.05$) on local EA-CSF at 6 months in the HR population using ROI-based analysis based on different cortical parcellation atlases.

		ACC (%)	SEN (%)	SPC (%)	PPV (%)	NPV (%)
Surface Area	**Lobar ROIs**	**75.3**	**62.5**	**78.5**	**44.3**	**89.6**
	Desikan-Killiany ROIs	71.6	27.5	83.2	32.3	81.9
	AAL ROIs	73.8	37.5	83.3	46.1	84.5
	Destrieux ROIs	76.4	35.5	87.2	38.3	84.3
	Functional ROIs	75.9	35.5	87.3	42.2	83.7
Cortical Thickness	Lobar ROIs	75.4	31.7	86.6	33.8	83.5
	Desikan-Killiany ROIs	75.9	37.5	85.9	39.0	84.4
	AAL ROIs	**76.5**	**66.7**	**86.6**	**50.0**	**84.5**
	Destrieux ROIs	75.3	32.5	86.5	34.8	83.4
	Functional ROIs	77.5	34.2	88.5	47.8	84.0
Surface Complexity	Lobar ROIs	72.1	60.8	75.2	38.4	88.5
	Desikan-Killiany ROIs	79.1	51.7	85.9	52.4	83.0
	AAL ROIs	77.1	43.3	85.3	49.5	86.2
	Destrieux ROIs	**81.4**	**50.8**	**89.3**	**55.1**	**87.9**
	Functional ROIs	79.7	32.5	90.7	48.3	84.5
Local EA-CSF	Lobar ROIs	72.7	60.0	75.8	39.3	88.6
	Desikan-Killiany ROIs	74.8	43.33	83.2	52.58	85.2
	AAL ROIs	76.9	42.5	85.8	48.3	85.6
	Destrieux ROIs	73.8	43.33	81.8	47.0	84.7
	Functional ROIs	**77.5**	**38.3**	**87.9**	**53.4**	**84.8**

Table 4.2: Summary of the classification results obtained in the HR population using the mean ACC, SEN, SPC, NPV, and PPV scores. Each of the 4 cortical measures was entered independently as a single measure into a fully cross-validated 10-fold Deep-ASD prediction framework to predict ASD (i.e., HR-ASD from HR-Neg infants) diagnosis. The age at scan, sex, and ICV information was included in all prediction analyses. Bold rows indicate the parcellation atlases that were optimally selected for each of the 4 cortical measures to perform the subsequent analyses.

The proposed networks were compiled using Stochastic Gradient Descent with Nesterov Momentum optimization and a weighted binary cross-entropy loss function. The network architecture (e.g., hidden layers nodes, activation function) and training parameters (e.g., dropout rate, learning rate, ℓ_2 regularization weight, batch size, epochs) were determined using a grid search procedure. To determine the optimal parcellation atlas for each of the 4 cortical measures, the network performance was evaluated using the mean ACC, SEN, SPC, PPV, and NPV scores. PPV was emphasized as the primary evaluation metric, as it reflects the potential clinical relevance of the predictive test for infants born into a high-risk family. It is important to note that PPV is influenced by the prevalence of ASD in the sample tested. Therefore, a predictor with higher PPV was considered a more accurate test if it also maintained high ACC, SEN, SPC, and NPV.

As shown in Table 4.2, SCI (Destrieux ROIs: 81.4% ACC; 55.1% PPV) and local EA-CSF (Functional ROIs: 77.5% ACC; 53.4% PPV) were the two measures with the highest PPV and ACC scores. On the other hand, CT (AAL ROIs: 66.7% SEN) and SA (Lobar ROIs: 62.5% SEN) provided the highest SEN results; however, with a significantly lower PPV and ACC scores. Overall, the obtained results suggest that an ASD diagnosis test based on a single predictor would not yield accurate enough classification results for useful clinical application. However, the complementary information provided by the different cortical measures suggests that combining such weaker classifiers in an ensemble model would improve the overall predictive performance. Next, I investigate such an observation by exploring different combinations of cortical measures.

Finding the Optimal Parsimonious Subset of Cortical Measurements: To improve the diagnostic results of the trained models, a classifier ensemble was constructed from the independent weaker classifiers highlighted in Table 4.2. In machine learning, a classifier ensemble refers to a collaborative decision-making system formed from N individual classifiers, in which a strategy is employed for combining the predictions of ensemble members to produce a single decision as output (Zhang and Ma, 2012). In this work, I apply a soft voting strategy to combine the probabilities produced each of the N individual classifiers with weights and threshold determined experimentally

via a grid search procedure (see Figure 4.10). A classifier ensemble is preferred to only using a single MLP classifier with a concatenated high-dimensional input feature vector as it maintains reasonable model complexity (i.e., number of trainable parameters) with available limited datasets.

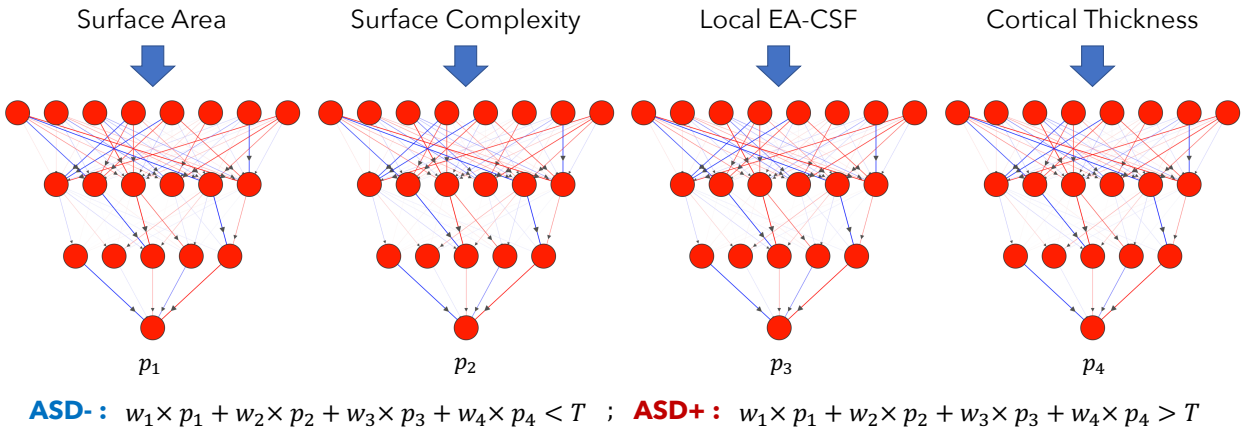


Figure 4.10: Illustration of the classifier ensemble formed from N individual classifiers to combine the predictions of ensemble members to produce a single decision as output. A soft voting strategy was utilized to mix the probabilities produced each of the N individual classifiers with w_i and T determined experimentally via a grid search procedure. Shown here a combination of all 4 cortical measures, i.e., $N = 4$; a similar design was used to construct a classifier ensemble with $N = 2, 3$.

As illustrated in Table 4.3, the 4 individual predictors were first evaluated in combinations of two. The combination of SCI + local EA-CSF yielded the best predictive performance with the highest PPV (76.5%) and ACC (87.2%); however, this combination did not yield the best SEN (60.8% as compared to 82.5% of SA + CT) and NPV (90.9% as compared to 96% of SA+CT). When predictors were evaluated in combinations of three, the combination of SA + SCI + local EA-CSF yielded the highest PPV (85.5%) and ACC (88.7%) but inferior SEN (68.3% as compared to 80.0% of SA + CT + SCI) and NPV (93% as compared to 95.5% of SA + CT + SCI). That improvement in ACC and PPV resulted from adding SA to the SCI + local EA-CSF was mainly driven by an increase in the number of True Positives (TP) and a reduction in the number of False Negatives (FN). Swapping the local EA-CSF with CT in the previous combination resulted in a significant SEN and NPV improvement driven by a substantial increase in the number of TP and the number of True Negatives (TN); however, a rise in FN and False Positives (FP) was also observed,

which explains the superior SEN and NPV and inferior ACC and PPV. Hence, mainly based on the obtained PPV, SA + SCI + local EA-CSF is considered the best performing 3-way combination.

Finally, combining all predictors (SA + CT + SCI + local EA-CSF) resulted in the best ACC (89.7%) with significant increase in SEN (78.3%) and NPV (95.2%) as compared to the SA + SCI + local EA-CSF results; still, a drop in PPV (80.2%) was observed. Such improvement in ACC, SEN, and NPV resulted from also adding the CT to the best performing 3-way combination was mainly driven by a decrease in the number of FN while the inferior PPV is attributed to a rise in the number of FP. As discussed in Section 4.2, CT information has the least overlap with the other cortical measures (SA, SCI, and local EA-CSF), which may explain the improved ACC propelled by the additional complementary information. However, since CT at 6 months is known to be noisy and less reliable, adding CT to the prediction model negatively affected the obtained PPV.

		ACC (%)	SEN (%)	SPC (%)	PPV (%)	NPV (%)
Two Predictors	SA + CT	77.1	82.5	75.8	64.7	96.0
	SA + SCI	83.9	66.7	88.7	77.2	92.1
	SA + EACSF	84.5	73.3	87.3	64.2	92.9
	CT + SCI	77.1	82.5	75.80	64.7	96.0
	CT + EACSF	85.1	52.5	93.3	74.7	88.8
	SCI + EACSF	**87.2**	**60.8**	**94.0**	**76.5**	**90.9**
Three Predictors	SA + CT + SCI	85.0	80.0	86.5	77.5	95.5
	SA + CT + EACSF	87.2	73.3	90.7	71.2	93.5
	SA + SCI + EACSF	**88.7**	**68.3**	**93.9**	**85.5**	**93.0**
	CT + SCI + EACSF	88.3	70.8	92.7	79.3	92.7
All Predictors	**SA + CT + SCI + EACSF**	**89.7**	**78.3**	**92.5**	**80.2**	**95.2**

Table 4.3: Summary of the classification results obtained in the HR population using the mean ACC, SEN, SPC, NPV, and PPV scores from a 10-fold cross-validation procedure. The 4 individual predictors were evaluated in combinations of 2, 3, and 4 using a classifier ensemble with the soft voting strategy employed for combining the predictions of different classifiers.

Despite such a drop in the obtained PPV (the primary performance metric), combining all predictors resulted in a notable improvement in the SEN while maintaining similar performance levels as compared to the previous sMRI prediction system proposed in (Hazlett et al., 2017) (91% ACC; 78% SEN; 94% SPC; 82% PPV; 91% NPV). Hence, applied to the HR population, a diagnostic test based on a prediction model utilizing all four cortical measures would be preferred in a clinical setting. In total, the obtained results are consistent with previous methods (Hazlett et al., 2017) that demonstrated the ability of brain MRI in infancy to predict ASD but substantially improve upon previous reports by achieving comparable accuracy ($\sim$80% PPV) at an earlier age by accurately predicting later ASD diagnosis with a conventional 6 months sMRI scan.

Using Trained HR-ASD/HR-Neg Network to Predict LR-Neg Subjects: The current study also examined a cohort of LR infants with no family history of ASD, intellectual disability, or major psychiatric disorder and who had an older sibling with typical development. LR infants were assessed in parallel with the HR infants (i.e., evaluated with the same neuroimaging and diagnostic assessments). All LR infants included in this study were confirmed as not having ASD at 24 months (LR-Neg). Thus, the LR-Neg cohort served as an important negative control group, providing an opportunity to validate the same prediction classifier by testing its ability to predict LR infants as being negative for ASD correctly. The identical fixed classifier ensembles highlighted in Table 4.3 (trained and cross-validated on the HR infants in the analyses above) were applied to the LR infants (the LR dataset was not used in the earlier training phase). A hard majority vote strategy was employed to combine predictions from the different models in the cross-validation evaluation.

As shown in Table 4.4, consistent with the ACC obtained in HR infants, applying the two-way combination of SCI + local EA-CSF in the LR-Neg group was able to correctly predict 87.3% of LR-Neg infants as negative for ASD (87.3% ACC obtained in the HR population). As shown in Figure 4.11, the obtained ASD-positive vote count generated by the SCI + local EA-CSF classifier ensemble indicates that an advanced voting strategy that employs a model selection procedure may help improve the current LR-Neg classification results. Similar results were observed for the combination of SA + local EA-CSF + SCI (86.3% ACC in the LR population is comparable

to the 88.7% ACC obtained in the HR population). However, a significant drop was observed for the classifier ensemble utilizing the 4 cortical measures (81.4% ACC in the LR population is significantly lower than the 89.7% ACC obtained in the HR population), suggesting that adding CT information resulted in an ensemble model that overfit to the HR population.

		ASD-Negative (%)	ASD-Positive (%)
Two Predictors	**SCI + EACSF**	**87.3**	**12.7**
Three Predictors	**SA + SCI + EACSF**	**86.3**	**13.7**
All Predictors	**SA + CT + SCI + EACSF**	**81.4**	**18.6**

Table 4.4: Summary of the classification results obtained in the LR population using different combinations of cortical measures. Reported here are the results of applying the already trained network (the HR population) to the LR sample. Since 10-fold cross-validation was first applied to the HR sample, this yielded 10 prediction networks that were then applied to the LR sample. The predicted outcomes of the LR sample were determined by a majority vote of these 10 networks.

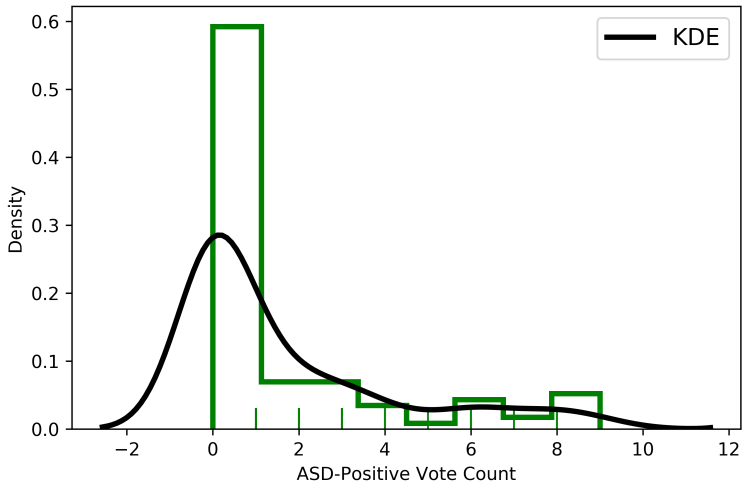


Figure 4.11: The distribution of ASD-Positive vote count generated by the SCI + local EA-CSF classifier ensemble. Using a simple majority voting (ASD-positive if > 5 votes), the proposed model was able to correctly predict 87.3% of LR-Neg infants as negative for ASD. Only a few samples were considered borderline (received 5 ASD-positive votes), confirming the usefulness of the proposed methods when deployed in a diagnostic setting. Also, none of the incorrectly classified LR-Neg infants received a unanimous vote, which suggests that a more advanced voting strategy that employs a model selection procedure could further improve the current classification results.

In total, both SCI + local EA-CSF and SA + SCI + local EA-CSF were able to provide ACC in LR-Neg infants similar to that obtained in the HR population. These findings suggest that the deep learning prediction network had effectively learned to distinguish differences in ASD-related brain features, even among infants who did not develop ASD. However, the absence of LR-ASD infants to perform adequate evaluation and the relatively low classification results (as compared to low ASD prevalence of $\sim 2\%$) made the proposed prediction models still not clinically useful in the LR population. With the collection of new data that provide a representative sample of the general population, the proposed methods combined with adequate transfer learning (Tan et al., 2018) may be able to achieve classification results that are clinically actionable in the future.

4.4 Summary and Discussion

This Chapter described a complete framework to generate, summarize, and optimally combine measures of cortical surface anatomy and CSF from a structural MRI scan at 6 months to accurately predict ASD diagnosis at 24 months in HR infants. The proposed methods also showed promising results in predicting 24-month outcomes of LR infants who did not develop ASD. These results add to previous research showing that brain imaging features acquired during the presymptomatic period in infancy can predict later diagnosis of ASD while substantially improving upon earlier approaches by downwardly extending the prediction to a past age (from 12 months to 6 months) using a conventional structural MRI scan. The current results suggest the feasibility of presymptomatic detection of ASD in HR infants, affording the possibility of intervention in the first year of life, approximately two years before treatment occurs in the community for most children.

Clinical Significance to Autism: Recent emphasis has been placed on the importance of identifying biological markers to aid in detecting early risk for ASD (Ruggeri et al., 2014). Early detection of ASD makes early intervention both feasible and efficient because children could be identified for intervention based on the presence of reliable risk markers in infancy. This approach would fully leverage the consensus in the field that 'earlier is better' in ASD: earlier treatment is more effective than later treatment in reducing impairments and improving long-

term outcomes for the child (Cidav et al., 2017). The brain is more malleable in infancy than at later ages and more responsive to treatment, and recent research has indicated the existence of distinct developmental windows for maximal therapeutic benefit in NDDs. Furthermore, initiating behavioral intervention after a child is already diagnosed with ASD may be less effective because brain changes are already consolidated, and the presence of the behavioral features of autism makes intervention more challenging. Behavioral markers in the first year of life have yet to demonstrate sufficient sensitivity or specificity at the individual-level for clinically useful prediction of later ASD diagnosis (Bussu et al., 2018). The current work showed that MRI-derived brain markers at 6 months could reliably predict which HR infants will be diagnosed with ASD with highly accurate sensitivity and specificity.

Strengths Over Previous Studies: First, based on the previous reports, I developed this approach guided by the hypothesis that regional CSF and cortical surface anatomy would be predictive of later ASD diagnosis. In two previous infant studies, global volume EA-CSF showed high sensitivity but insufficient specificity in predicting later ASD diagnosis (Shen et al., 2013, 2017). By generating localized EA-CSF measurements, the presented deep learning classifier was able to learn the specific regions of EA-CSF that provided better discrimination between HR-ASD and HR-Neg infants. Second, sMRI is more scalable to the community than more complex neuroimaging acquisitions (e.g., fMRI), which are sensitive to differences in scanner platforms, acquisition parameters, and image analysis methods. The cortical measures utilized in this work were generated from a conventional sMRI scan that could be acquired from standard clinical scanners, enabling these predictive measures to be used in general clinical settings. The current findings suggest that early abnormalities of cortical surface anatomy and CSF volume in ASD are present in infancy, are detectable by conventional sMRI, and could serve as early indicators of altered neurodevelopment preceding the manifestation of symptoms of ASD. If replicated, this may warrant discussions as to whether sMRI should be considered as standard practice for screening HR infants, since the scans in the current study were acquired during natural sleep and present a safe and informative index of early abnormal brain development.

Finally, though it is unknown whether the current framework could be applied to LR infants who develop ASD, there is evidence that LR children exhibit the same brain and CSF abnormalities found in the current study of HR children (Shen et al., 2018). Hence, sMRI in a population sample might serve as a second-level screening measure once cost-effective behavioral (Miller et al., 2017), neurophysiological (Levin et al., 2017), or genetic risk markers (Miller et al., 2010) are identified as first-level screeners, approximating the moderate level of risk seen in siblings. This two-level approach would first narrow down which infants in the population are suitable candidates for a second level screen for ultra-high risk of later ASD, using conventional sMRI. This work serves as proof-of-principle that quantitative brain differences in ASD are detectable by sMRI at 6 months and can distinguish infants who will be diagnosed with ASD at 24 months from those who will not.

Conclusions: The current findings add to a growing literature demonstrating that neuroimaging can detect brain differences in the first year of life with sufficient accuracy to predict ASD before infants show behavioral symptoms. The current approach generated multiple measures of infant brain structure, across thousands of anatomical points along the cortical surface, taking full advantage of a novel computational framework. The proposed Deep-ASD framework contributes to an emerging field that uses computational neuroscience to bolster ongoing efforts in clinical neuroscience to ultimately improve clinical outcomes of brain-based disorders. Such efforts are mainly focused on utilizing deep learning to leverage multiple sources of biological variance to separate individuals into clinically meaningful groups with clinically actionable prognoses.

CHAPTER 5: CONVOLUTIONAL LEARNING ON CORTICAL BRAIN SURFACES

5.1 Overview

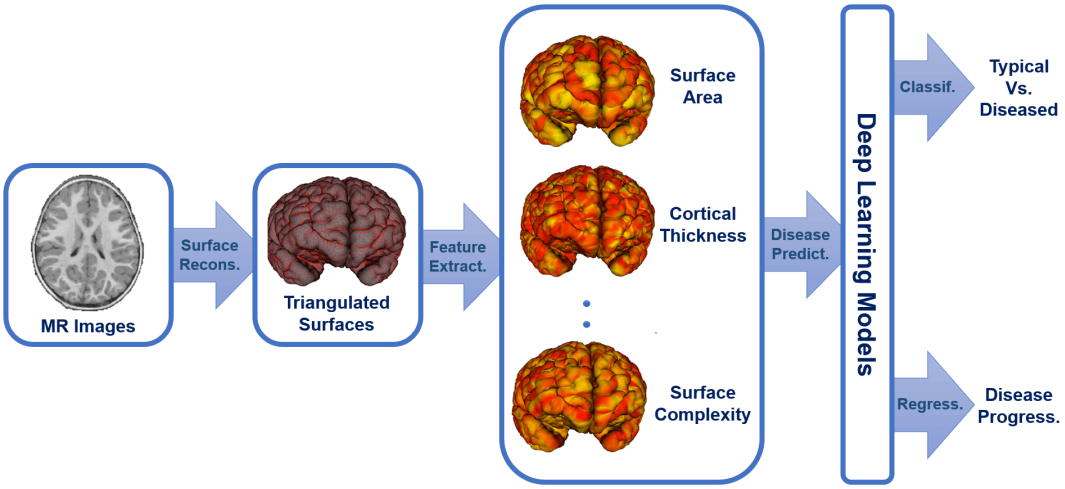


Figure 5.1: Deep learning-based framework for disease prediction where high-dimensional (100,000 or more) cortical measurements extracted from sMRI and analyzed on a per-vertex basis.

Cortical features, such as Cortical Thickness (CT) and Surface Area (SA) have been proposed as biomarkers for detection of various neurodevelopmental disorders (Hazlett et al., 2017; Hartberg et al., 2018) and neurodegenerative diseases (Schwarz et al., 2016; Racine et al., 2018). Such cortical measures are extracted from finely-sampled 3D cortical surfaces, forming a high-dimensional feature list with 100,000 or more features (see Figure 5.1). Based on the extracted features, deep learning classifiers are used to predict diagnostic outcomes. Such classifiers undergo a learning process in which they learn the true category labels for each set of features. However, high-dimensional clinical datasets are also limited in size, leading to a degradation in the accuracy and efficiency of the prediction system as a consequence of the curse of dimensionality. High-dimensional datasets lead to a drop in the performance of classifier performance as the increased model complexity would make it harder to train a stable classifier. Overfitting is another problem that arises when dealing

with high dimension features spaces, particularly when the number of features is high compared to the number of training samples. The resulting classifier ends up memorizing training examples, resulting in a good performance on training data but inferior performance on unseen testing data.

Dimensionality Reduction: The dimensionality of input high-dimensional feature space is typically reduced to avoid overfitting. The objective of dimensionality reduction is three-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying generating process. Supervised or unsupervised feature extraction techniques have been employed to achieve that task (Van Der Maaten et al., 2009). However, unsupervised methods run the risk of losing relevant information while supervised methods tend to be more biased and, therefore, harder to generalize. Alternatively, if possible, dimensionality can be reduced by summarizing the features via prior anatomical knowledge, such as summarizing vertex-wise cortical features via a cortical subdivision that divides the cortex into a mosaic of anatomically and/or functionally distinct, spatially adjacent areas using prior parcellation atlases (Glasser et al., 2016). However, current parcellation atlases are generic and not optimized to a given study (i.e., subjects' age or pathology). This is problematic in infant studies as currently only adult parcellations are employed due to the scarcity of infant cortical surface parcellations. Hence, there is a need for data-driven classifiers that can directly learn from a high-dimensional cortical feature space without employing a separate dimensionality reduction.

CNNs on Non-Euclidian Domains: One way to address this problem is to exploit CNNs' ability to extract hierarchical abstractions directly from raw high-dimensional data with almost no prior knowledge and with fewer parameters. However, extending the use of CNNs to applications where the input features live in irregular non-Euclidean domains is still challenging. These challenges stem from the missing notion of a grid on a non-Euclidean surface, in addition to the need to adapt the convolutional filters locally while sliding across the input surface. Notably, the design of the sliding kernels needs to be adapted locally to account for the changing underlying geometry of the surface. To solve this problem, a generalization of the CNN paradigm to non-Euclidean manifolds was introduced (Masci et al., 2015) based on a local geodesic system of polar coordinates to extract

"patches," which are then passed through a cascade of filters and linear and non-linear operators. The coefficients of the filters and linear combination weights are optimization variables that are learned to minimize a task-specific cost function. Although this implementation is promising, it usually fails if the surface mesh is very irregular or if the radius of the geodesic patches is large compared to a curvature radius of the shape. To solve these drawbacks, one would need to generate highly uniform surface representations as well as to decrease the geodesic path radius, which in turn might be a problem in terms of expected computational complexity. Another CNN generalization was presented (Boscaini et al., 2015) based on localized frequency analysis (a generalization of the windowed Fourier transform to manifolds) that is used to extract the local behavior of some dense intrinsic descriptor, roughly acting as an analogy to patches in images. The resulting local frequency representations are then passed through a bank of filters whose coefficients are determined by a learning procedure minimizing a task-specific cost. In their implementation, the authors addressed some of the limitations described in the previous paper; however, that work was only designed to capture shape descriptors on the surface and is not suitable to be used with other surface measures.

Proposed Surface-CNN: To overcome the above problems, I propose a novel CNN extension to non-Euclidean manifolds that utilizes geodesic-based kernels in learning the optimal features and brain regions in a data-driven way[1] (see Figure 5.2). This chapter's contributions are three-fold:

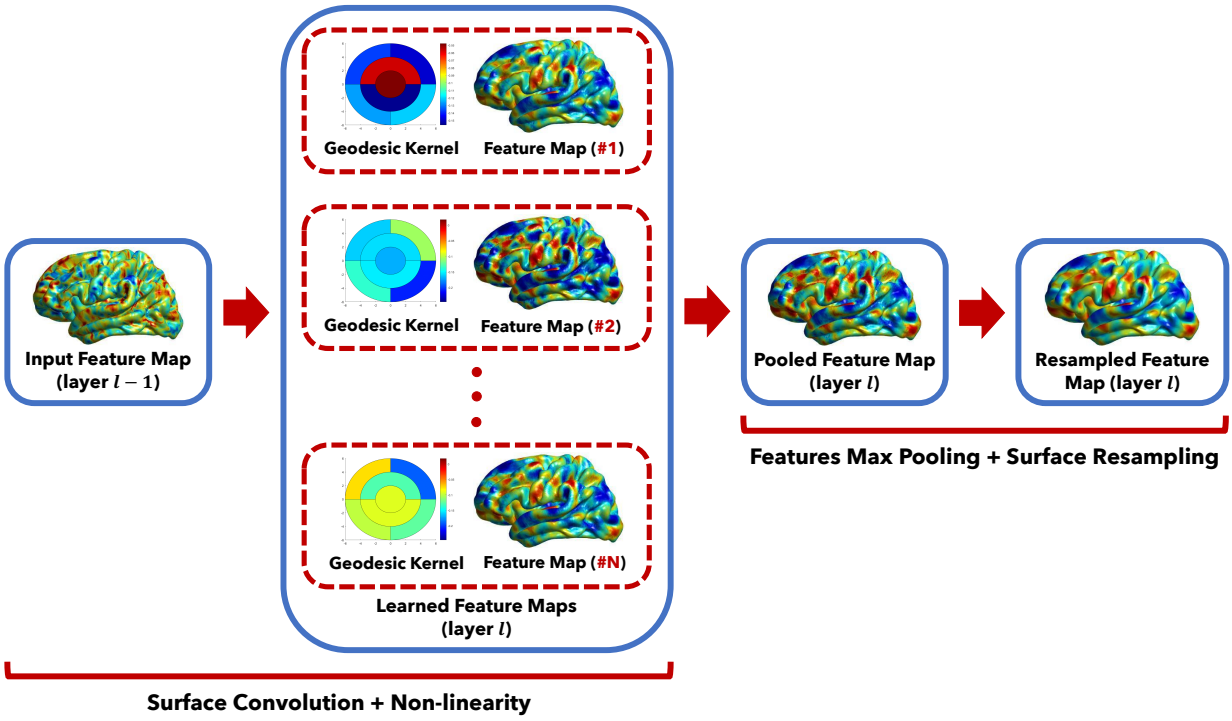
[1] A general definition of kernels on non-Euclidean cortical surfaces using a locally constructed geodesic grid, which in turn can be used to extract corresponding patches on the manifold.

[2] A general framework for learning on cortical surfaces using novel surface convolution, surface subsampling, and bottleneck layers designed to control the model complexity.

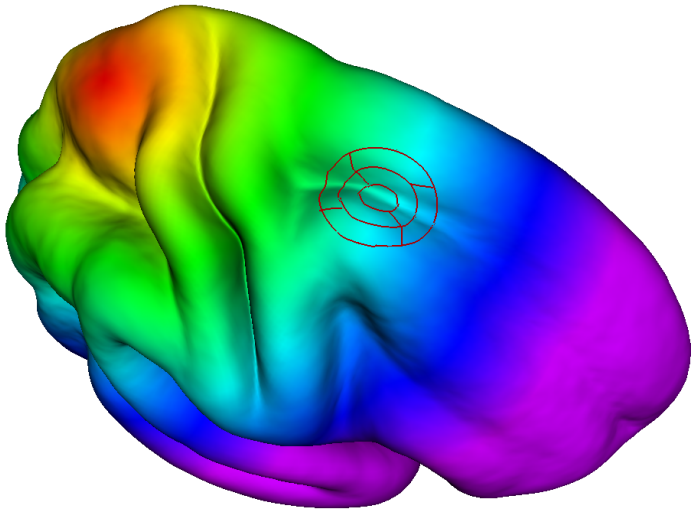
[3] An accurate disease classification using high-dimensional cortical measurements without the need for a separate dimensionality reduction typically employed to avoid model overfitting.

Details of the proposed CNN extension on cortical surfaces will be provided in the next section.

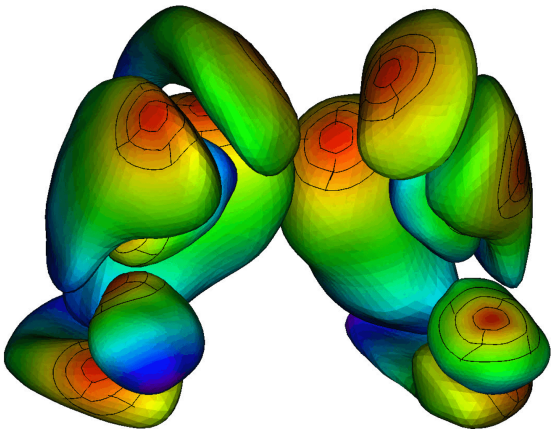
[1]The work is based on the previously published paper (Mostapha et al., 2018a). This chapter partially adapts text descriptions and figures from the published paper.



(a) Surface Convolution Block



(b) Geodesic Grid on a Cortical Surface



(c) Geodesic Grid on Subcortical Surfaces

Figure 5.2: The proposed extension of CNNs to non-Euclidean manifolds such as cortical brain surfaces. A Surface-CNN architecture is defined using a series of (a) surface convolution blocks, which consist of surface convolution, pooling, and resampling layers that are applied consecutively. The proposed convolution block relies on local surface kernels that are reconstructed using a local system of geodesic coordinates created on (b) cortical or (c) subcortical brain surfaces.

5.2 Methods

As shown in Figure 5.2, an extension to the classical CNN framework on cortical surfaces (Surface-CNN) is proposed, which includes newly developed surface convolution blocks that are applied consecutively. In this section, I provide details of the proposed surface convolution, surface subsampling, and bottleneck layers used to form a surface convolution block.

5.2.1 Convolutional Learning on Surfaces

Convolution of data living on cortical surfaces is defined as a correlation with a surface kernel that is used to extract corresponding patches on the manifold. A localized grid is established at each surface point in a way that considers the intrinsic shape of the underlying manifold. To achieve that, I rely on the geodesic (shortest path) distances computed at each surface point to reconstruct a local system of geodesic coordinates. The geodesic distance to a collection of points satisfies a non-linear differential equation. This so-called the Eikonal equation gives the viscosity $\phi(x, y)$ according to

$$||\nabla\phi|| = F. \tag{5.1}$$

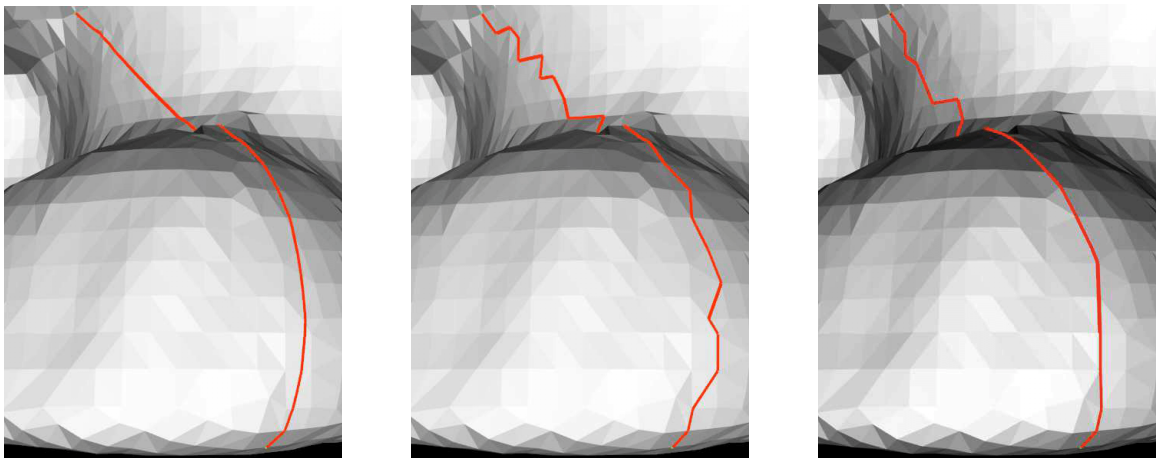
This ϕ can be interpreted as a weighted distance map from an initial seed, where the weights are given by the function $F(x, y)$ which is a scalar positive function (Peyre and Cohen, 2006). It follows that, the curve, $C(t)$ giving the level set of distance, defined as points on the front of the function ϕ at time t propagates following the evolution equation

$$\frac{d}{dt}C(t) = \frac{\vec{n}_{xy}}{P(x, y)}, \tag{5.2}$$

where $\vec{n}_{xy}$ is the exterior unit normal vector to the curve at the point $C(t) = (x, y)$. The function $F(x, y) = 1/P(x, y)$ is the propagation speed of the front, $C(t)$.

Geodesic Computation: When computing geodesic distances on a 2D manifold surface represented by a triangular mesh $M : (V, E, F)$, computational efficiency is preferred over accuracy (Novotni and Klein, 2002). Hence, instead of optimal solutions provided by slow $(O(V^2 \log V))$

exact methods (Mitchell et al., 1987), efficiently computable high-quality approximations of the true geodesic distance field for each vertex $v \in V$ are utilized in this work. This is typically achieved by formulating the problem in terms of a weighted graph, where the shortest path between the source vertex and all the other vertices can be computed using Dijkstra's algorithm (Dijkstra et al., 1959). However, since Dijkstra's algorithm updates rely on the edges of the mesh, it computes only a rough approximation of the correct geodesic distances (Figure 5.3c). Therefore, I utilize the Fast Marching Method (FMM) (Sethian, 1998; Kimmel and Sethian, 1998a) to provide a fast and high-quality approximation of the true geodesic distance field in $O(V \log V)$ time. The FMM is essentially a wavefront propagation algorithm, which relies on an upwind, finite difference approximation to the gradient and a resulting causality relationship, resulting in a Dijkstra-like algorithm but with update steps for triangles instead of edges (Figures 5.3a and b).

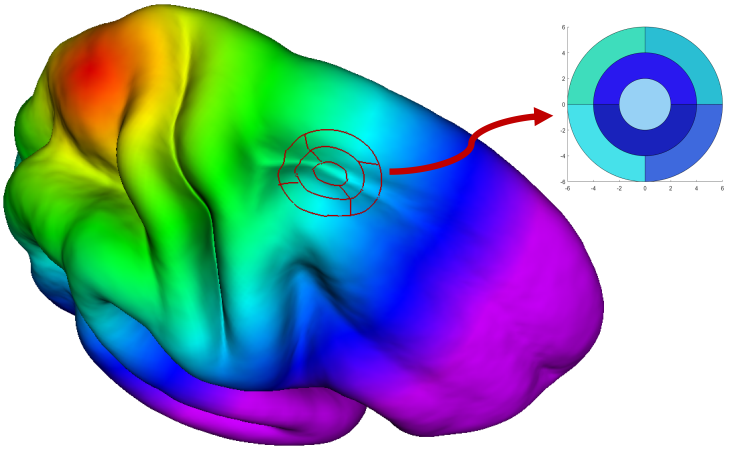


(a) First-order FMM (b) Zero-order FMM (c) Zero-order Dijkstra

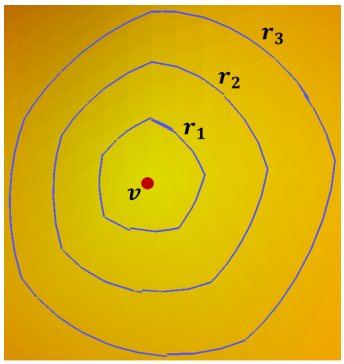
Figure 5.3: Geodesic paths created using Dijkstra and Fast Marching Method (FMM) to compute the shortest path between two surface vertices. Paths can be either generated so that they traverse between mesh vertices (first-order or linear interpolation) or restricted to mesh vertices (zero-order or nearest neighbor interpolation). Observe that a first-order FMM provides a better approximation of the geodesic path as compared to a (b) zero-order FMM and a (c) zero-order Dijkstra method.

Geodesic Grid: Surface kernels are then defined based on a locally constructed geodesic grid to extract surface patches before applying a cascade of filters and non-linear operators (Figure 5.4a). The parameters of these filters are optimized to minimize a task-dependent loss function. The radial coordinates of the geodesic grid relative to a vertex v are defined as the level curves of the geodesic

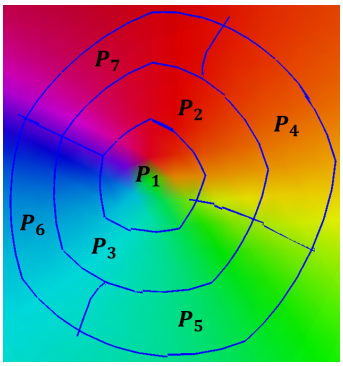
distance function $\phi = r_i$, where r_i is the radius of the i^{th} geodesic ring (Figure 5.4b). A grid is now formed at v (Figures 5.4c and d). Angular partitions are then created by splitting the curve of the geodesic distance function $\phi = r_i$ into segments of equal length and propagating directions from these points inward along the gradient $\nabla_M \phi$. This geodesic grid at v is formed from partitions or regions P_k with a similar surface area, and the order of these partitions is established using the spherical parameterization of the surface (Figure 5.4c), which is typically provided by surface reconstruction pipelines (see Chapter 2, Section 2.2). Finally, the feature values over the points in each partition P_k are summarized using a specified number of sampled points m_{kj} (Figure 5.4d).



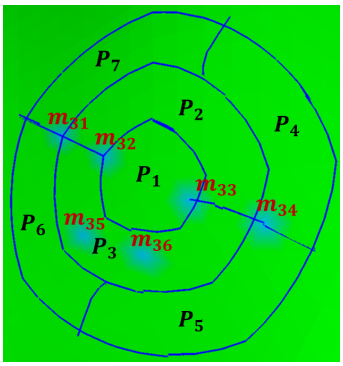
(a) Local Surface Patch



(b) Radial Coordinates



(c) Angular Partitions



(d) Regional Summaries

Figure 5.4: (a) A local geodesic kernel created on cortical surfaces to extract local surface patches. (b) Geodesic distance level curves. (c) Angular partitions created using inward ray shooting and ordered by the surface spherical parameterization. (d) Regional feature summaries are created by averaging six sampled feature values (four measurements at the region corners and two measurements that straddle each region). Showing here the measurements locations for P_3 as an example.

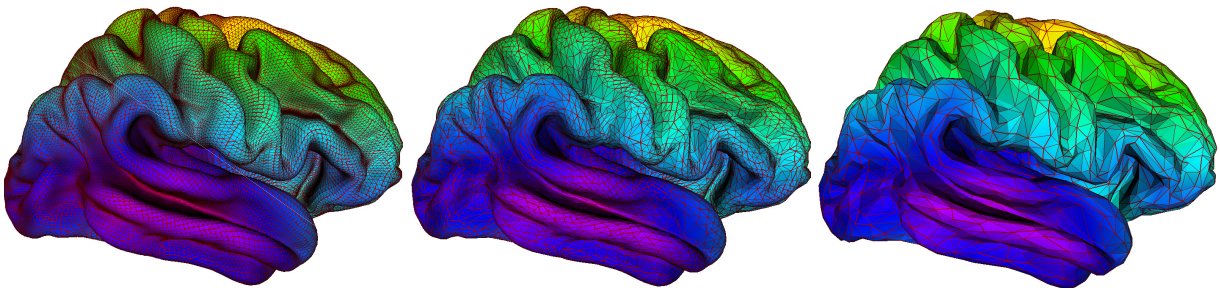
Model Complexity: In the training phase, N geodesic kernels are used to produce the corresponding feature maps $f_1, \ldots, f_N$. The number of filters used in convolutional layers typically increases with the depth of the network, increasing the number of resulting feature maps and associated trainable parameters (weights). This large number of parameters would require a large training dataset (typically not available in medical imaging) to ensure the generalizability of deep learning models. To address such a problem, the model complexity is controlled using bottleneck layers that are used to produce a fused feature map f by selecting the maximum feature response for each v on the surface from $f_1, \ldots, f_N$.

5.2.2 Surface Subsampling

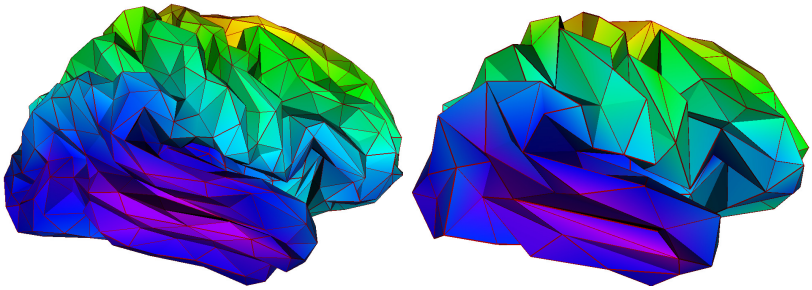
To further reduce the number of training parameters and to learn an abstracted form of the representation, the feature dimensionality is reduced via surface subsampling. This is done by reducing the number of triangles in the input triangle mesh, forming a good approximation to the original geometry. A surface simplification algorithm based on repeated edge collapses is employed (Garland and Heckbert, 1997; Hoppe, 1999). Edges are placed in a priority queue based on a quadric error measure Q that is associated with each vertex $v : (p, s)$ of the surface, where p is the geometric position and s is a set of attribute scalars. Let $v' : (p', s')$ be the projection of v onto the associated affine subspace, and let Q be given by

$$Q(v) = ||p - p'||^2 + \lambda||s - s'||^2, \tag{5.3}$$

the weighted sum of the geometric distance error $||p-p'||^2$ and the attribute deviation error $||s-s'||^2$. As edges are deleted, the quadric error measures associated with the two endpoints of the edge are summed, and an optimal collapse point is computed. This edge collapse process is repeated until the desired surface resolution level is reached or topological constraints are violated. In this paper, Gaussian and mean curvatures along with the geodesic distances are used as attributes in the quadric error measure to ensure the uniformity and integrity of the resampled surfaces. Figure 5.5 shows a resampled surface using the proposed surface simplification layer.



(a) 40K Vertices (b) 10K Vertices (c) 2.5K Vertices



(d) 0.62K Vertices (e) 0.16K Vertices

Figure 5.5: An example of a resampled surface using repeated edge collapses, preserving shape and geodesic distances at (a)-(e) different surface resolution levels (subsampling factor of 4). Surfaces were colored based on geodesic distances computed from the seed vertex shown in Figure 5.4.

Surface-CNN: The proposed convolution and subsampling techniques enable the development of network architectures capable of extending convolutional learning to non-Euclidean surfaces. The proposed Surface-CNN allows for learning hierarchical task-specific features on non-Euclidean manifolds for applications such as shape classification. The proposed model is very generic and flexible, and can be made arbitrarily complex by stacking multiple surface convolution blocks.

5.3 Experimental Results

The usability of the proposed Surface-CNN framework is demonstrated in two applications. First, Surface-CNN is used to discriminate Alzheimer's Disease patients from healthy controls using measurements extracted from cortical brain surfaces. Second, Surface-CNN is used to predict ASD diagnosis at 24 months using measures derived from 6-month-old subcortical brain surfaces.

5.3.1 Alzheimer's Disease Classification

5.3.1.1 Materials

ADNI3 Dataset: The Proposed Surface-CNN architecture was applied to a subset of the structural MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`adni.loni.usc.edu`). Notably, the first T1w scans of AD patients and Control (CN) patients were included from the ADNI3 dataset, which was acquired at 59 research centers in the United States and Canada. This resulted in a dataset of 86 subjects of two classes: 39 AD patients (25 males, 14 females; 75.1$\pm$ 8.2 years) and 47 CN patients (23 males, 24 females; 73.2$\pm$ 5.7 years). Note that the diagnostic categories used here were provided with the imaging dataset based on clinical assessments outlined in the ADNI3 protocol (Weiner et al., 2017). The CN group did not include any borderline cases (i.e., patients with Significant Memory Concern (SMC) were not included). Using 3T Siemens Prisma, Verio, Skyra, TrioTim scanners, the T1w structural MRI were acquired using the 3D MP-RAGE pulse sequence (208 sagittal slices, TR = 2300 ms, TE = 2.95 or 2.98 ms, TI=900 ms, flip angle = 9°, resolution = 1.0 $\times$ 1.0 $\times$ 1.0 or 1.2 mm^3, field of view = 256 $\times$ 240).

5.3.1.2 Data Preprocessing

For each subject, the T1w MRI was processed using FreeSurfer pipeline (Fischl, 2012) for intensity inhomogeneity correction, skull stripping and cerebellum removal, tissue segmentation, and cortical surface reconstruction. The resampled cortical surfaces consisted of 81,920 high-resolution triangles (40,962 vertices) per hemisphere. Vertex-wise CT and SA measures were extracted from the reconstructed cortical surfaces and then were summarized using the Desikan-Killiany parcellation atlas (70 ROIs) (Desikan et al., 2006) as well as the Destrieux parcellation atlas (148 ROIs) (Destrieux et al., 2010). All features were standardized using Z-score normalization, which rescales the features so that they will have zero mean and unit variance.

5.3.1.3 Network Architecture and Training

As shown in Figure 5.6, geodesic convolution was applied first on features that live on a high-resolution average surface (40,962 vertices) using 64 geodesic kernels (2 mm step size for each geodesic ring) with trainable weights and non-linearity in the form of the Hyperbolic Tangent activation function. The newly generated features were then summarized into a single feature map using a maximum feature pooling layer. A surface resampling followed this with a factor of 4, where feature values are mapped to the lower resolution surface by averaging information from the k-nearest neighbors, with $k = 5$ was chosen based on the neighborhood structure commonly encountered in the average surfaces. This process was repeated for the new lower resolution surface but with proportionally (based on average edge length) increased step size. After reaching a reasonable number of features, fully connected layers were applied with ReLU neurons.

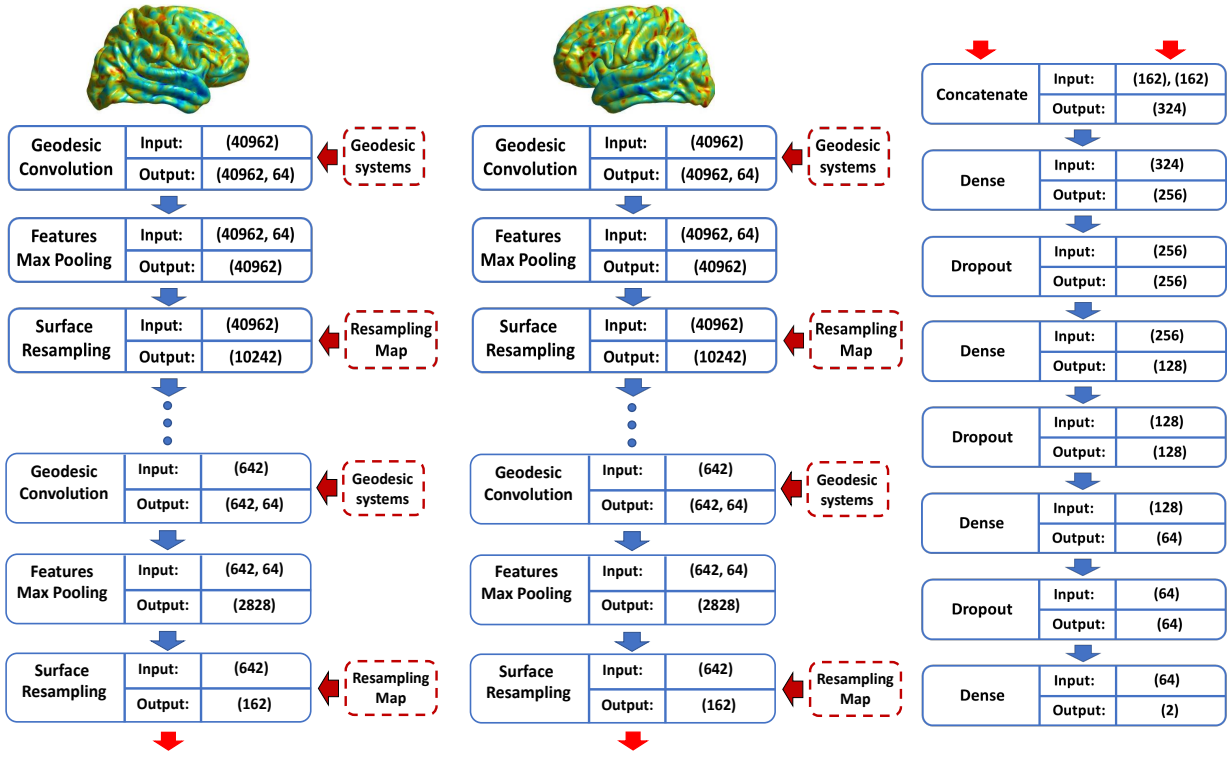


Figure 5.6: The proposed Surface-CNN network architecture for learning on cortical surfaces. 3 surface convolutional blocks are applied to each hemisphere in parallel, followed by fully connected and dropout layers. Note that the geodesic grid at each surface resolution level and the resampling maps were computed offline and fed to the network during training (dashed red boxes).

To overcome the overfitting problem associated with deep neural nets with many parameters, dropout techniques were used after each hidden layer (dropout rate of 0.2). Max-norm regularization was also used to constrain kernel weights, which is useful in deep neural networks. The proposed Surface-CNN was compiled using the Adam optimization algorithm (Kingma and Ba, 2014) and the binary cross-entropy loss function. The network was trained using the back-propagation algorithm for 500 epochs and a batch size of 1. To handle the classes' imbalance in the training phase, the SMOTE method (Chawla et al., 2002) was used to oversample the minority class to match the number of samples in the majority class with 5 nearest neighbors used to construct these synthetic samples. Only the non-synthetic minority class samples were used in the testing phase.

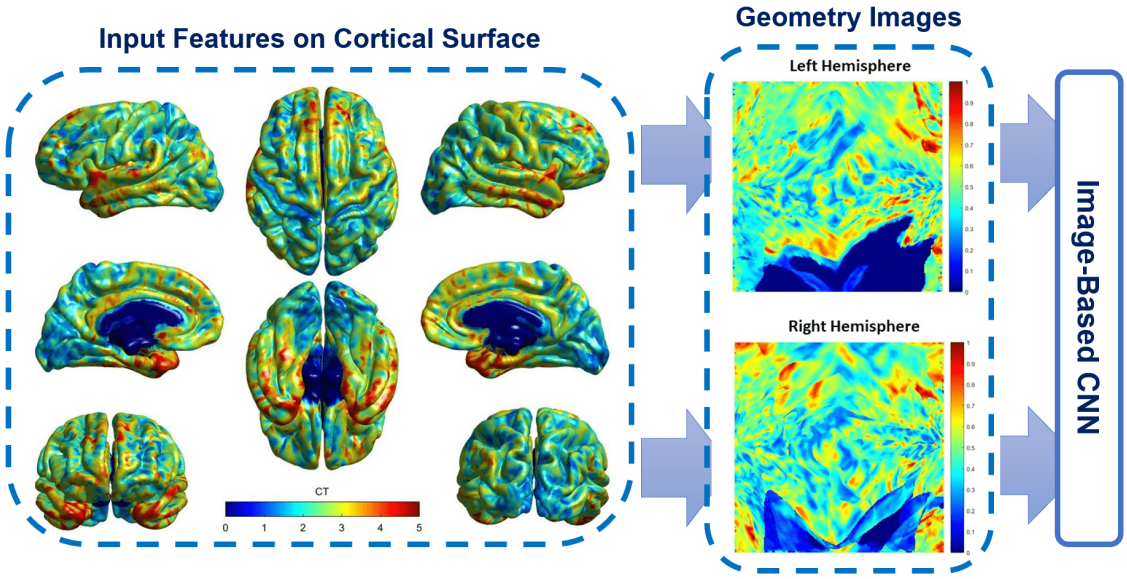


Figure 5.7: An alternative to the proposed Surface-CNN architecture on cortical surfaces. Surfaces can be re-meshed onto a completely regular structure called a geometry image (Gu et al., 2002), which allows learning using conventional CNN architectures.

5.3.1.4 Performance Evaluation

The proposed techniques were tested by applying them to each feature extracted (CT and SA) to classify AD subjects from CN subjects. The network performance was evaluated using 10-fold cross-validation using the number of TP, FP, TN, and FN as well as the mean Area Under the ROC Curve (AUC), ACC, SEN, and SPC summary scores. To confirm the proposed Surface-

CNN's ability to extract efficient representations from the input high-dimensional feature space, I compare the Surface-CNN performance against a similar DNN with only fully connected layers that were trained using features summaries provided by Desikan-Killiany (Desikan et al., 2006) and Destrieux (Destrieux et al., 2010) parcellation atlases. Also, traditional machine learning methods were explored to perform the required dimensionality reduction of cortical measurements. Sparse coding and dictionary learning (Mairal et al., 2009) has been previously proposed to learn an over-complete set of basis vectors (also called dictionary) to represent input vectors efficiently and concisely. Hence, I also compare Surface-CNN against a DNN that is trained using encoded representations from a surface-based sparse coding and dictionary learning feature reduction. Finally, rather than applying CNNs directly on the surface, one can also define a mapping to bring the data into space where conventional CNN methods can be applied. For example, closed surface data can be mapped onto a regular 2D image structure called a geometry image (Gu et al., 2002) (see Figure 5.7). This allows the straightforward use of image-based CNN to learn directly from the high-dimensional input features without the complexities of the above-described surface CNN.

Table 5.1 provides a summary of the classification results obtained from applying the Surface-CNN and other competing methods to the CT and SA extracted from the ADNI3 dataset. The proposed Surface-CNN performs notably better than competing methods, according to every performance metric. Experiments with geometry images of features living on surfaces show that the performance level achieved by such a simplified convolutional approach is still significantly inferior to convolutional learning directly on the manifold. Such a performance drop can be attributed to the non-isometric nature of the mapping into the 2D image space, which results in a surface-specific distortion of the surface feature maps (see Figure 5.8). To show the ability of the proposed CNN extension to provide an insight on which brain regions are contributing more in separating the two classes, Figure 5.9 shows the average features learned after applying the first block of convolution, pooling, and resampling layers. Figure 5.10 show those regions showing statistically significant group differences. For an example of such brain areas, Figure 5.9 highlights a mean response difference between AD and CN groups in a region in the medial limbic cortex called the entorhinal

cortex, which has been reported to be atrophied in AD (Du et al., 2001) patients. The ability clearly confirms the potential of the proposed Surface-CNN in dynamically learning clinically relevant brain regions that are specific to the training dataset and the investigated disease.

		TP	FP	TN	FN	AUC	ACC (%)	SEN (%)	SPC (%)
Cortical Thickness	**Proposed Surface-CNN**	**36**	**4**	**43**	**3**	**0.94**	**91.86**	**92.31**	**91.49**
	Desikan-Killiany Atlas + DNN	31	8	39	8	0.84	81.40	79.49	82.98
	Destrieux Atlas + DNN	26	6	41	13	0.87	77.91	66.67	87.23
	Sparse Coding + DNN	33	6	41	6	0.90	86.05	84.62	87.23
	Geometry Images + CNN	28	8	39	11	0.80	77.91	71.79	82.98
Surface Area	**Proposed Surface-CNN**	**33**	**6**	**41**	**6**	**0.87**	**86.05**	**84.62**	**87.23**
	Desikan-Killiany Atlas + DNN	23	12	35	16	0.67	67.44	58.97	74.47
	Destrieux Atlas + DNN	24	16	31	15	0.67	63.95	61.54	65.96
	Sparse Coding + DNN	22	14	33	17	0.73	63.95	56.41	70.21
	Geometry Images + CNN	22	18	29	17	0.62	59.30	56.41	61.70

Table 5.1: Summary of the AD vs. CN classification results in ADNI3 dataset using the number of TP, FP, TN, and FN as well as the mean AUC, ACC, SEN, and SPC summary metrics.

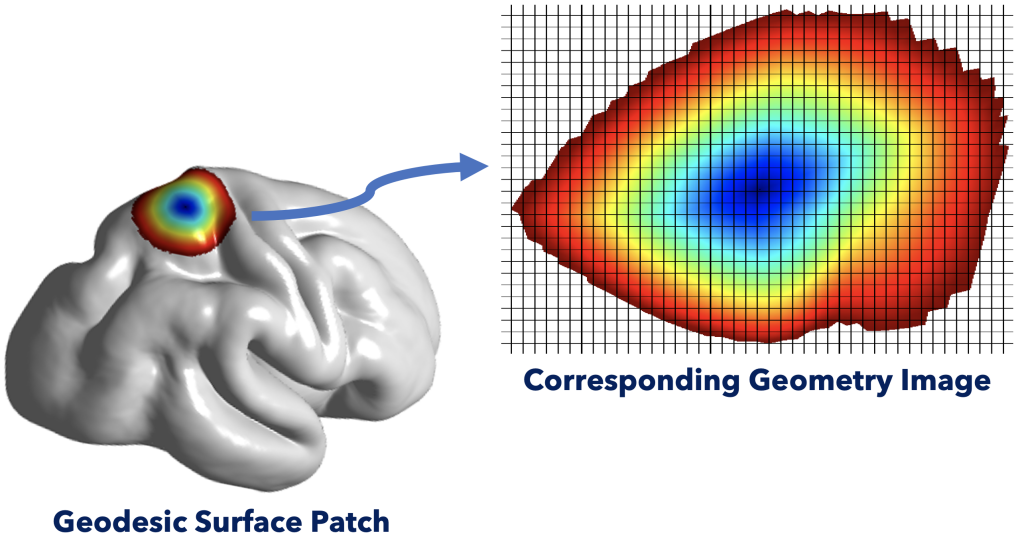


Figure 5.8: Illustration of the distortion introduced by geometry images. A geodesic surface patch (15mm) was mapped to a regular 2D grid in a way that does not guarantee the preservation of distances and angles. Hence, standard (Euclidean) CNN architectures applied to such 2D geometry images would result in a performance degradation because of the introduced mapping distortions.

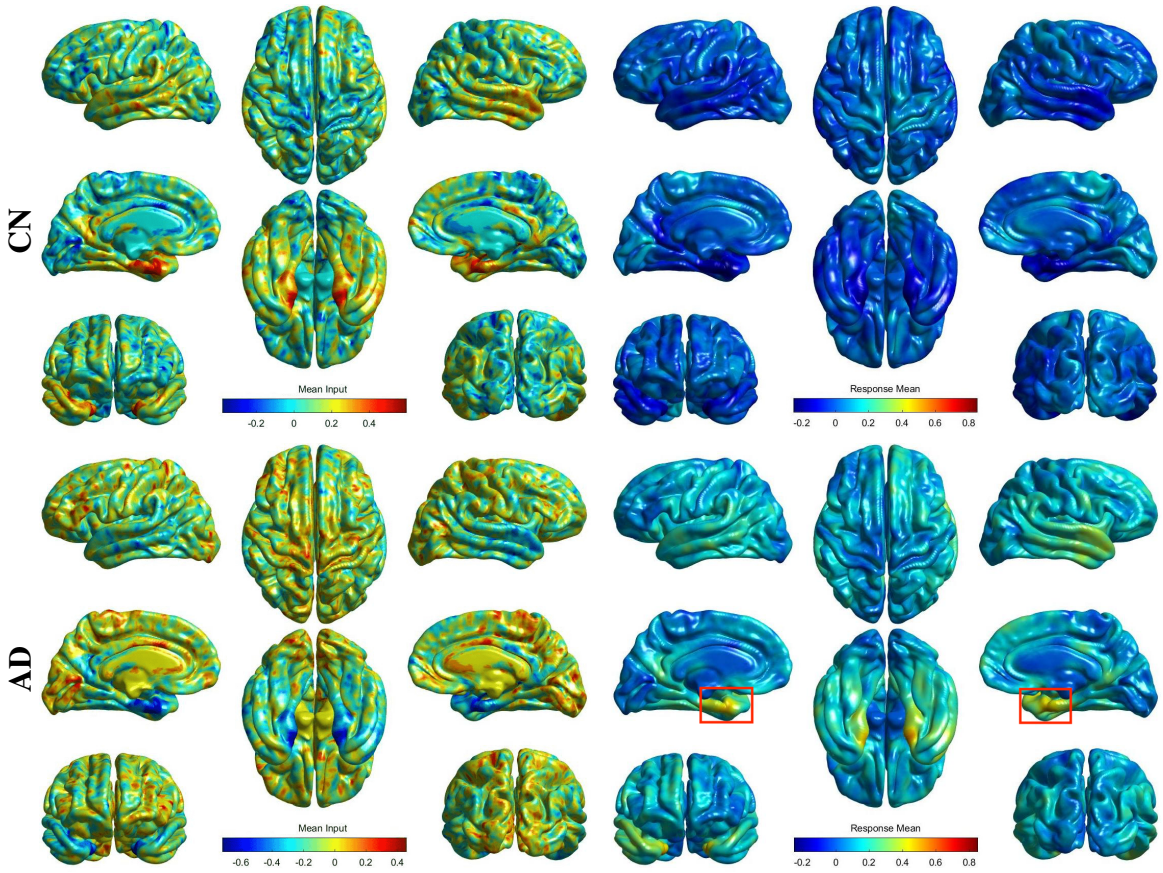


Figure 5.9: Visualization of the mean features learned by the proposed Surface-CNN for both CN (first row) and AD (second row). The first column shows the mean input CT while the second column shows the mean filters response after applying the first surface convolution block of layers.

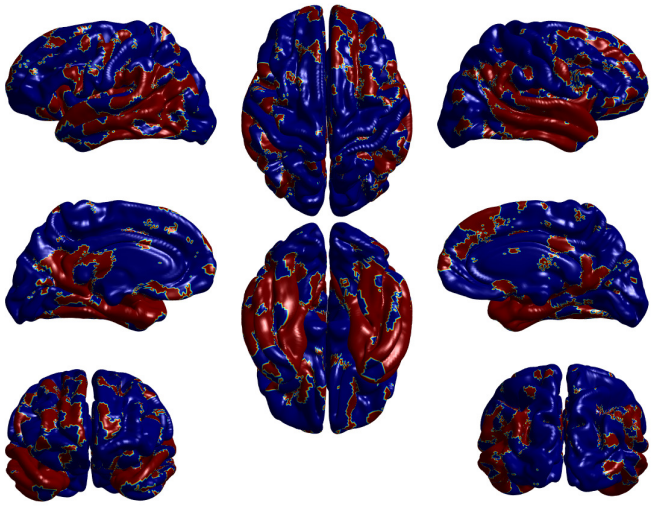


Figure 5.10: Uncorrected statistical t-test results on the first geodesic block output indicated that several brain regions (36% of the whole brain showing here in red) showed statistically significant (p-value $<$ 0.05) group differences.

5.3.2 Early Prediction of Autism

5.3.2.1 Materials

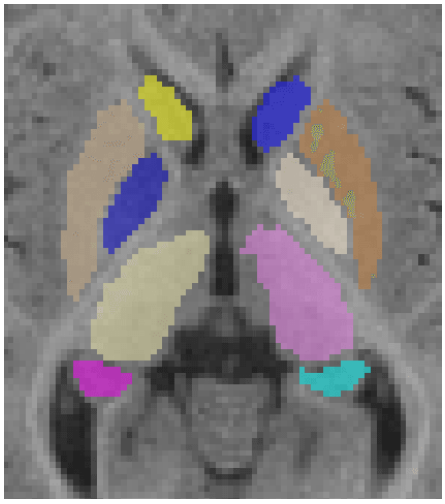
IBIS Dataset: To demonstrate the generalizability of the proposed Surface-CNN, it was used to investigate the early prediction of ASD using features extracted from 6-month-old's subcortical brain surfaces (rather than cortical surfaces). From the IBIS dataset, I included 151 infants at High-familial Risk (HR) for developing ASD, by virtue of having an older sibling with the disorder. Younger siblings of children with ASD demonstrate a high recurrence rate of ASD (20%), and therefore the prospective study of these high-risk infant siblings represents an ideal study design to identify early features of ASD before diagnosis. These infant participants were prospectively followed starting with a MRI scan at 6 months of age, until 24 months of age when they were clinically classified as either meeting criteria for ASD (HR-ASD; n=32; 19.9%; 28 males, 4 females; 6.7$\pm$ 0.7 months) or not (HR-Neg; n=121; 80.1%; 66 males, 55 females; 6.6$\pm$ 0.7 months). The MRI scans were acquired at 4 different sites (University of North Carolina at Chapel Hill, University of Washington at Seattle, Washington University at Saint Louis and Childrens Hospital of Philadelphia), each equipped with 3T Siemens Tim Trio scanners. The scan sessions included T1w (160 sagittal slices with TR=2400ms, TE=3.16ms, flip angle=8°, field of view 224 $\times$ 256) and T2w (160 sagittal slices with TR=3200ms, TE=499ms, flip angle=120°, field of view 256 $\times$ 256) MRI scans. All datasets possess the same spatial resolution of 1 $\times$ 1 $\times$ 1 mm^3.

5.3.2.2 Data Preprocessing

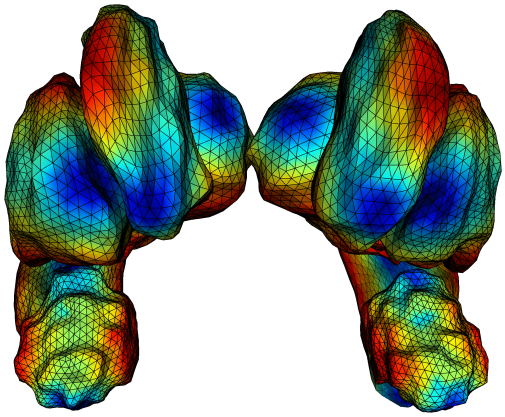
All T1w and T2w images were corrected for geometric distortions (Fonov et al., 2010) and intensity non-uniformity (Tustison et al., 2010). T1w and T2w images were rigidly transformed to a prior pediatric 1-year-old atlas in stereotaxic space. The brain mask necessary to perform skull stripping was performed using a multi-atlas approach that combines multiple candidate brain masks obtained via ANTs deformable registration (Avants et al., 2011) of a prior set of atlases (FSL-BET (Smith, 2002) atlas, two in-house prior atlases, and two atlases of CIVET pipeline (Kim

et al., 2005)). The fusion of the candidate brain masks was performed via a straightforward majority vote. Segmentation of subcortical brain structures was performed using a graph-based multi-atlas approach implemented in the AutoSeg toolkit (Wang et al., 2014). First, all atlases and subject MRI were paired and co-registered using ANTS deformable registration (Avants et al., 2011). Then, a directed graph was constructed with weights based on intensity and shape similarity between all atlases and the subject MRI. The atlases were then clustered based on the shortest path from each atlas to each subject MRI. The final segmentation was produced by fusing the propagated label files of the neighboring atlases via weighted majority voting (see Figure 5.11a).

Surface meshes of the segmented subcortical structures were then generated using the SPHARM-PDM shape analysis tool (Styner et al., 2006). Mainly, object segmentations first represented by binary labels were converted into surfaces with corresponding surface sampling via spherical harmonic description (SPHARM) based on area-preserving, distortion-minimizing mapping. SPHARM models were then sampled into point distribution models (PDM) or triangulated surfaces using a uniform icosahedron subdivision (1002 vertices). The correspondence problem was solved by aligning each SPHARM-PDM surface to the mean surface created from all subjects in the parametrization space using the centroid and longitudinal axes of a first-order ellipsoid. Vertex-wise shape index and curvedness information (Koenderink and Van Doorn, 1992) were computed from the obtained SPHARM-PDM coefficients (Styner et al., 2006). Local medial thickness information (Paniagua et al., 2013) was also extracted based on mean latitude axis that was also directly computed from the spherical parameterization (see Figure 5.11b)



(a) Subcortical Segmentation



(b) SPHARM-PDM Surfaces

Figure 5.11: An example of the segmented subcortical structures with corresponding surface meshes. (a) 12 subcortical structures (left and right amygdala, caudate, globus pallidus, hippocampus, putamen, thalamus) were segmented using a graph-based multi-atlas approach implemented in the AutoSeg toolkit (Wang et al., 2014). Only 10 structures are shown here, as it was not possible to show all structures on a single axial slice. (b) Binary segmentations of different subcortical structures were converted into a corresponding SPHARM PDM (Styner et al., 2006) surface meshes. Surfaces were colored based on local distances to the structure's medial axis (e.g., thickness).

5.3.2.3 Network Architecture and Training

As shown in Figure 5.12, the architecture of the proposed Surface-CNN for learning on subcortical surfaces is similar to that applied for learning on cortical surfaces (Figure 5.6). Particularly, 3 surface convolution blocks (surface resampling with a factor of 2) were applied to each subcortical structure in parallel, followed by classification and dropout layers. 64 geodesic kernels were used at each surface resolution level with an adaptive geodesic grid (initial step size of 2mm at highest surface resolution). Hyperbolic Tangent non-linearity was used as an activation function for surface convolution blocks while ReLU activation was used for the fully connected layers. Dropout layers (dropout rate of 0.2) and Max-norm regularization were used to avoid model overfitting. The network was trained with a batch size of 1 of using Adam optimization (Kingma and Ba, 2014) and a binary cross-entropy loss function for 500 epochs. Classes were balanced during training using SMOTE (Chawla et al., 2002) oversampling based on 5 nearest neighbors minority samples.

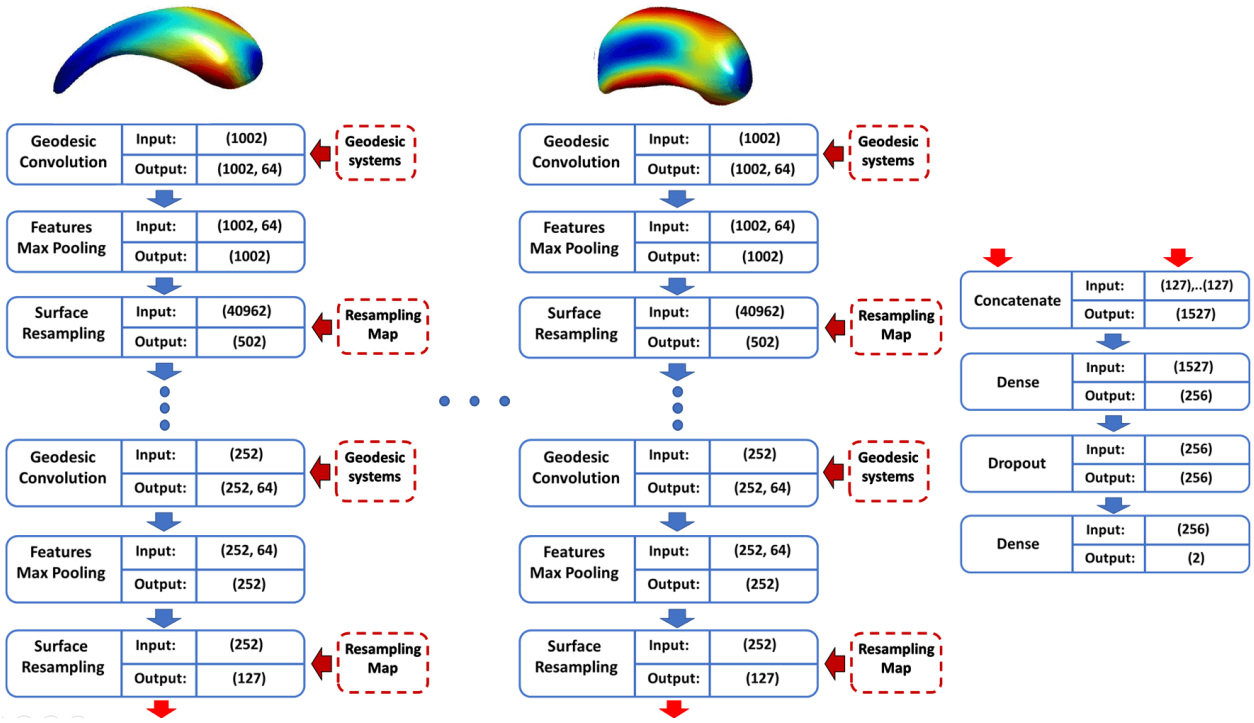


Figure 5.12: The proposed Surface-CNN network architecture for learning on subcortical surfaces. 3 surface convolutional blocks are applied to each subcortical structure in parallel, followed by fully connected and dropout layers. The geodesic grid at each surface resolution level and the resampling maps were computed offline and fed to the network during training (dashed red boxes).

5.3.2.4 Performance Evaluation

The proposed Surface-CNN network was tested by applying it to the input features, namely shape index, curvedness, and thickness measures computed at each vertex of 6-month-old subcortical surfaces to discriminate HR-ASD subjects from HR-Neg subjects. The network performance was evaluated using 10-fold cross-validation via the TP, FP, TN, FN, AUC, ACC, SEN, and SPC scores. The Surface-CNN performance was also compared against an image-based CNN that was trained using geometry images (Gu et al., 2002) resulting from mapping the investigated features to a regular 2D image grid. Table 5.2 provides a summary of the prediction results obtained from applying the Surface-CNN and the image-CNN approach to the 6-month-old subcortical features extracted from the IBIS dataset. Similar to the previously obtained results from learning on cortical surfaces, the proposed Surface-CNN performs better than the combined geometry images and

image-CNN approach, which further confirms the need for learning methods that operate directly on the extracted manifold. However, the performance of the proposed Surface-CNN does not reach clinically acceptable levels (low sensitivity), which could be attributed to the lack of discriminative ability of the extracted features.

		TP	FP	TN	FN	AUC	ACC (%)	SEN (%)	SPC (%)
Thickness	**Proposed Surface-CNN**	**11**	**12**	**109**	**21**	**0.75**	**78.02**	**34.69**	**90.22**
	Geometry Images + CNN	9	21	100	23	0.58	70.85	28.57	82.75
Shape Index	**Proposed Surface-CNN**	**13**	**11**	**110**	**19**	**0.77**	**79.82**	**40.81**	**90.80**
	Geometry Images + CNN	4	15	106	29	0.61	70.85	12.24	87.35
Curvedness	**Proposed Surface-CNN**	**14**	**15**	**106**	**17**	**0.76**	**78.47**	**44.89**	**87.93**
	Geometry Images + CNN	4	15	106	29	0.58	70.85	12.24	87.35

Table 5.2: Summary of the ASD prediction results in IBIS dataset using the number of TP, FP, TN, and FN as well as the mean AUC, ACC, SEN, and SPC summary metrics.

5.4 Summary

In this chapter, I proposed a novel data-driven generalization of CNNs for use on non-Euclidean manifolds, such as cortical and subcortical surfaces. The proposed CNN extension relies on adaptive surface kernels that are defined using a locally constructed geodesic grid. A surface convolution block is defined using novel surface convolution, pooling, and resampling layers, which allow for the construction of a Surface-CNN capable of directly learning on non-Euclidean surfaces without the need for a separate dimensionality reduction step. The proposed Surface-CNN learns the most powerful features and brain regions from the extracted large dimensional feature space, thus creating a new feature space in which the dimensionality is reduced and feature distributions are better separated. The proposed CNN extension applies to various kinds of clinical applications that involve learning from high-dimensional features living on non-Euclidean manifolds.

The high performance of the proposed techniques was first demonstrated in classifying AD patients from CN patients using cortical measurements living on cortical surfaces. The proposed

Surface-CNN framework outperformed competing methods employing a variety of parcellation atlases or using traditional machine learning methods such as sparse coding to perform the required feature reduction before applying deep learning models. Surface-CNN also demonstrated superior performance as compared to an alternative approach where an image-based CNN was applied to cortical features mapped to a regular 2D image grid, thereby confirming the need for convolutional learning directly on the manifold. Moreover, Surface-CNN showed an ability to transform high-dimensional features into a compact high-level representation in a way that highlights critical surface regions affected by the investigated disease. Surface-CNN generalizability was also demonstrated in predicting ASD diagnosis at 24 months using measures derived from 6-month-old subcortical brain surfaces. Though not clinically significant, the obtained results by the proposed Surface-CNN further confirmed the need for learning methods operating directly on the data manifold.

CHAPTER 6: ROBUST MULTI-TASK MULTI-MODAL DEEP LEARNING

6.1 Overview

Deep learning algorithms and in particular, those using CNNs have shown tremendous success in MRI image analysis applications, including image segmentation (Akkus et al., 2017), classification (Hoo-Chang et al., 2016), and regression (Yang et al., 2017) tasks. An effective deep learning algorithm needs to be accurate in order to able to detect subtle changes in the volumes and shapes of extracted brain structures in patients with neurodevelopmental disorders (Hazlett et al., 2017) or neurodegenerative diseases (Moradi et al., 2015). Moreover, such deep learning tools need to be fast and memory-efficient to support large multimodal datasets that may be adopted in real-time applications. MRI-based deep learning algorithms also need to be robust to variations in image intensities and contrasts originating from using different scanner models, acquisition parameters, and image resolutions (see Figure 6.1). The robustness of deep learning models is a major concern as multi-center studies are increasingly implemented while standardization techniques are still unable to achieve perfect data harmonization (Thompson et al., 2014).

MRI-based CNNs Robustness: With recent advancements of powerful supervised CNNs, it is now possible to produce excellent accuracy results within seconds in a variety of medical imaging tasks (Litjens et al., 2017), including image segmentation (Akkus et al., 2017), image registration (Yang et al., 2017), image fusion (Suk et al., 2014), lesion detection (Kooi et al., 2017), and computer-aided diagnosis (Hoo-Chang et al., 2016). However, the robustness of CNNs is still an open question as it was shown that such models could be unstable to small perturbations in the input data (Elsayed et al., 2018). The generalization of CNNs depends on the availability of sufficient labeled data that represents well the variability expected to be seen at testing time. Due to the costly and time-consuming nature of collecting labeled MRI datasets, current MRI-based

CNNs are typically trained only on limited datasets that were acquired using the same scanner and with a homogeneous MRI acquisition protocol. Since variability in acquisition protocols highly depends on the performance site, CNNs that are trained on data gathered from only a small number of sites might not be representative of the imaging variability that can be expected at test time in a broadly deployed application. Hence, there is still a need to design deep learning models and training schemes to address the current MRI-based CNNs robustness problem.

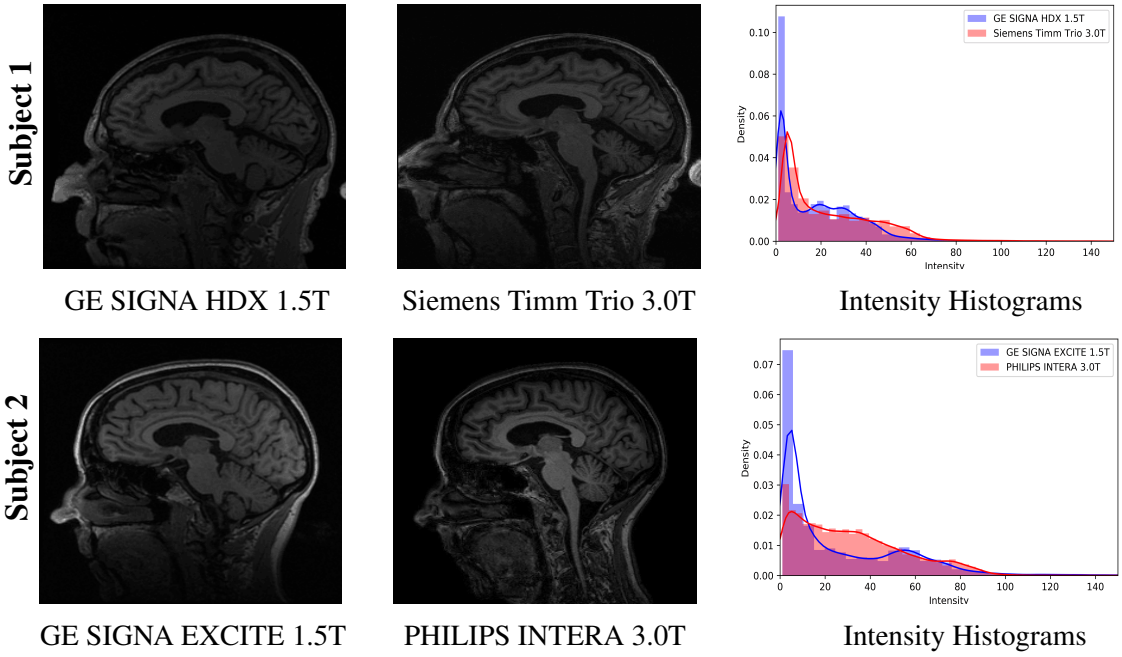


Figure 6.1: Image slices (left and middle column) and corresponding histograms (right column) of T1w MRIs obtained from using different scanners and using similar acquisition parameters. Though high-level information looks similar, there is considerable intensity and contrast variability that affects decisions made by image analysis algorithms. Note that using the same scanner model and varying acquisition settings (e.g., field strength, coil type, acquisition plan), pulse sequence parameters (e.g., TE, TR, TI, flip angle), or geometric information (e.g., filed of view, resolution) would result in similar intensity distribution differences (see Section 6.3 for more details).

Scan Resilient CNNs: Recent efforts have been made to develop CNNs that are resilient to variation in the input T1w MRI acquisition. Such efforts were mainly focused on the application of whole-brain segmentation from a T1w MRI as such a step is critical in neuroimaging processing pipelines (Fischl, 2012). For example, a lifelong learning approach for segmenting brain MRI consistently across multi-scanner and multi-protocol data was proposed (Karani et al., 2018). This

approach relies on a single CNN that adds a new set of batch normalization parameters as new scanners or acquisition protocols are encountered. However, to learn these added parameters, such an approach requires the availability of a few labeled images from the added domain; thus limits its application in time-sensitive realistic settings. Alternatively, Jog et al. (Jog et al., 2019) proposed a pulse sequence adaptive CNN approach that relies on an augmentation scheme to generate a wide range of synthetic training data to build a robust CNN model. Synthetic variants of the training images are created by applying pulse sequence approximated forward equations, with uniformly sampled parameters, to the training MR relaxation parameter maps. Despite its ability to produce accurate segmentation across different pulse sequences without the need for additional labeled data, this approach is limited by the need for training MR relaxation parameter maps. Moreover, the success of the utilized approximated forward models is restricted to specific pulse sequences.

Current Limitations: Current scan resilient CNNs also fail to utilize a lot of relevant non-imaging information (i.e., meta-data) that is also often available. Examples of that include patient demographic information (e.g., sex, age), scanner information (manufacturer model, field strength), geometric image parameters (e.g., the field of view, image resolution), and scan parameters (e.g., MRI sequence description). In the image synthesis field, a common technique for mixing imaging and non-imaging data consists of expanding scalar meta-data into constant images, and then concatenating them as additional channels to the image data (Reed et al., 2016). Because such an approach is computationally inefficient, the meta-data is typically only made available at the lowest possible resolution of the network. Since the availability of relevant meta-data at different scales was shown to be useful in style transfer (Karras et al., 2019); I suggest that such benefits may be extended to other tasks (e.g., image segmentation), driven by a potential increase in the representation capacity of such deep learning models. Current efforts to design robust CNNs were limited to MRI image segmentation, and there is a need to extend them to perform other tasks, including image classification and regression. In particular, developing a robust Multi-Task Learning (MTL) (Ruder, 2017) framework to learn multiple tasks simultaneously is desirable because of its ability to exploit the underlying similarities between correlated tasks. By allowing the

network to learn a common feature representation for multiple tasks, an improvement in the network generalization performance is gained. Additional MTL benefits include improvements in prediction accuracy and network scalability as well as a reduction in computational complexity (Ruder, 2017).

Proposed Multi-Task Multi-Modal Networks: To overcome the above problems of current MRI-based CNNs, I introduce the notion of a multi-modal block as a generalization of convolutional (Conv) or fully connected (FC) layers for learning from paired data (meta-data, images). For robust MRI tissue segmentation, multi-modal networks (M-Nets) are implemented as a combination of such multi-modal blocks with scanner information, MRI sequence parameters, geometrical information, and tissue information used as meta-data. By forcing the network to learn representations that work well in expectation over different settings, improvements are gained in the overall performance as well as in the ability to generalize to new acquisition protocols. I also extend the proposed M-Net architecture to multi-task learning (MM-Net) for the joint learning of segmentation, classification, and regression tasks in heterogeneous datasets. The proposed MM-Net is designed to model the coexisting modality and task heterogeneities, leading to further improvements in network performance and generalizability. My main contributions are summarized as follows:

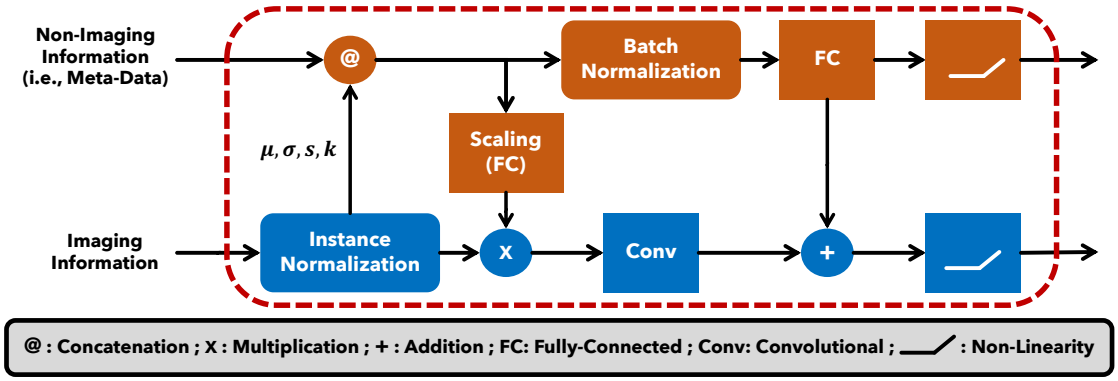
[1] A definition of a multi-modal block as a generalization of convolutional or fully connected layers for efficient representations learning from images and corresponding meta-data information.

[2] A description of multi-modal networks as a combination of such multi-modal blocks and other deep learning layers, allowing for the availability of relevant meta-data at different scales.

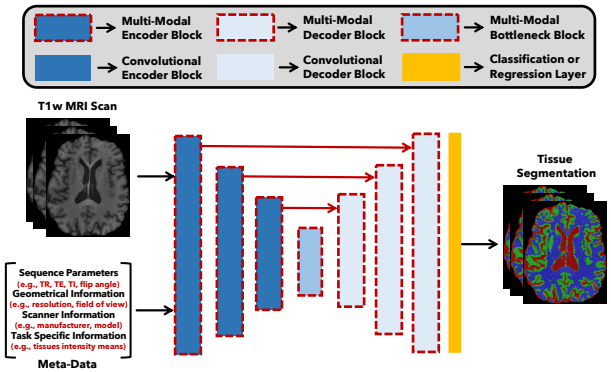
[3] A method of accurate, fast, automatic segmentation of brain structural MRI in a wide range of acquisition protocols using a multi-modal version of popular segmentation architectures.

[4] A framework of unified multi-task multi-modal learning for the robust joint learning of classification, regression, and segmentation tasks in multi-modal heterogeneous MRI datasets.

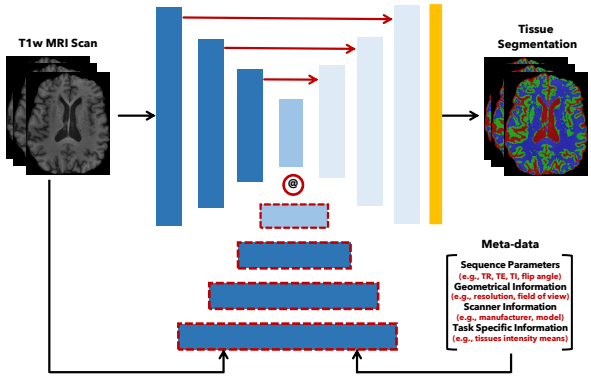
In the next section, I provide details of the proposed multi-modal block for joint processing of imaging and corresponding non-imaging (meta-data) information. I also describe how the proposed multi-modal blocks are combined in multi-modal networks for robust representation learning.



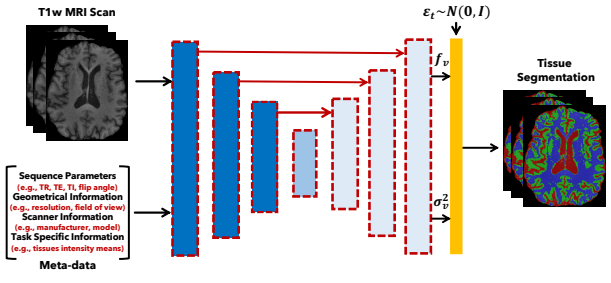
(a) The proposed Multi-Modal Block Design



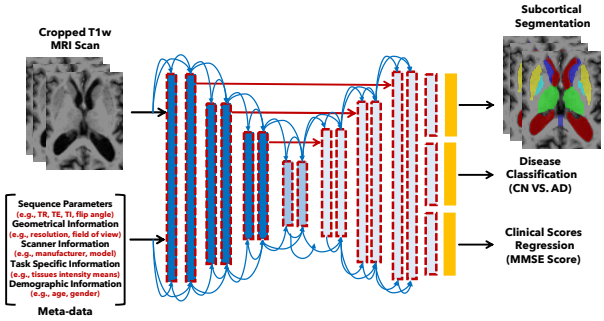
(b) Multi-Modal U-Net



(c) Conditional Multi-Modal U-Net



(d) Uncertainty Multi-Modal U-Net



(e) Multi-Task Multi-Modal Dense-U-Net

Figure 6.2: The proposed multi-modal networks, which utilizes new multi-modal blocks for the joint processing of imaging and non-imaging (i.e., meta-data) information. (a) A multi-modal block is defined using convolutional, fully connected, normalization, and non-linear layers, allowing for efficient information flow between imaging and non-imaging data. (b)-(d) The proposed multi-modal networks for the robust segmentation of brain tissues from a T1w MRI, where commonly available scan meta-data information is leveraged to improve the network performance and robustness. (e) The proposed unified multi-task multi-modal learning framework for predicting diagnostic outcomes as well as continuous scores in a disease-related domain from input MRI and associated meta-data (scan information and patient demographic data). The MRI subcortical segmentation task was additionally defined as an auxiliary task to improve the performance of main prediction tasks.

6.2 Methods

I propose a design for deep multi-modal networks as a combination of multi-modal blocks that all have paired data (meta-data, images) as inputs and outputs (Figure 6.2). In this section, I introduce such multi-modal blocks and describe their combinations for improving the robustness and overall performance of MRI segmentation tools as well as multi-task prediction frameworks.

6.2.1 Multi-Modal Learning Block

As shown in Figure 6.2a, multi-modal blocks are defined as a generalization of the convolutional or fully connected deep learning blocks composed of batch normalization, instance normalization, linear, and non-linearity layers. The instance normalization layer (Ulyanov et al., 2016) is extended to be used on the image data both as a normalization layer and as an embedding layer. In addition to normalizing the channels of given input image data, this layer also concatenates the computed channel means μ_i and standard deviations σ_i, as well as higher-order statistics (i.e., skewness s_i and kurtosis k_i) to the input meta-data. Mathematically, for an input image $x = [x_1,\, x_2, \cdots, x_n]$ with n channels with dimension D, such a layer operates on image x and meta-data m as follows

$$\begin{gathered}
\mu_i = \tfrac{1}{D}\sum_{d=1}^{D} x_i^d \quad ; \quad \sigma_i^2 = \tfrac{1}{D}\sum_{d=1}^{D}(x_i^d - \mu_i)^2 \quad ; \quad \hat{x_i} = \tfrac{x_i - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}} \\
s_i \propto \sum_{d=1}^{D}(x_i^d - \mu_i)^3/(\sigma_i^2 + \epsilon) \quad ; \quad k_i \propto \sum_{d=1}^{D}(x_i^d - \mu_i)^4/(\sigma_i^2 + \epsilon)^{\frac{3}{2}} \\
\hat{m} = [m,\, \mu_1, \cdots, \mu_n, \sigma_1, \cdots \sigma_n, s_1, \cdots, s_n, k_1, \cdots, k_n].
\end{gathered} \tag{6.1}$$

The information flow from the updated meta-data to the input image is accomplished using two mixing connections. First, a multiplicative connection is also established between meta-data and image data to give the layer the ability to invert the normalization performed by the instance norm. Second, an additive connection is utilized to sum the output of a fully-connected layer to the output of the convolutional layer; this is equivalent to the method of expanding and concatenating the meta-data to the images but is more efficient since the meta-data are processed as scalars.

The addition of such embedding layers at different levels will give the network capacity to learn to encode deep shift-covariant and shift-invariant features in an efficient way. Moreover, since the network acquires meta-data from the instance normalization layers, the designed model could be used even in the absence of input meta-data. This ability can be useful for learning disentangled representations, where one wants to learn scalar features from the input images.

6.2.2 Multi-Modal Segmentation Networks

Once the multi-modal block is designed, all the standard combination methods of deep learning can be extended to be deep multi-modal networks. As shown in Fig. 6.2b, a fully multi-modal network design based on a U-Net (Ronneberger et al., 2015) architecture is proposed to achieve the required MRI tissue segmentation task. The proposed architecture consists of a contracting path to capture context through compact representations, an expanding path that allows for accurate localization, and skip connections to enable the network to learn from fewer training images. However, unlike a U-Net, convolutional blocks are replaced by multi-modal blocks to allow for learning robust representations from input MRI and corresponding meta-data information. Moreover, pooling layers are only applied to the image data while meta-data are left untouched. Since the proposed fully multi-modal architecture can provide both scalar and image outputs, it can be naturally extended to multitask learning. Additionally, any of the network outputs can be ignored to obtain a network with a similar API to current classification or image-to-image deep networks.

As shown in Figure 6.2c, I also introduce a partially multi-modal architecture in which meta-data and images are passed through a multi-modal encoder before being concatenated at the bottleneck layer of a U-Net network. Despite the representation capacity loss, such an architecture is more memory efficient for larger models that utilize high-dimensional meta-data information. Finally, to handle noisy data, a fully multi-modal architecture is designed to model heteroscedastic aleatoric uncertainty. Unlike epistemic uncertainty, modeling aleatoric uncertainty is essential as it can not be explained away with a large training dataset (Kendall and Gal, 2017). As shown in Figure 6.2d, in addition to a vector of unaries f_v predicted by the network for each voxel v, a measure of aleatoric

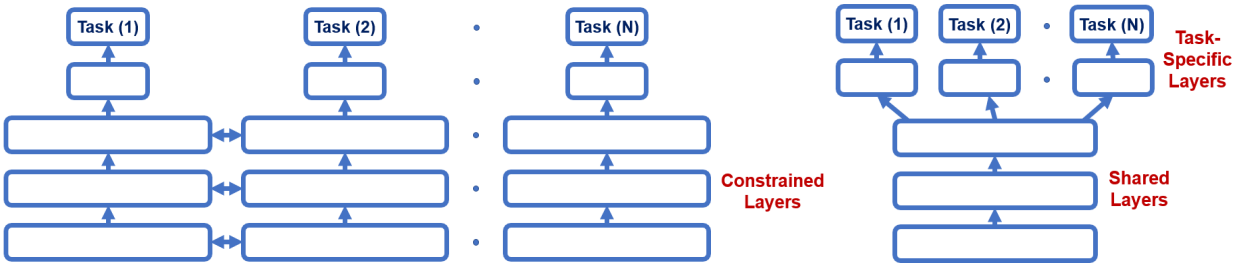
uncertainty given by the variance σ_v^2 is also produced. Monte Carlo integration is then used to sample unaries from the placed Gaussian prior through the softmax function. The network is trained using the following stochastic loss:

$$\hat{y}_{v,t} = f_v + \sigma_v \varepsilon_t, \quad \varepsilon_t \sim N(0,1)$$
$$L_y = \sum_v \log \tfrac{1}{T} \sum_T \exp(\hat{y}_{v,t,c} - \log \sum_{c'} \exp \hat{y}_{v,t,c'}) \tag{6.2}$$

where $y_{v,t,c'}$ is the c' element in the logit vector $y_{v,t}$.

6.2.3 Multi-Task Multi-Modal Network

Current deep learning prediction frameworks need to move beyond categorical diagnostic outcomes to predict continuous dimensions of disease-related features. This can be accomplished using an MTL framework, which is typically performed in DNNs with either soft or hard parameter sharing of hidden layers. A soft parameter sharing (Figure 6.3a) allows each task to have its model and parameters, with regularization techniques (e.g., L_2 norm (Duong et al., 2015) or trace norm (Yang and Hospedales, 2016)) employed to keep the model's parameters similar. On the other hand, in hard parameter sharing (Figure 6.3b) the hidden layers are shared between all tasks while several output layers are used for each task separately (Caruna, 1993).



(a) Soft Parameter Sharing (b) Hard Parameter Sharing

Figure 6.3: Conventional approaches to performing multi-task learning in deep neural networks, namely (a) soft or (b) hard parameter sharing of hidden layers. A hard parameter sharing strategy is utilized in this chapter to force the network to learn generalized representations across different tasks, leading to improvements in the robustness and performance of the multi-task learning network.

In the context of learning from heterogeneous and limited MRI datasets, a hard parameter sharing strategy is typically preferred. Particularly, forcing the network to learn different tasks simultaneously will lead to relying on generalized representations, which reduces the chance of overfitting to a single task (Baxter, 1997). As shown in Figure 6.2e, I propose a multi-modal multi-task network (MM-Net), which extends a Dense U-Net architecture (Huang et al., 2017) to the simultaneous learning of classification, regression, and segmentation tasks from MRI datasets with paired imaging and non-imaging information. The proposed network utilizes a hard parameter sharing with a fully multi-modal design for improved network performance and robustness.

Multi-Task Losses: Choosing an appropriate loss or prioritization strategy is crucial when applying MTL to avoid learning imbalances resulting from some tasks dominating others during training (Guo et al., 2018). This dominance can be associated with differences in task heuristics, such as their complexities, uncertainties, and magnitudes of losses. Hence, instead of using a fixed weighted loss function with weights determined through costly hyperparameters tuning, automatic methods are being proposed to weight each task loss dynamically (Kendall et al., 2018; Chen et al., 2017; Liu et al., 2019; Sener and Koltun, 2018). I first investigate the use of the homoscedastic uncertainty of tasks to dynamically weight each task loss (Kendall et al., 2018). In Bayesian modeling, homoscedastic uncertainty remains constant for all network input data while varying among different training tasks, thereby describing a task-dependent uncertainty. A multi-task loss function can be derived by maximizing the Gaussian likelihood with homoscedastic uncertainty under given some sufficient statistics (Kendall et al., 2018). This results in the following formulation:

$$L(W, \sigma_1, \sigma_2, \sigma_3) \approx \frac{1}{2\sigma_1^2} L_1(W) + \frac{1}{2\sigma_2^2} L_2(W) + \frac{1}{2\sigma_3^2} L_3(W) + \log \sigma_1 + \log \sigma_2 + \log \sigma_3, \quad (6.3)$$

where W are the learned network parameters and σ_1, σ_2, and σ_3 are the model learned observation noise parameters of the classification, regression, and segmentation tasks. $\log \sigma_i$ can be viewed as a regularization term that penalizes the loss function when σ_i becomes too large.

Instead of using the scale-sensitive arithmetic mean to combine the losses of each task, I explore using a geometric loss combination that is invariant to the scale of the task losses. Hence, the multi-task loss function can be expressed as

$$L(W) = \sqrt[3]{L_1(W)L_2(W)L_3(W)}. \tag{6.4}$$

Note that the multi-loss functions described in Equations 6.4 and 6.3 are differentiable and can be optimized using an stochastic gradient descent optimizer such as Adam (Kingma and Ba, 2014). I finally propose a dynamic task prioritization scheme, where each task priority is inversely proportional to the task performance as measured using a running average of the task loss.

6.3 Experimental Results

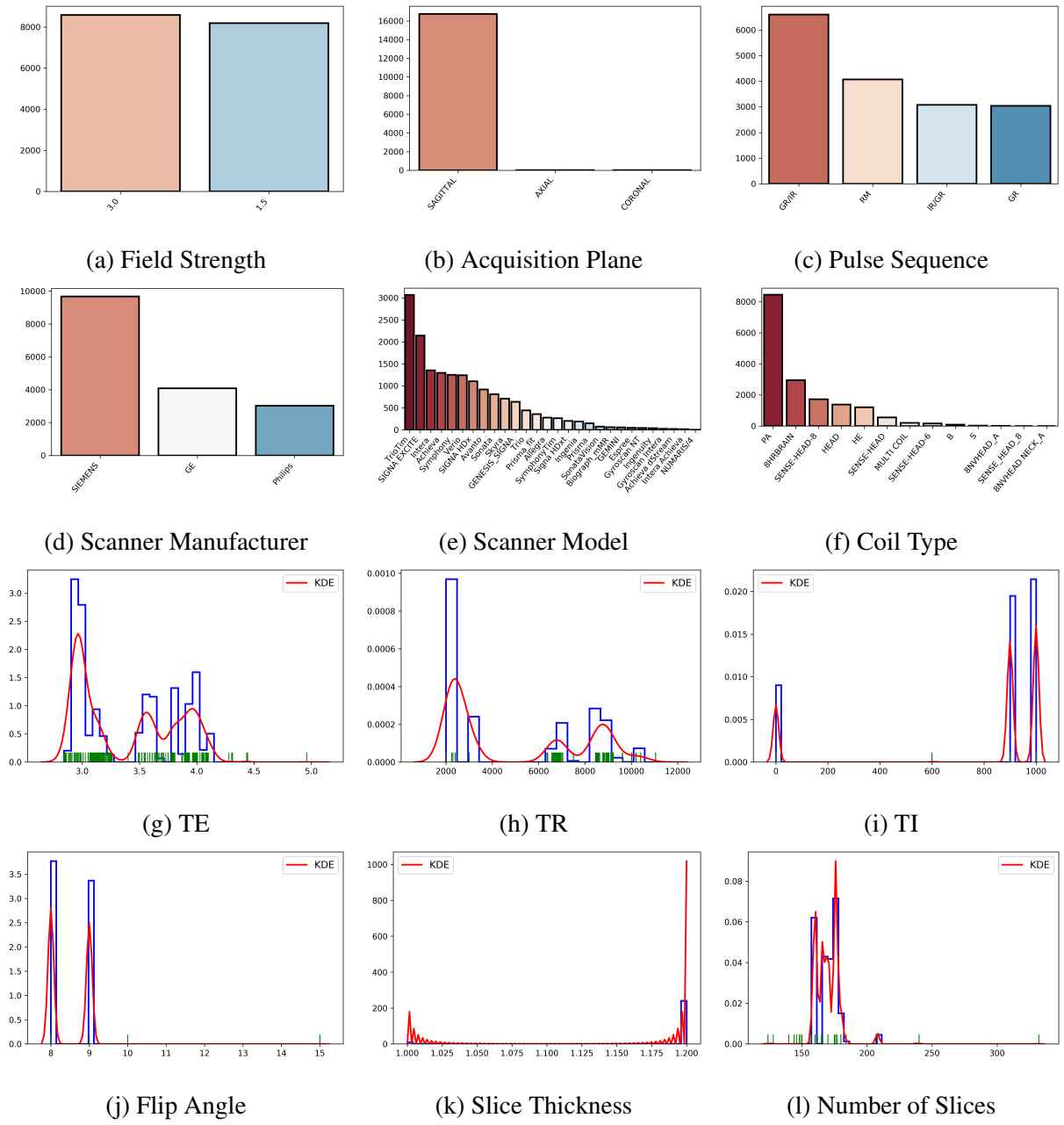
The proposed multi-modal networks (M-Nets) have been tested in the application of segmenting CSF, GM, and WM brain tissues from a large and highly heterogeneous MRI dataset. The proposed multi-task multi-modal network (MM-Net) was applied for classifying AD patients from healthy controls as well as predicting the associated Mini-Mental State Examination (MMSE) cognitive scores with MRI subcortical segmentation defined as an additional auxiliary task.

6.3.1 MRI Tissue Segmentation

6.3.1.1 Materials

ADNI Dataset: The proposed M-Nets were used to segment structural MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`adni.loni.usc.edu`). Specifically, I included all 16,788 3D MP-RAGE T1w brain images available in ADNI1, ADNI2, and ADNI3 datasets, acquired at 63 sites in the US and Canada. Related scan meta-data were extracted from the DICOM files; namely, I included information associated with pulse sequence parameters (i.e., TE, TR, TI, and flip angle), coil type, acquisition plane, scanner manufacturer, scanner model, field strength, image resolution, and field of view (see Figure 6.4). Additional sources of data

heterogeneity stem from using different age at scan (75.6± 7.4 years), sex (55% Males, 45% Females), and disease progression status (31.6% CN, 51.21% MCI, 13.76% AD, 3.43% SMC).



(a) Field Strength (b) Acquisition Plane (c) Pulse Sequence

(d) Scanner Manufacturer (e) Scanner Model (f) Coil Type

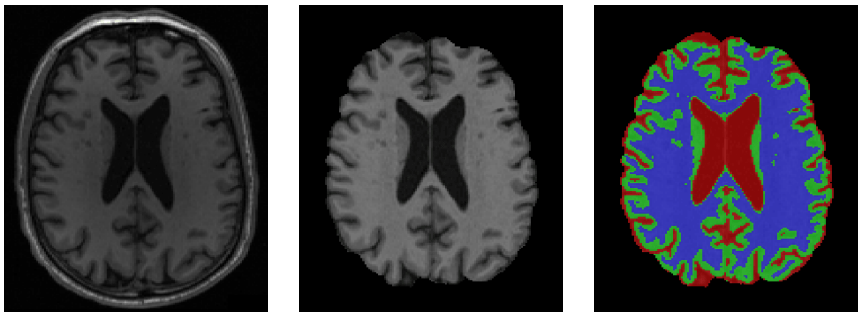
(g) TE (h) TR (i) TI

(j) Flip Angle (k) Slice Thickness (l) Number of Slices

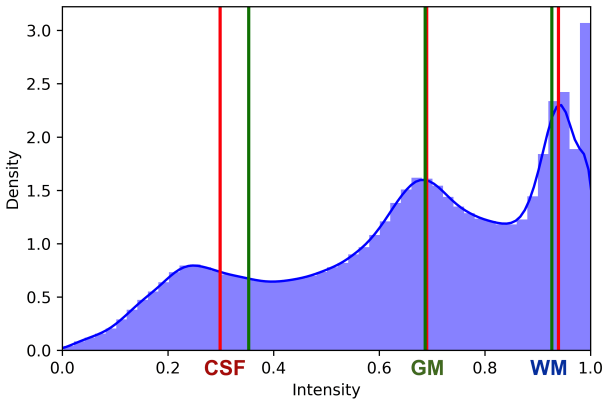
Figure 6.4: Examples of (a-f) categorical and (g-l) continuous meta-data in the combined ADNI1, ADNI2, and ADNI3 datasets, where significant variability can be observed, which is modeled by the proposed multi-modal networks to account for the associated intensity variability.

6.3.1.2 Data Preprocessing

Due to the lack of publicly available manually segmented brain MRI datasets, pseudo-ground-truth annotations were utilized. As shown in Figure 6.5, to generate such labels, I first apply N4ITK bias correction (Tustison et al., 2010), then BET2 skull stripping (Jenkinson et al., 2005), and finally, a segmentation of CSF, GM, and WM brain tissues using FSL (Zhang et al., 2001). While labels from human experts would be ideal, I believe that FSL annotations can serve as a reasonable proxy to evaluate the proposed methods. No image pre-processing was performed except resizing the bias-corrected images to $128 \times 128 \times 112$ and linearly scaling intensities to the [0,1] range. Besides the previously mentioned meta-data, I also included the mean intensity of CSF, GM, and WM tissues estimated using a Gaussian Mixture Model (GMM) with 3 components (see Figure 6.5d).



(a) Input T1w MRI (b) Skull Stripping (c) Tissue Segmentation



(d) Normalized Intensity Histogram

Figure 6.5: Applied to the (a) input T1w MRI, (b) N4ITK bias correction (Tustison et al., 2010) followed by BET2 skull stripping (Jenkinson et al., 2005) were applied before (c) brain tissues were segmented using FSL (Zhang et al., 2001) to generate pseudo-ground-truth labels. Besides the scan meta-data, (d) the mean intensity of CSF, GM, and WM tissues estimated using a GMM (shown in red), provide a reasonable approximation to the true intensity means (shown in green).

The categorical meta-data were encoded using a one-hot representation while continuous meta-data were min-max normalized (meta-data vector with 72 dimensions). Figure 6.6 shows the intensity variability of CSF, GM, and WM brain tissues for selected acquisition settings, illustrating the usefulness of feeding such meta-data to the proposed M-Nets to extract relevant information.

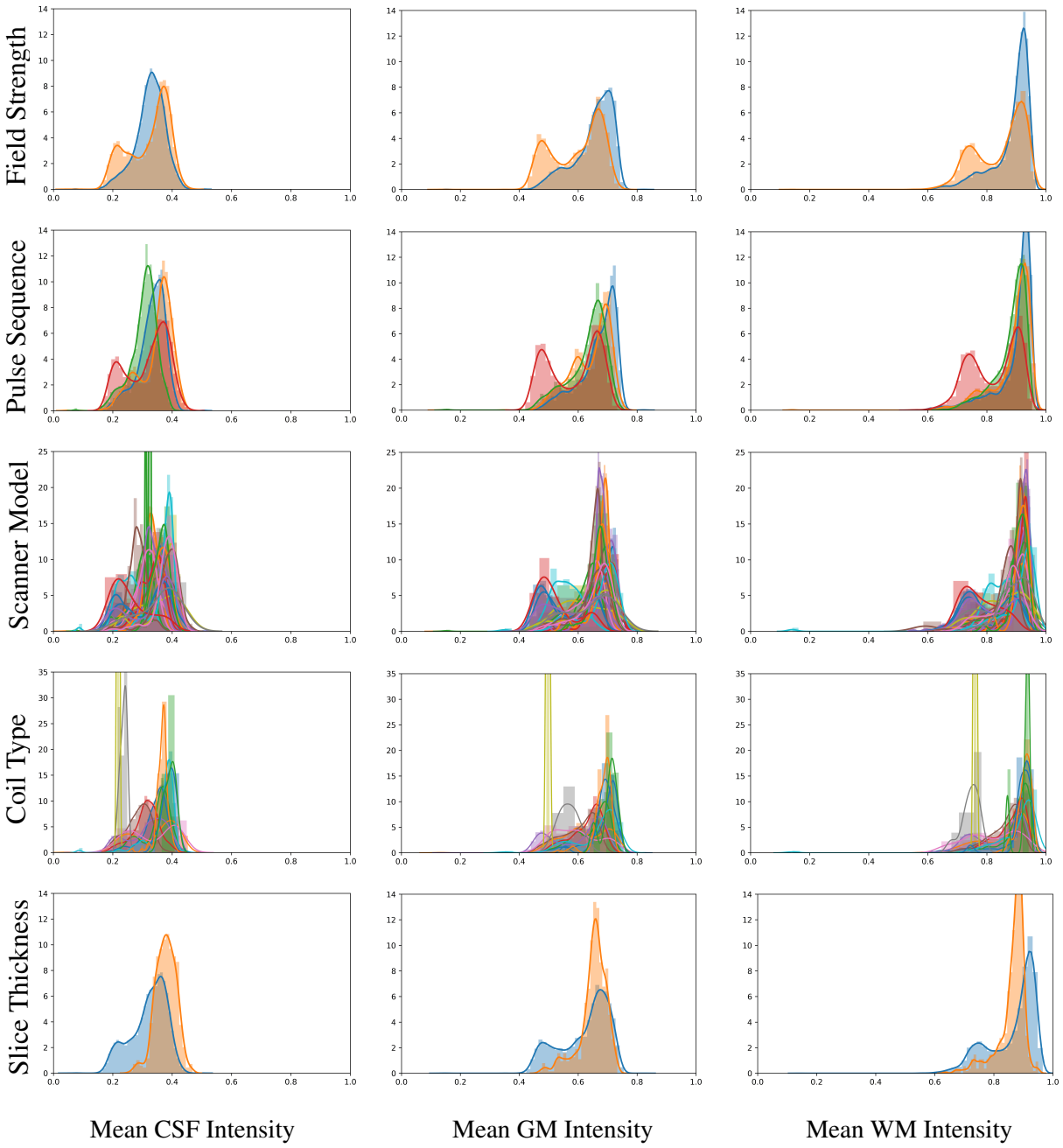


Figure 6.6: The distributions of normalized mean brain tissues (i.e., CSF, GM, and WM) intensity for selected acquisition settings. Significant intensity variations can be observed in the marginal distributions, which indicate the usefulness of incorporating such meta-data in the learning process.

6.3.1.3 Networks Architecture and Training

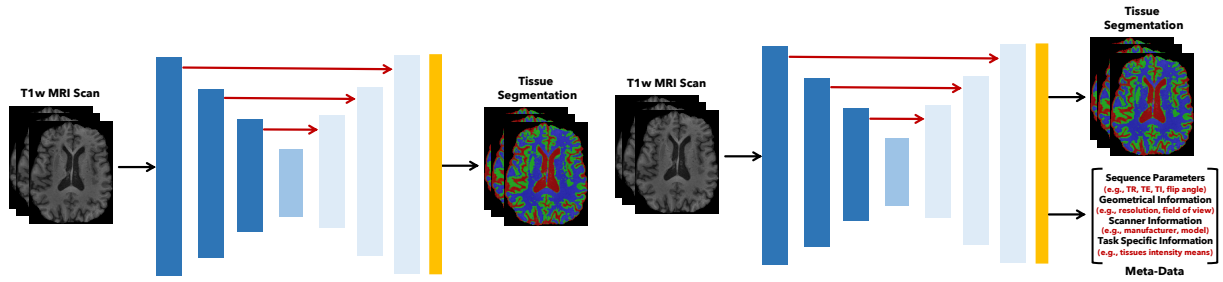
Following the base multi-modal block defined as shown in Figure 6.2a, a fully multi-modal U-Net architecture was designed for the robust segmentation of CSF, GM, and WM brain tissues from a T1w MRI scan, as illustrated in Figure 6.2b. As shown in Figure 6.2c, a conditional multi-modal U-Net that partially composed from multi-modal blocks was also investigated. Moreover, a multi-modal U-Net architecture that also trained to model the heteroscedastic aleatoric uncertainty was additionally investigated, as demonstrated in Figure 6.2d. The proposed multi-modal networks were trained on 50% of the combined ADNI dataset while the 25% were used for validation (optimal model selected based on the validation loss) and the remaining 25% were used for testing. The key parameters used to build and train the proposed multi-modal networks are given in Table 6.1.

Levels of Pooling	3	Kernel Size	3	Growth Rate	4
Normalization	Instance	Pooling Type	Max	Activations	Leaky ReLU
Final Activation	Softmax	Loss	Weighted CE	Batch Size	32
Optimizer	Adam	Learning Rate	0.00001	Epochs	200

Table 6.1: Key parameters used to build and train the proposed multi-modal networks.

6.3.1.4 Performance Evaluation

The performance of the proposed multi-modal networks was evaluated using the Dice Similarity Coefficient (DSC) (Zou et al., 2004), computed using the network prediction and corresponding pseudo-ground-truth. To confirm the usefulness of incorporating meta-data during training, I compared the proposed multi-modal network against a similar meta-data free U-Net (Figure 6.7a). The proposed methods were also compared against a multitask U-Net (Figure 6.7b), which provides an alternative approach to incorporate meta-data information as auxiliary outputs. Table 6.2 provides a summary of the segmentation results using the mean Dice scores while Figure 6.8 shows the proposed multi-modal U-Net results for randomly selected examples from the testing set.



(a) Meta-Data Free U-Net (b) Multi-Task U-Net

Figure 6.7: Tissue segmentation networks used to generate baseline results. The usefulness of incorporating meta-data during training is illustrated by comparing against (a) a meta-data free U-Net. I also examine the proposed multi-modal networks against (b) a multitask U-Net that provides an alternative approach to incorporate meta-data information as auxiliary outputs.

		CSF	GM	WM
Baseline Networks	Meta-Data Free U-Net	0.867	0.869	0.886
	Multitask U-Net	0.870	0.875	0.890
Proposed Networks	**Multi-Modal U-Net**	**0.903**	**0.901**	**0.907**
	Conditional Multi-Modal U-Net	0.850	0.853	0.868
	Uncertainty Multi-Modal U-Net	0.840	0.890	0.882

Table 6.2: Summary of the segmentation results of the proposed multi-modal networks and the competing baseline networks using the mean Dice scores for CSF, GM, and WM brain tissues.

As documented in Table 6.2, the proposed fully multi-modal U-Net performs notably better than the U-Net and multitask U-Net baseline networks. This confirms that the proposed segmentation network benefits from the availability of relevant meta-data information, which is reflected in its ability to generalize better in a heterogeneous setting. The obtained results of the multitask learning approach are promising, and such a network could be useful in cases where corresponding meta-data information is not available at test time. However, such a solution is still limited by the ambiguity in choosing from the various weighting schemes in the multitask loss function. On the other hand, the lower Dice scores obtained using the partially multi-modal U-Net architecture may be explained by the network ignoring meta-data information. Degradation in performance level was observed when

aleatoric uncertainty was modeled, despite the performance gains promised in computer vision tasks (Kendall and Gal, 2017). The heterogeneous nature of the dataset plus the noisy labels of the pseudo-ground-truth might explain the network's inability to learn an accurate probabilistic model.

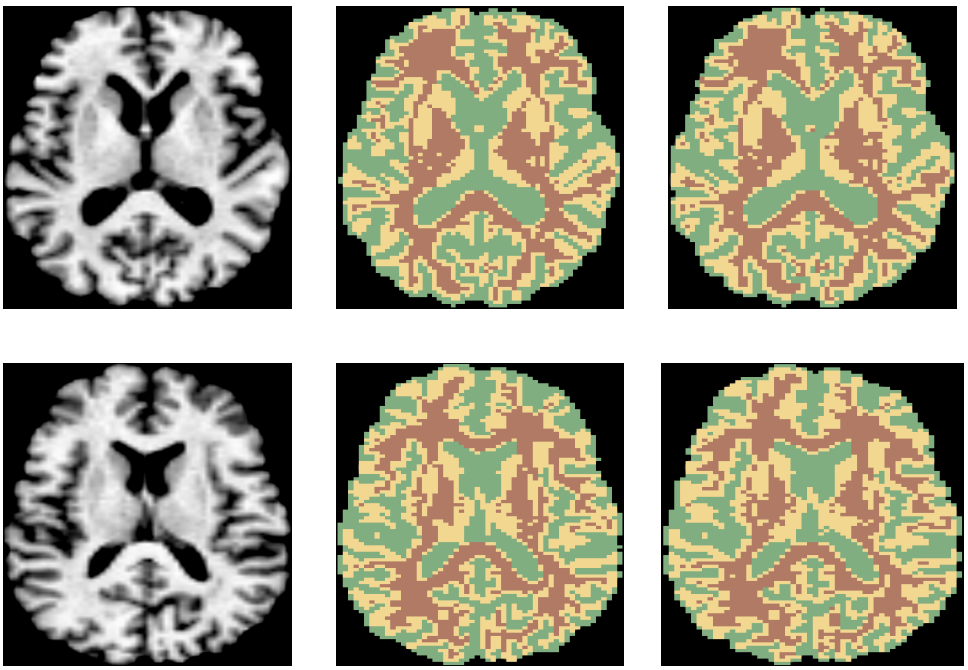


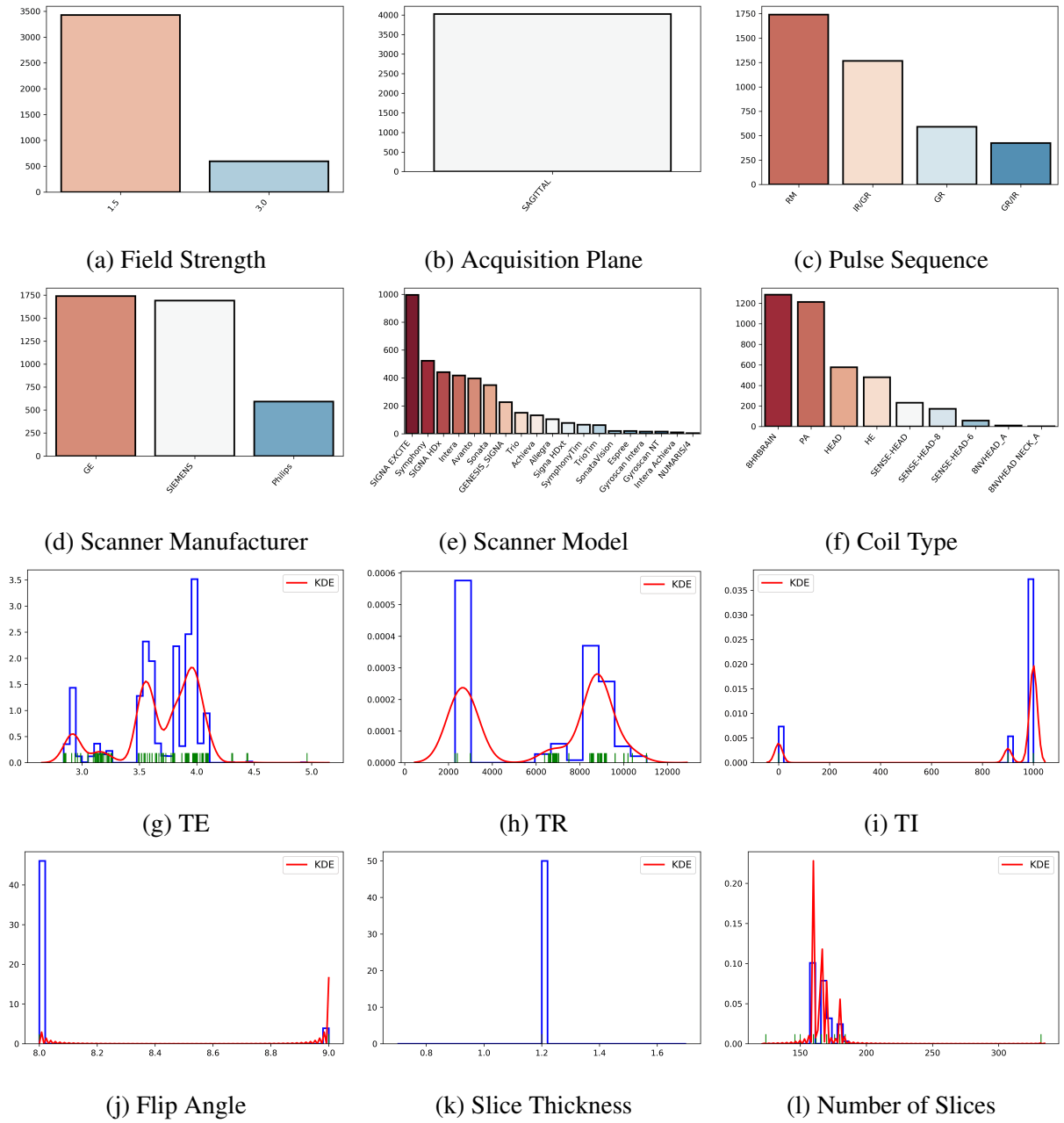
Figure 6.8: Randomly selected examples from the testing set showing the segmentation results of the proposed multi-modal U-NET as compared to the pseudo-ground-truth segmentation.

6.3.2 Alzheimer's Disease Classification and Cognitive Scores Prediction

6.3.2.1 Materials

ADNI1 Dataset: The Proposed MM-Net was applied to a subset of the structural MRI data from the ADNI database (`adni.loni.usc.edu`). Notably, I included all T1w scans of AD patients and CN patients from the ADNI1 dataset, which were acquired at 63 research centers in the United States and Canada. This resulted in a dataset of 4023 3D MP-RAGE T1w scans with corresponding scan meta-data extracted from the DICOM files, including pulse sequence parameters (i.e., TE, TR, TI, and flip angle), coil type, acquisition plane, scanner manufacturer, scanner model, field strength, image resolution, and field of view (see Figure 6.9). Moreover, the sex (50.8% males,

49.2% females) and age at scan (75.1± 6.2 years) demographic information (see Figure 6.10a,b) were also collected to aid in the prediction tasks; namely, classifying AD patients (63.2%) from CN patients (36.8%) (see Figure 6.10c) and predicting associated MMSE cognitive scores (26.3± 4.7 points) (see Figure 6.10d).



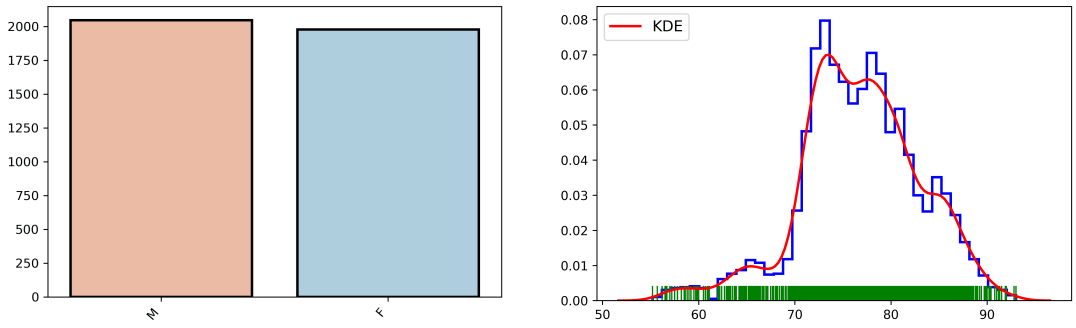
(a) Field Strength (b) Acquisition Plane (c) Pulse Sequence

(d) Scanner Manufacturer (e) Scanner Model (f) Coil Type

(g) TE (h) TR (i) TI

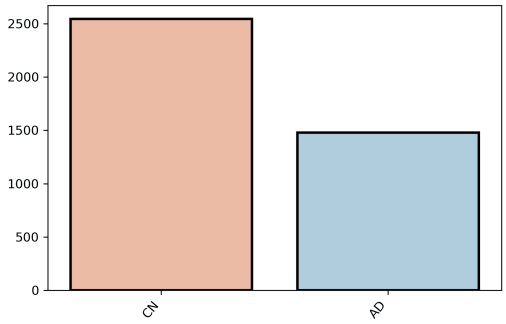
(j) Flip Angle (k) Slice Thickness (l) Number of Slices

Figure 6.9: Examples of (a-f) categorical and (g-l) continuous meta-data associated with the scans included from the ADNI1 dataset. Similar to the combined ADNI dataset, significant variability is observed across different meta-data except for acquisition plane and slice thickness information.

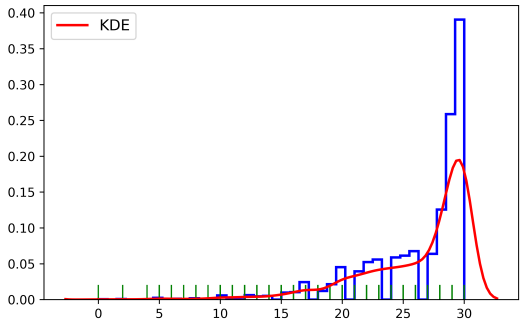


(a) Sex

(b) Age at Scan



(c) Diagnostic Group



(d) MMSE Cognitive Score

Figure 6.10: The (a) sex and (b) age at scan demographic information were also incorporated in the meta-data fed to the proposed MM-Net, which was trained to simultaneously (c) classify AD patients from CN patients and (d) predict the associated MMSE cognitive scores.

6.3.2.2 Data Preprocessing

In addition to the defined classification and regression prediction tasks, the proposed MM-Net was also trained to segment subcortical structures from the input T1w MRI. To tackle the lack of publicly available MRI datasets with manually labeled subcortical structures, I utilized pseudo-ground-truth labels for training the proposed MM-Net. As shown in Figure 6.11, the FreeSurfer pipeline (Fischl, 2012) was used to generate such labels. In particular, the following preprocessing steps were applied subsequently to the input T1w MRI, namely, motion correction, affine transformation to Talairach image space, intensity inhomogeneity correction, and removal of non-brain tissues. The intensity of the extracted brain volume was then normalized before applying a non-linear warping of the atlas brain image to produce the required subcortical structures labels.

As shown in Figure 6.11c, 7 subcortical structures were segmented, namely, lateral ventricle (LV), thalamus (TH), caudate (CA), putamen (PU), pallidum (PA), hippocampus (HI), amygdala (AM). The obtained pseudo-ground-truth labels for subcortical structures were significantly noisy. However, a rough localization of subcortical structures can still be useful in improving the performance of the proposed MM-Net in the classification and regression prediction tasks.

(a) Input T1w MRI (b) Cropped Input T1w MRI (c) Subcortical Segmentation

Figure 6.11: Subcortical structures (lateral ventricle, thalamus, caudate, putamen, pallidum, hippocampus, amygdala) were segmented using the FreeSurfer pipeline (Fischl, 2012) to provide the required pseudo-ground-truth labels. Note that the T1w MRIs and corresponding label maps were cropped using a common maximum bounding box estimated from the training dataset.

To control the network complexity, a common maximum bounding box ($128 \times 96 \times 96$) was learned from the training data to crop the input T1w MRI to limit the network training to brain regions containing subcortical structures (see Figures 6.11a and b). Additionally, the intensities of the bias-corrected images were linearly scaled to the [0,1] range. In addition to the meta-data utilized in the MRI tissue segmentation application, I also included the sex and age at scan demographic information. Categorical meta-data were encoded using a one-hot representation while continuous meta-data were min-max normalized, resulting in a meta-data vector with 52 dimensions.

6.3.2.3 Network Architecture and Training

Following the base multi-modal block is defined as shown in Figure 6.2a, a fully multi-modal multi-task Dense-U-Net architecture was designed for performing classification, regression, and segmentation tasks jointly from T1w MRI (see Figure 6.2e). The proposed networks were trained

and validated on 80% of the ADNI1 dataset while the remaining 20% were used for testing. To handle the class imbalance problem for the classification task, in the training phase, the dataset was balanced through oversampling by duplicating some of the original samples of the minority (AD) class. The key parameters used to build and train the proposed MM-Net are given in Table 6.3.

Levels of Pooling	4	Layers per Block	4	Task Blocks	1
Growth Rate	1	Kernel Size	3	Normalization	Instance
Pooling Type	Max	Activations	Leaky ReLU	Class. Activation	Softmax
Reg. Activation	Linear	Seg. Activation	Softmax	Class. Loss	CE
Reg. Loss	Huber	Seg. Loss	Weighted CE	Batch Size	4
Optimizer	Adam	Learning Rate	0.00001	Epochs	200

Table 6.3: Key parameters used to build and train the proposed multi-modal multi-task network.

6.3.2.4 Performance Evaluation

The DSC (Zou et al., 2004) scores, computed using the network prediction and corresponding pseudo-ground-truth, were used to evaluate the segmentation task performance by the proposed MM-Net. To assess the prediction tasks, the mean ACC was used to evaluate the classification task while Mean Absolute Error (MAE) was used to evaluate the regression task. To illustrate the usefulness of jointly training the proposed MM-Net on multiple tasks, I first trained the network to perform every single task separately. The proposed MM-Net was also trained to perform segmentation, classification, and regression tasks jointly with different strategies for combining the losses of each task. In particular, I explore using uncertainty, geometric, and priority multi-task loss strategies (see Subsection 6.2.3 for more details). Table 6.4 provides a summary of the obtained segmentation, classification, and regression results using the proposed MM-Net when trained to perform each task separately and when trained in a multi-task learning setting.

As documented in Table 6.4, when trained on the segmentation task only, the proposed MM-Net provided good segmentation quality for larger subcortical structures (e.g., lateral ventricle and

thalamus). However, due to the limited resolution of the input T1w MRI ($1 \times 1 \times 1.2$ mm^3) and the noisy nature of the ground truth labels, the DSC segmentation results were relatively low for smaller subcortical structures (e.g., pallidum and amygdala). An MM-Net with larger capacity (i.e., increased number of blocks, number of layers per block, and layers growth rate) would be able to improve on current segmentation results; however, the current model capacity was limited by the memory size of available GPUs (NVIDIA TITAN RTX 24GB was utilized for this application). On the other hand, the proposed MM-Net trained on a single prediction task was able to classify 93.9% of samples in the testing set correctly and accurately predict the associated MMSE cognitive scores with MAE of 1.63 points. When trained to perform the segmentation, classification, and regression tasks jointly, the proposed MM-Net was able to improve its performance on the prediction tasks as indicated by the obtained ACC and MAE scores. Since the segmentation task was only defined as an auxiliary task, one can tolerate the drop in the segmentation results were observed across the different multi-task loss strategies. Among the investigated multi-task loss strategies, utilizing the homoscedastic uncertainty of tasks to dynamically weighting each task loss resulted in the best prediction performance (ACC score of 98.9% and MAE score of 1.36 points) with a slight drop in the obtained DSC segmentation scores.

	DSC Scores for Segmented Subcortical Structures							Prediction Accuracy	
	LV	TH	CA	PU	PA	HI	AM	ACC (%)	MAE
Segmentation Task	0.937	0.813	0.758	0.784	0.689	0.742	0.644	-	-
Classification Task	-	-	-	-	-	-	-	93.91	-
Regression Task	-	-	-	-	-	-	-	-	1.63
Uncertainty Multi-Task	0.92	0.793	0.703	0.737	0.639	0.683	0.570	**98.90**	**1.36**
Geometric Multi-Task	0.770	0.534	0.485	0.516	0.462	0.389	0.367	97.80	1.37
Priority Multi-Task	0.720	0.556	0.444	0.493	0.435	0.357	0.306	97.40	1.38

Table 6.4: Summary of the segmentation, classification, and regression results obtained using the proposed multi-task learning framework. The mean Dice Similarity Coefficient (DSC) scores were used to asses the segmentation task, the mean Accuracy (ACC) scores to asses the classification task, and the Mean Absolute Error (MAE) scores to asses the regression task.

6.4 Summary

This chapter proposed a novel design for a multi-modal block as a generalization of convolutional or fully connected layers for efficient representations learning from paired data (meta-data, images). I also described multi-modal networks (M-Nets) as a combination of such multi-modal blocks and other deep learning layers, allowing for the availability of relevant meta-data at different scales. For robust MRI tissue segmentation, M-Nets utilizing a 3D U-Net architecture were implemented as a combination of such multi-modal blocks with scanner information, MRI sequence parameters, geometrical information, and task-specific prior information used as meta-data. The proposed methods were extended in a unified multi-task multi-modal learning framework (MM-Net) that utilizes a 3D Dense U-Net architecture for the robust joint learning of classification, regression, and segmentation tasks in multi-modal heterogeneous structural MRI datasets.

The robustness of the proposed M-Nets was demonstrated in providing fast and accurate tissue segmentation results in a large and highly heterogeneous (wide range of MRI protocols) MRI dataset. The proposed fully multi-modal U-Net outperformed the competing segmentation networks, confirming its ability to learn robust representations efficiently. The high performance of the proposed MM-Net was tested in jointly classifying AD patients from CN patients and predicting associated MMSE cognitive scores using input T1w MRI and linked meta-data (i.e., scan parameters and patient demographic information). Moreover, subcortical structures segmentation was additionally defined as an auxiliary task to improve the overall performance of the multi-task learning framework. The obtained results from the jointly trained MM-Net confirmed its ability to exploit the underlying similarities between correlated tasks to improve the generalization and accuracy of prediction models. I also demonstrated that modeling the homoscedastic uncertainty of tasks to dynamically weight each task loss provided superior performance to competing geometric and priority multi-task loss strategies. In total, the proposed methods presented in this chapter for improving the robustness and the performance of deep learning models may help increase the confidence of the clinical decisions of neuroimaging pipelines utilizing these models.

CHAPTER 7: OUT-OF-SAMPLE DETECTION IN HETEROGENEOUS SETTINGS

7.1 Overview

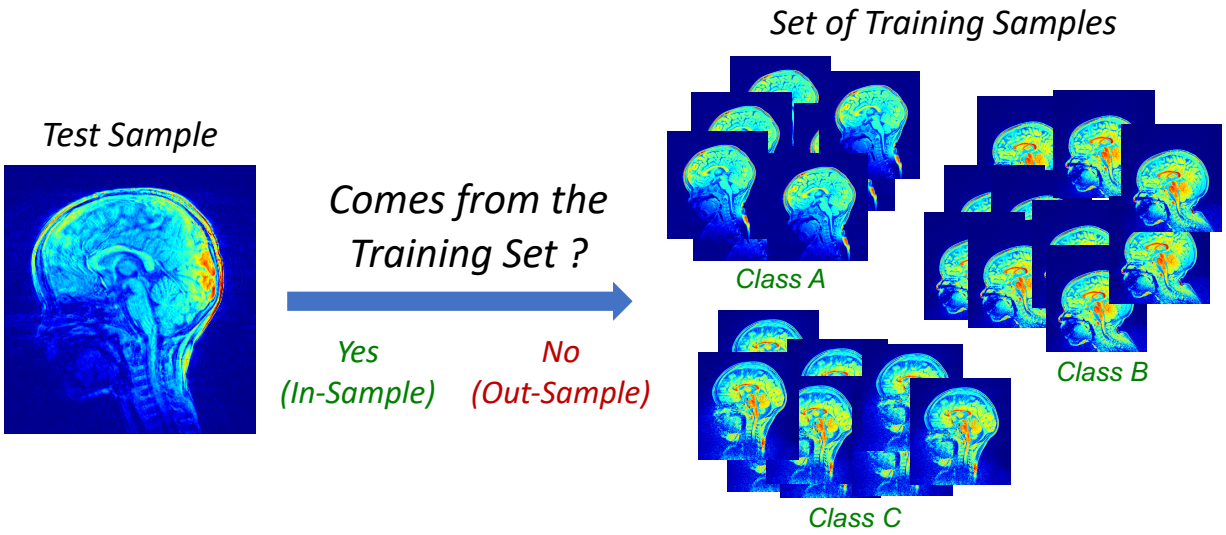


Figure 7.1: Illustration of the out-of-sample/ anomaly detection problem, which refers to the issue of determining whether an unseen test sample is from a different distribution from the training data, i.e., can we predict if it is from within the same distribution?

Discriminative deep learning models have shown notable success in many medical image analysis applications, including segmentation and disease prediction (Litjens et al., 2017). However, in problems that involve learning from noisy datasets with highly imbalanced classes, their success is limited (Litjens et al., 2017). The class imbalance problem has a substantial negative impact on training deep learning models, as learning would focus on samples from the majority classes, leading to poor performance for the weakly represented classes. This problem is typically resolved by modifying the representation of the minority classes using data augmentation or oversampling techniques (Krawczyk, 2016). Such a strategy fails if the minority classes are also inherently heterogeneous, and some classes missing from the training dataset. Thereby, the trained models can

provide incorrect labels with very high confidence for underrepresented samples (Elsayed et al., 2018). Generative models parameterized using deep learning models that can exploit both unlabeled and labeled data can provide relief to solve this problem. By only observing samples from the majority classes, these models can learn the underlying data distribution, which can be used to identify samples that are unlikely have been generated by the same distribution (see Figure 7.1). The application of out-of-sample/ anomaly detection methods to medical imaging is discussed next.

Application to Medical Imaging: Several medical imaging out-of-sample methods have been proposed that utilize deep generative architectures such as VAEs or GANs. VAEs explicitly try to approximate the data distribution of the training data both in the latent and the original space by maximizing a variational lower bound (Kingma and Welling, 2013). In the original space out-of-sample detection can be performed using the reconstruction score (Lu and Xu, 2018) or distance- or density-based approaches (Vasilev et al., 2018). Anomalies can also be detected in the learned latent space using VAE regularizer scores (Lu and Xu, 2018) or an offline one-class support vector machines classifier (El Azami et al., 2016). Compared to VAEs, GANs avoids any strong distribution assumptions by implicitly specifying probabilistic models without requiring a likelihood function to be defined. GANs only define a generating procedure to generate sharp and compelling higher quality images directly (Goodfellow et al., 2014). GAN-based anomaly detection for testing samples is performed based on post hoc estimated likelihood scores (Schlegl et al., 2017). Nonetheless, GANs are still challenging to train, and the post hoc score estimators are very slow and tend to produce inaccurate likelihood estimates.

Combining VAE and GAN: In 2018, architectures utilizing both a VAE and a GAN were introduced for pixel-wise anomaly detection. An adversarial autoencoder (Makhzani et al., 2015) with latent space consistency constraints was proposed to identify anomalies based on created reconstruction error maps (Chen and Konukoglu, 2018). Similarly, an architecture that combines a VAE and a GAN was also proposed (Baur et al., 2018) to detect within image anomalies, which showed improvements over a GAN-only approach (Schlegl et al., 2017). However, such models were designed to detect only pixel-wise anomalies in limited 2D homogeneous datasets and failed

to incorporate external information to organize the learned image manifolds. Hence, there is a need for methods to provide sample-wise anomaly scores in 3D high-dimensional heterogeneous data.

Structural MRI Quality Control: Acquiring an appropriate T1w MRI is critical to neuroimaging studies as T1w images are necessary not just in structural MRI workflows but also in functional and diffusion connectivity ones. Thus, during the scanning session MR technicians need to assess whether an acquired T1w image is appropriate for quantitative neuroimaging workflows or whether a repeat T1w scan is required. Outside of academic research centers, MRI technicians, though, are often not sufficiently trained for this task, such that the availability of a fast within-session automatic Quality Control (QC) method to accept/reject T1w images is highly desired. The method needs to be robust to different settings, including scanner models, acquisition parameters, and subject age. Discriminative approaches would fail to generalize due to data imbalance (ratio$\approx$1:20) and the heterogeneity of the failed classes (failure due to different MRI artifacts such as motion, noise, poor contrast, etc.). A deep generative model utilized for out-of-sample detection by training only on QC-passed MR images could be a potential solution to this problem.

Proposed Semi-Supervised VAE-GAN: The current work aims to provide a fast within-session automatic sMRI QC tool that relies on a novel generative approach to tackle the limitations mentioned above of current medical imaging anomaly detection methods. In particular, I propose a novel semi-supervised out-of-sample detection framework that robustly identifies samples outside the trained classes in complex sMRI datasets[1]. As shown in Fig. 7.2, the proposed architecture (Anomaly-VAEGAN) extends the VAE-GAN model (Larsen et al., 2015) to overcome the limitations of similar detection methods. My main contributions are summarized as follows:

[1] Instead of the element-wise reconstruction scores (ℓ_2 or ℓ_1), scores are assigned using a learned high-level similarity metric and invariant representations for accounting for spatial variability.

[2] The VAE-GAN architecture (Larsen et al., 2015) is adapted using a semi-supervised discriminator to stabilize network training and ensure a meaningful learned similarity metric.

[1]The work is based on the previously published paper (Mostapha et al., 2019). This chapter partially adapts text descriptions and figures from the published paper.

[3] The encoder is jointly trained with an additional predictor model to enforce the learned representations to be correlated with external desired properties (e.g., non-imaging information).

[4] A one-class classifier is jointly trained with the VAE-GAN to produce scores that can be used to detect anomaly samples at the testing time while also influencing the learned representations.

[5] A novel application of a deep generative method for the automatic QC of structural MRI, which I additionally evaluate in a realistic setting (i.e., applied to a large and heterogeneous dataset).

Details of the proposed out-of-sample detection framework will be provided next.

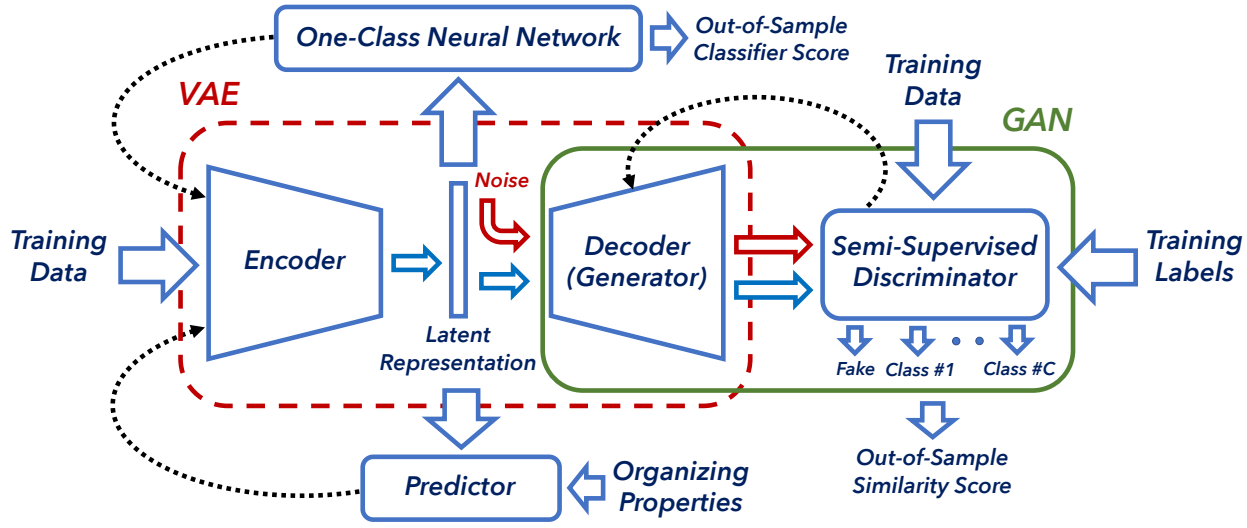


Figure 7.2: Illustration of the proposed out-of-sample detection framework in which VAE-GAN is adapted for semi-supervised learning by adding a semi-supervised discriminator and a properties predictor. Anomaly samples are detected using reconstruction scores based on a learned high-level similarity metric or using scores from a one-class classifier that is jointly trained. Dotted black lines indicate gradient feedback during training.

7.2 Methods

In this section, I briefly overview the related background on VAEs and GANs (see Chapter 2, Section 2.3 for complete background information) and also describe how they were combined in the proposed semi-supervised out-of-sample detection framework.

7.2.1 Variational Autoencoder

A VAE (Kingma and Welling, 2013) consists of an encoder network that maps a data sample $x \in \mathbb{R}^d$ to latent representation $z \in \mathbb{R}^m \sim \text{Enc}(x) = p(z|x)$ and a decoder network that attempts to recapture the original input $\hat{x} \in \mathbb{R}^d \sim \text{Dec}(z) = q(x|z)$. The VAE regularizes the encoder by forcing z to follow some prior distribution $p(z)$, which is chosen to be $\mathcal{N}(0, I)$ in this work. This will allow the decoder to generate new samples by stochastically sampling in the latent space using $p(z)$. Since the data distribution $p(x)$ is intractable, variational inference methods are introduced to approximately estimate $p(x)$ by maximizing the evidence lower bound (ELBO):

$$\log p(x) \geq -KL(q(z|x)||p(z)) + \mathbb{E}_{q(z|x)}[\log p(x|z)], \tag{7.1}$$

where KL is the Kullback-Leibler divergence. I choose $q(z|x) \sim \mathcal{N}(\mu_{enc}(x), \Sigma_{enc}(x))$ and $p(x|z) \sim \mathcal{N}(\mu_{dec}(z), \sigma^2 I)$ to be diagonal Gaussian distributions, where $\mu_{enc}(x)$ and $\Sigma_{enc}(x)$ are learned by the encoder, $\mu_{dec}(z)$ is learned by the decoder, and σ is pre-determined. Using the reparameterization trick ($z = \mu_{enc}(x) + \epsilon\Sigma^{\frac{1}{2}}(x)$, where $\epsilon \sim \mathcal{N}(0, I)$), the network is trained from end to end by minimizing the following VAE loss which is equivalent to maximizing Eqn. (7.1):

$$\begin{aligned}\mathcal{L}_{vae} &= KL(\mathcal{N}(\mu_{enc}(x), \Sigma_{enc}(x))||\mathcal{N}(0, I)) + \mathbb{E}_{\epsilon\sim\mathcal{N}(0,I)}\left[\frac{(x - \mu_{dec}(z))^T(x - \mu_{dec}(z))}{2\sigma^2}\right] \\ &= \frac{1}{2}(1 + \log(\Sigma_{enc}(x)) - \mu_{enc}^2(x) - \Sigma_{enc}(x)) + \frac{1}{2\sigma^2}\|x - \hat{x}\|^2 = \mathcal{L}_{prior} + \lambda\mathcal{L}_{recons}^{pixel}.\end{aligned} \tag{7.2}$$

7.2.2 Generative Adversarial Network

A GAN (Goodfellow et al., 2014) consists of a generator network, which attempts to map the latent code $z \sim p(z)$ to a realistic image $\text{Gen}(z)$, and a discriminator, which assigns a probability $y = \text{Disc}(x) \in [0, 1]$ to input images from the training data $x \sim p(x)$ and a probability $1 - y$ to images generated by the model $x = \text{Gen}(z)$. This introduces a min-max optimization problem,

which can be expressed as

$$\min_{Gen} \max_{Disc} \mathcal{L}_{gan} = \mathbb{E}_{x\sim p(x)}[\log \text{Disc}(x)] + \mathbb{E}_{z\sim p(z)}[\log(1 - \text{Disc}(\text{Gen}(z))]. \tag{7.3}$$

7.2.3 Semi-Supervised VAE-GAN for Out-of-Sample Detection

The proposed framework is based on a VAE-GAN architecture (Larsen et al., 2015) that was shown the ability to generate higher quality images as compared to VAEs while also yielding more diverse images compared to GANs. Next, details are provided on how the VAE-GAN model is adapted for out-of-sample detection.

Learned Similarity Metric: The pixel-wise reconstruction loss (e.g., mean square error), described in Eqn. (7.2), is not suitable for image-based applications as it does not provide invariance to spatial variability or intensity variations expected when learning from noisy heterogeneous datasets. Such element-wise similarity metrics do not take into account any peculiarity of the human visual system, and their use often leads to an inadequate evaluation of generated images, leading to blurry reconstructions. VAE-GAN solves this problem by replacing the VAE decoder with a generator of a GAN and adjusting the loss function to be computed using a discriminator. The motivation behind this modification is to exploit the discriminator's ability to learn an image-specific similarity metric to better asses the reconstructed images. In particular, a feature-wise reconstruction loss is defined to measure image similarity using high-level representations learned at the dth layer of the discriminator:

$$\mathcal{L}_{recons}^{\text{Disc}_d} = -\mathbb{E}_{q(z|x)}[\log p(\text{Disc}_d(x)|z)],\ p(\text{Disc}_d(x)|z) \sim \mathcal{N}(\text{Disc}_d(x)|\text{Disc}_d(\hat{x}), I). \tag{7.4}$$

Semi-Supervised Discriminator: Since the discriminator is only trained for recognizing real images, there is no guarantee that the learned similarity metric would capture subtle differences in image quality. One way to combat that is to restrict the discriminator architecture to convolutional layers that are effective for transfer learning. Additionally, to deal with this problem, I adapt

the VAE-GAN architecture for semi-supervised learning by extending the discriminator to also determine the class for the training examples. In addition to ensuring a more meaningful learned similarity metric, supervision with class labels helps to handle the GAN non-convex cost, thereby stabilizing the network during training. In this work semi-supervised learning is performed using a standard classifier that is trained on an extended dataset that includes images produced by the generator with corresponding additional "fake" class label ($y = C + 1$, where C is the number of classes in the training set). Then, the proposed semi-supervised discriminator can be trained to minimize the cross-entropy loss between the true class labels $y \in \{1, ..., C + 1\}$ and the predictive distribution $p_{\text{disc}}(y|x)$ produced by a softmax function at the last layer of the discriminator. Hence, the GAN cost in Eqn. (7.3) can be expressed as

$$\begin{aligned} \mathcal{L}_{gan} &= \mathbb{E}_{x,y\sim p(x,y)}[\log p_{\text{disc}}(y|x)] + \mathbb{E}_{\hat{x}\sim Gen}[\log(p_{\text{disc}}(y|\hat{x}, y = C + 1)] \\ &= \mathcal{L}_{supervised} + \mathcal{L}_{unsupervised}, \end{aligned} \tag{7.5}$$

where $\mathcal{L}_{supervised}$ is the standard supervised loss expressed as

$$\mathcal{L}_{supervised} = \mathbb{E}_{x,y\sim p(x,y)}[\log p_{\text{disc}}(y|x, y \in \{1, ..., C\})] \tag{7.6}$$

while $\mathcal{L}_{unsupervised}$ is the standard GAN loss that is modified to provide a learning signal by discriminating based on samples from latent code from $p(z)$ as well as $q(z|x)$, resulting in the following unsupervised loss:

$$\begin{aligned} \mathcal{L}_{unsupervised} = \mathbb{E}_{x\sim p(x)}[\log(1 - (p_{\text{disc}}(y|x, y = C + 1))] + \\ \mathbb{E}_{z\sim p(z|x)}[\log(p_{\text{disc}}(y|\text{Gen}(z), y = C + 1)] + \mathbb{E}_{z_p\sim p(z)}[\log(p_{\text{disc}}(y|\text{Gen}(z_p), y = C + 1). \end{aligned} \tag{7.7}$$

Constrained Latent Representations: A limitation of current generative approaches is their lack of consistency in their encoded representations. This can be explained by the absence of constraints that would encourage samples with similar properties to be mapped to nearby points

in the latent space. To avoid this problem, the input images encoded in the latent space of our model need to be correlated with some target properties that we are seeking to use to organize the latent space. Therefore, I added a supervised predictor to our model that maps the learned mean encoded latent vector $\mu_{enc}(x)$ to target properties $\hat{p_t} = \text{Pred}(\mu_{enc}(x))$. During training, the VAE-GAN network is trained with an additional task to minimize a regression loss (continuous) or a classification loss (discrete) between the predictions $\hat{p_t}$ and the true proprieties p_t

$$\begin{aligned} & \mathcal{L}_{pred} = \mathbb{E}_{x,p_t \sim p(x,p_t)}[\log p_{\text{pred}}(y|\mu_{enc}(x), y \in \{1, ..., P\})], \\ \text{and/or} \quad & \mathcal{L}_{pred} = \mathbb{E}_{x,p_t \sim p(x,p_t)}[\|p_t - \hat{p_t}\|^2]. \end{aligned} \tag{7.8}$$

One-Class Neural Network: One way to detect anomalies from the learned latent representations is to use a one-class classifier such as one-class SVM (OC-SVM) (Scholköpf and Smola, 2002). OC-SVM is a particular case of SVM in which a maximal margin decision surface is learned to optimally separate the data points from the origin in a reproducing kernel Hilbert space (see Figure 7.3). A drawback of this hybrid OC-SVM is that it can lead to sub-optimal solutions as OC-SVM cannot influence the latent representations learned by the encoder. Recently, it was demonstrated that the OC-SVM objective could be expressed using a fully-connected neural network with a single hidden layer with either a linear or sigmoid activation g and an output node with linear activation (Chalapathy et al., 2018). This One-Class Neural Network (OC-NN) is combined with the VAE-GAN to learn a decision surface that is customized for out-of-sample detection. Applied to $\mu_{enc}(x)$, the OC-NN objective is

$$\min_{w,V,r} \mathcal{L}_{oc} = \frac{1}{2}\|w\|^2 + \frac{1}{2}\|V\|^2 + \frac{1}{\nu}\,\mathbb{E}_{x \sim p(x)}[(max(0, r - \langle w, g(V\mu_{enc}(x))\rangle)] - r, \tag{7.9}$$

where w is the weight vector from hidden layer to the output node, V is the weight matrix from input to hidden layer, r is the bias of the decision surface, and $\nu \in [0, 1]$ controls the number of false positives in the training data. This non-convex objective is solved using an alternate minimization approach, where r is first fixed and w, V are optimized using backpropagation and

then the new values of w, V are used to compute an optimal value of r as the ν-quantile of $\langle w, g(V\mu_{enc}(x))\rangle$ (Chalapathy et al., 2018). In summary, the proposed semi-supervised VAE-GAN model is jointly trained according to Algorithm 1.

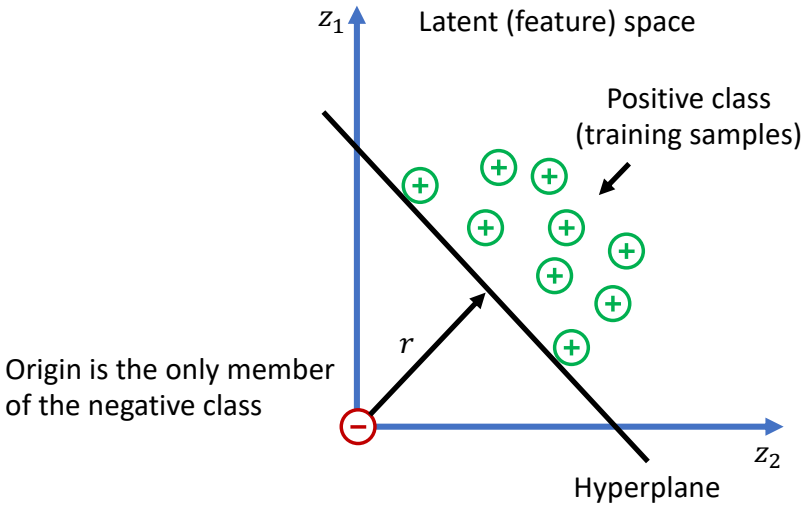


Figure 7.3: An illustration of one-class SVM (Scholköpf and Smola, 2002), which tackles the lack of negative class samples by maximizing the margin of a hyperplane to the origin. Green positive circles show the training samples, the red negative circle shows the anomaly sample, and the black line indicates the decision boundary captured by the one-class classifier.

Out-of-Sample Detection Scores: The trained semi-supervised VAE-GAN maps each test image x to a latent representation z that is sampled from $q(z|x) \sim \mathcal{N}(\mu_{enc}(x), \Sigma_{enc}(x))$, which is then passed through the decoder network to produce a reconstructed image. The degree of the anomaly of a new, unseen test sample can be characterized using the following scores:

(1) Similarity Score: As the model converges, it is expected that it has learned how to reconstruct data from the trained classes well but that it may not be good at reconstructing samples from the untrained classes leading to higher reconstruction errors. I utilize the following anomaly score based on assessing the reconstructed image quality using the learned similarity metric (Eqn. (7.4)):

$$S_{sim}(x) = -\frac{1}{N}\sum_{i=1}^{N} \log p(\text{Disc}_d(x)|z_i), \tag{7.10}$$

where N is the number of samples drawn from $q(z|x)$.

(2) One-Class Classifier Score: Anomalous samples are also detected in the latent space using the jointly trained OC-NN. The following anomaly score is defined based on the distance to the learned classifier decision surface:

$$S_{oc}(x) = \langle w, g(V\mu_{enc}(x))\rangle - r. \tag{7.11}$$

Algorithm 1 Summary of training the semi-supervised VAE-GAN

$\theta_{enc}, \theta_{dec}, \theta_{disc}, \theta_{pred}, w, V, r \leftarrow$ initialize network parameters

repeat

$X \leftarrow$ random mini-batch

$Y \leftarrow$ corresponding class labels

$P_t \leftarrow$ desired target properties

// Compute losses of the semi-supervised VAE-GAN model

$\mathcal{L}_{prior} \leftarrow KL(\mathcal{N}(\mu_{enc}(x), \Sigma_{enc}(x))||\mathcal{N}(0, I))$ Equation (7.2)

$\mathcal{L}_{recons}^{\text{Disc}_d} \leftarrow -\mathbb{E}_{q(z|x)}[\log p(\text{Disc}_d(x)|z)]$ Equation (7.4)

$\mathcal{L}_{gan} \leftarrow \mathbb{E}_{x,y\sim p(x,y)}[\log p_{\text{disc}}(y|x)] + \mathbb{E}_{\hat{x}\sim Gen}[\log(p_{\text{disc}}(y|\hat{x}, y = C+1)]$ Equation (7.5)

$\mathcal{L}_{pred} \leftarrow \mathbb{E}_{x,p_t\sim p(x,p_t)}[\log p_{\text{pred}}(y|\mu_{enc}(x), y \in \{1, ..., P\})]$ Equation (7.8)

$\mathcal{L}_{oc} \leftarrow \frac{1}{2}\|w\|^2 + \frac{1}{2}\|V\|^2 + \frac{1}{\nu}\,\mathbb{E}_{x\sim p(x)}[(max(0, r - \langle w, g(V\mu_{enc}(x))\rangle)] - r$ Equation (7.9)

// Update parameters according to computed gradients

$\theta_{enc} \overset{+}{\leftarrow} -\nabla_{\theta_{enc}}(\mathcal{L}_{prior} + \lambda\mathcal{L}_{recons}^{\text{Disc}_d} + \alpha\mathcal{L}_{pred} + \beta\mathcal{L}_{oc})$

$\theta_{dec} \overset{+}{\leftarrow} -\nabla_{\theta_{dec}}(\gamma\mathcal{L}_{recons}^{\text{Disc}_d} - \mathcal{L}_{gan})$

$\theta_{disc} \overset{+}{\leftarrow} -\nabla_{\theta_{disc}}(\mathcal{L}_{gan})$, $\theta_{pred} \overset{+}{\leftarrow} -\nabla_{\theta_{pred}}(\mathcal{L}_{pred})$

$w_{oc} \overset{+}{\leftarrow} -\nabla_w(\mathcal{L}_{oc})$, $V \overset{+}{\leftarrow} -\nabla_V(\mathcal{L}_{oc})$, $r = \nu$-quantile of$\langle w, g(V\mu_{enc}(X))\rangle$

until deadline

7.3 Experimental Results

7.3.1 Materials

The proposed methods were applied to the automatic QC of T1w MR brain images collected from the Baby Connectome Project (BCP) (Howell et al., 2019) and the Early Brain Development Study (EBDS) (Jha et al., 2018) datasets. Below, I briefly describe the BCP and EBDS datasets, the rating system employed to acquire the QC labels, and the minimal image processing performed.

BCP Dataset: 757 T1w MR brain images (1-5 years) were included from BCP, which is an ongoing study that is jointly conducted by investigators at the University of North Carolina at Chapel Hill and the University of Minnesota (Howell et al., 2019). The MRI scans were acquired at 2 different sites (Center for Magnetic Resonance Research at the University of Minnesota and the Biomedical Research Imaging Center at the University of North Carolina at Chapel Hill), each equipped with 3T Siemens Prisma scanners using a 32 channel head coil. Children up to 36 months of age were scanned unsedated during natural sleep while 36 months and older subjects were allowed to be scanned while awake based on caregiver preference. T1w structural pulse sequences were a 3D MP-RAGE (208 sagittal slices, TR = 400 ms or 1060 ms, TE = 2.24 ms, flip angle = 8°, resolution = $0.8 \times 0.8 \times 0.8$ mm^3, field of view = 256×256).

EBDS Dataset: 1697 T1w MR brain images (1-6 years) were included from large prospective studies of EBDS in healthy singletons and twins, administered at the University of North Carolina at Chapel Hill (Knickmeyer et al., 2008; Gilmore et al., 2010; Knickmeyer et al., 2016; Jha et al., 2018). MR images were acquired on both a Siemens Allegra and a Siemens Timm Trio head-only 3T scanner. Children were scanned unsedated during natural sleep, fitted with ear protection, and with their heads secured in a vacuum-fixation immobilization device before the scan. T1w structural pulse sequences were a 3D MP-RAGE (TR = 1900 ms, inversion time = 1100 ms, TE = 3.74 (Timm Trio) or 4.38 (Allegra) ms, flip angle = 7°, resolution = $1 \times 1 \times 1$ mm^3, field of view = 208×256).

Quality Control Scores: The combined dataset contains 2454 scans (1-6 years) with MRI variability, including scanner models, acquisition parameters, and image resolutions. MRI experts

assigned QC labels according to a 4-point rating scale (Blumenthal et al., 2002) (pass: $\{1, 2, 3\}$, fail: $\{4\}$). QC scores from two raters were available for the EBDS dataset, while only one QC score was available for the BCP dataset. Images with QC scores of 3 and 4 that showed disagreement between the two raters were excluded from the dataset. For 3 and 4 QC scores, scores from one of the evaluators were selected when a discrepancy was observed. As shown in the distribution of the obtained QC scores (Figure 7.4), the resulting dataset is highly imbalanced with only 3.5% scans failed image QC. Moreover, the QC-failed class is also inherently heterogeneous due to MRI artifact heterogeneity such as motion, noise, poor contrast, etc. (Figure 7.5). The proposed Anomaly-VAEGAN was trained on 80% of the QC-passed images while the remaining 20% and the QC-failed scans were used for testing.

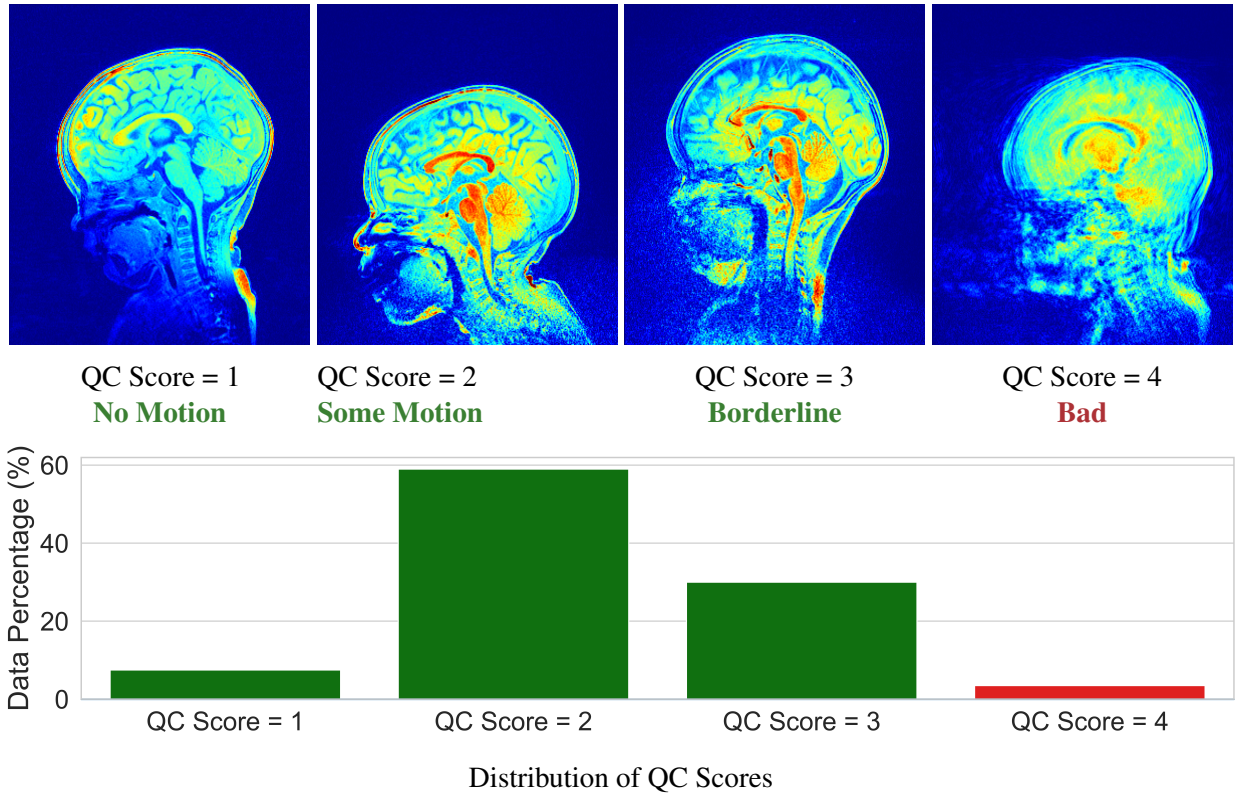


Figure 7.4: Examples from the used 4-point QC scoring system used in this chapter. The dataset is highly imbalanced as only 3.5% of scans failed QC (QC score = 4). Images are color-coded based on image intensities to highlight motion distortions.

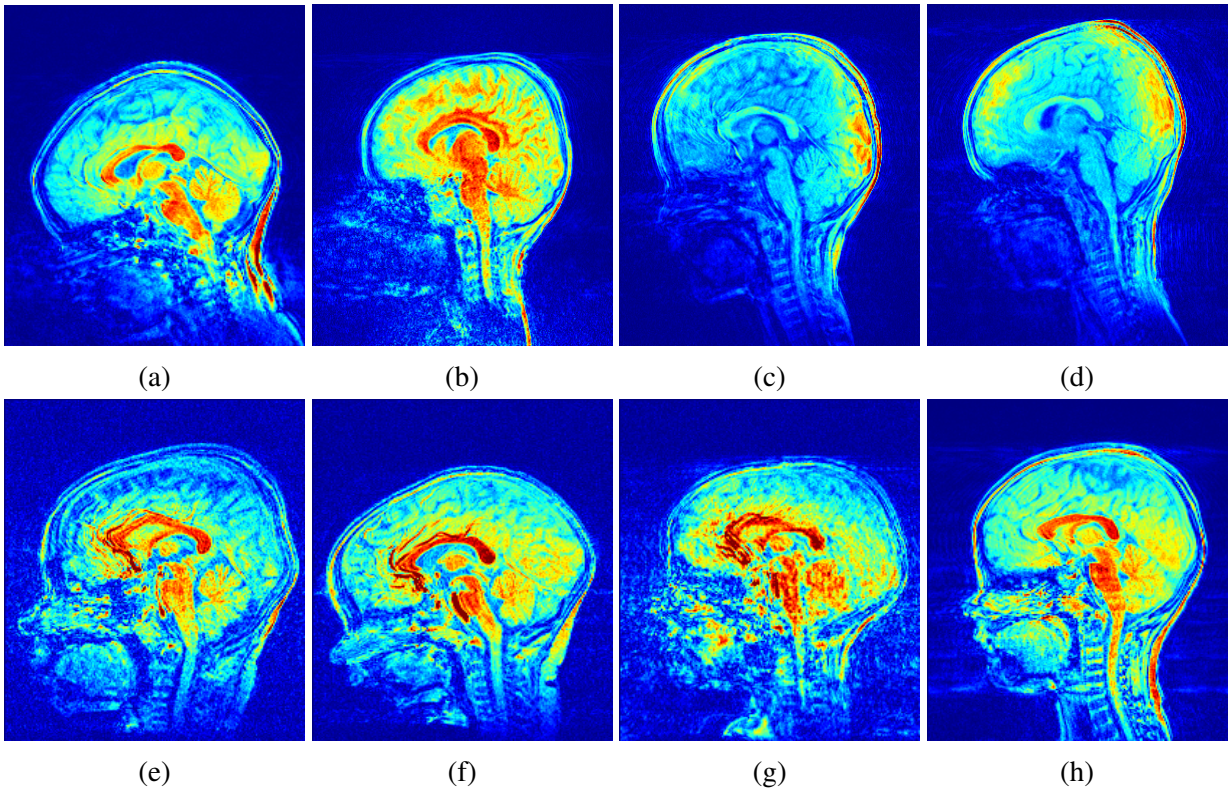


Figure 7.5: Additional examples from the QC-failed classes (QC score = 4), showing that minority classes in this imbalanced dataset are also inherently highly heterogeneous (QC failure due to numerous MRI artifacts such as motion, noise, poor contrast, etc.).

Image Preprocessing: To emulate a realistic setting for the proposed Anomaly-VAEGAN, a minimal image preprocessing was applied to enable fast within-session automatic MRI QC. No preprocessing was performed except resampling the input images to $128\times128\times128$ grid using a linear interpolation and linearly rescaling image intensities to the [-1,1] range.

7.3.2 Model Architecture

As shown in Table 7.1, the proposed Anomaly-VAEGAN design is based on a 3D version of DCGAN (Radford et al., 2015) with five convolution blocks (based on image size) and latent code dimension of 300. We used a shallow design for the predictor and the one-class classifier (single hidden layer). Weight decay regularization was used to constrain all kernel weights. Dropout layers (dropout rate $= 0.6$) and one-sided label smoothing (0.9 instead of 1) were additionally used for the discriminator. ADAM was used as the optimizer ($\eta = 10^{-4}$, batch size $= 8$, epochs $= 500$) with

network training hyperparameters: $\lambda = 1, \alpha = 20, \beta = 1, \gamma = 10^{-4}$, and $N = 20$. Hyperparameters selection was based on a manual search that was applied to optimize the network performance.

Encoder Network	**Decoder (Generator) Network**
$X \rightarrow (3 \times 3 \times 3)$ 32 Conv $(2 \times 2 \times 2) \downarrow$, BN, LR	$Z, Z_p \rightarrow (3 \times 3 \times 3) \times 512$ FC, BN, LR
$(3 \times 3 \times 3)$ 64 Conv $(2 \times 2 \times 2) \downarrow$, BN, LR	$(3 \times 3 \times 3)$ 512 Conv $(2 \times 2 \times 2) \uparrow$, BN, LR
$(3 \times 3 \times 3)$ 128 Conv $(2 \times 2 \times 2) \downarrow$, BN, LR	$(3 \times 3 \times 3)$ 256 Conv $(2 \times 2 \times 2) \uparrow$, BN, LR
$(3 \times 3 \times 3)$ 256 Conv $(2 \times 2 \times 2) \downarrow$, BN, LR	$(3 \times 3 \times 3)$ 128 Conv $(2 \times 2 \times 2) \uparrow$, BN, LR
$(3 \times 3 \times 3)$ 512 Conv $(2 \times 2 \times 2) \downarrow$, BN, LR	$(3 \times 3 \times 3)$ 64 Conv $(2 \times 2 \times 2) \uparrow$, BN, LR
300 FC $\rightarrow \mu_{enc}$; 300 FC $\rightarrow \log(\Sigma_{enc})$	$(3 \times 3 \times 3)$ 32 Conv $(2 \times 2 \times 2) \uparrow$, LR
	$(3 \times 3 \times 3)$ 1 Conv $(2 \times 2 \times 2) \uparrow$, tanh $\rightarrow \hat{X}$

Semi-Supervised Discriminator Network
$X, Y, \hat{X}, \hat{X}_p \rightarrow (3 \times 3 \times 3)$ 32 Conv $(2 \times 2 \times 2) \downarrow$, LR, DO
$(3 \times 3 \times 3)$ 64 Conv $(2 \times 2 \times 2) \downarrow$, BN, LR, DO
$(3 \times 3 \times 3)$ 128 Conv $(2 \times 2 \times 2) \downarrow$, BN, LR, DO
$(3 \times 3 \times 3)$ 256 Conv $(2 \times 2 \times 2) \downarrow$, BN, LR, DO
$(3 \times 3 \times 3)$ 512 Conv $(2 \times 2 \times 2) \downarrow$, BN, LR, DO
4 FC, Softmax $\rightarrow \hat{Y}$

Predictor Network	**One-Class Neural Network**
$\mu_{enc} \rightarrow$ 128 FC, LR	$\mu_{enc} \rightarrow$ 128 FC, Sigmoid
3 FC, Softmax $\rightarrow \hat{P}_t$	1 FC $\rightarrow \hat{Y}_{oc}$

Table 7.1: The network architecture of the proposed out-of-sample detection framework. Key: Conv: convolution; BN: Batch Normlization; LR: Leaky ReLU; FC: Fully-Connected; Do: Dropout; $\downarrow$ and $\uparrow$: represent down- and upsampling using strided convolution respectively; $\rightarrow$: represents network inputs and outputs. Network architecture is based on a 3D version of DCGAN (Radford et al., 2015) with a manual hyperparameters search applied to optimize network performance.

7.3.3 Performance Evaluation

The proposed Anomaly-VAEGAN was trained using images from QC-passed classes with class labels fed to the discriminator and the predictor networks. Upon convergence, the trained model was applied to both training and testing datasets to produce the two proposed anomaly scores S_{sim} and S_{oc} as well as their normalized sum $S_{sim} + S_{oc}$. The network performance on the testing set is evaluated using the area under the ROC curve (AUC) as well as the SEN and SPC values generated using the threshold maximizing their sum. To highlight the benefits obtained from measuring image similarity using a higher-level and invariant representation, the proposed methods were assessed by comparison against a VAE-based approach (Lu and Xu, 2018), where reconstruction scores were generated using typically used pixel-wise mean square error S_{mse}. To demonstrate that an online OC-NN can produce data representations that are customized for anomaly detection, I compare the proposed Anomaly-VAEGAN to hybrid approaches where a dimensionality reduction is first applied and then fed to a separate one-class classifier. In addition to features extracted using a VAE model, non-linear dimensionality reduction performed using a manifold learning method named Uniform Manifold Approximation and Projection (McInnes et al., 2018) was also investigated. A OC-SVM (Scholköpf and Smola, 2002) is utilized for anomaly detection using the extracted features. Finally, experiments were conducted to compare against the GAN-based approach in (Schlegl et al., 2017). These experiments failed due to learning instability when dealing with high-dimensional data, which is a problem that previously reported (Chen and Konukoglu, 2018)).

The normalized histograms of different out-of-sample scores generated for both training and testing data by the proposed Anomaly-VAEGAN and the VAE-based approach (Lu and Xu, 2018) are shown in the left column in Figure 7.6. Compared to the S_{mse} scores provided by the VAE-based method, the distributions produced by the proposed S_{sim} and S_{oc} scores provides better separation between the samples from the trained classes and samples from untrained classes. The best separation was obtained using $S_{sim} + S_{oc}$ sum score, which is also demonstrated by the corresponding ROC curves computed on the test set (see the right column of Figure 7.6). As shown in Table 7.2, results obtained using the proposed Anomaly-VAEGAN framework are superior to the

results obtained by compared methods, according to every performance metric. Particularly, the best performance was achieved using Anomaly-VAEGAN and a combined $S_{sim} + S_{oc}$ sum score (AUC = 0.93, SPC = 95%, and SEN = 81%), illustrating that assessment of test samples should be performed in both the original and in the latent spaces for optimal anomaly detection. Figure 7.7 provides a visualization of the learned encoding space by projecting the learned latent representations of the testing samples from the original 300 dimensions to a 2D grid using UMAP (McInnes et al., 2018). 2D visualization of the encoding space demonstrates that the latent representations learned by the proposed Anomaly-VAEGAN are better organized as compared to the latent codes produced by the VAE-based approach (Lu and Xu, 2018). This confirms that constraining the latent representation using the proposed predictor model allows for a better separation between the samples from untrained classes from the training data in the encoding space.

Method	AUC	SPC	SEN	Method	AUC	SPC	SEN
Anomaly-VAEGAN-$\mathbf{S_{sim} + S_{oc}}$	**0.93**	**0.95**	**0.81**	VAE-S_{mse}	0.63	0.93	0.42
Anomaly-VAEGAN-$\mathbf{S_{sim}}$	0.90	0.96	0.76	VAE+OC-SVM	-	0.83	0.57
Anomaly-VAEGAN-$\mathbf{S_{oc}}$	0.86	0.96	0.64	UMAP+OC-SVM	-	0.88	0.52

Table 7.2: Summary of the anomaly detection results using the mean AUC, SEN, and SPC scores.

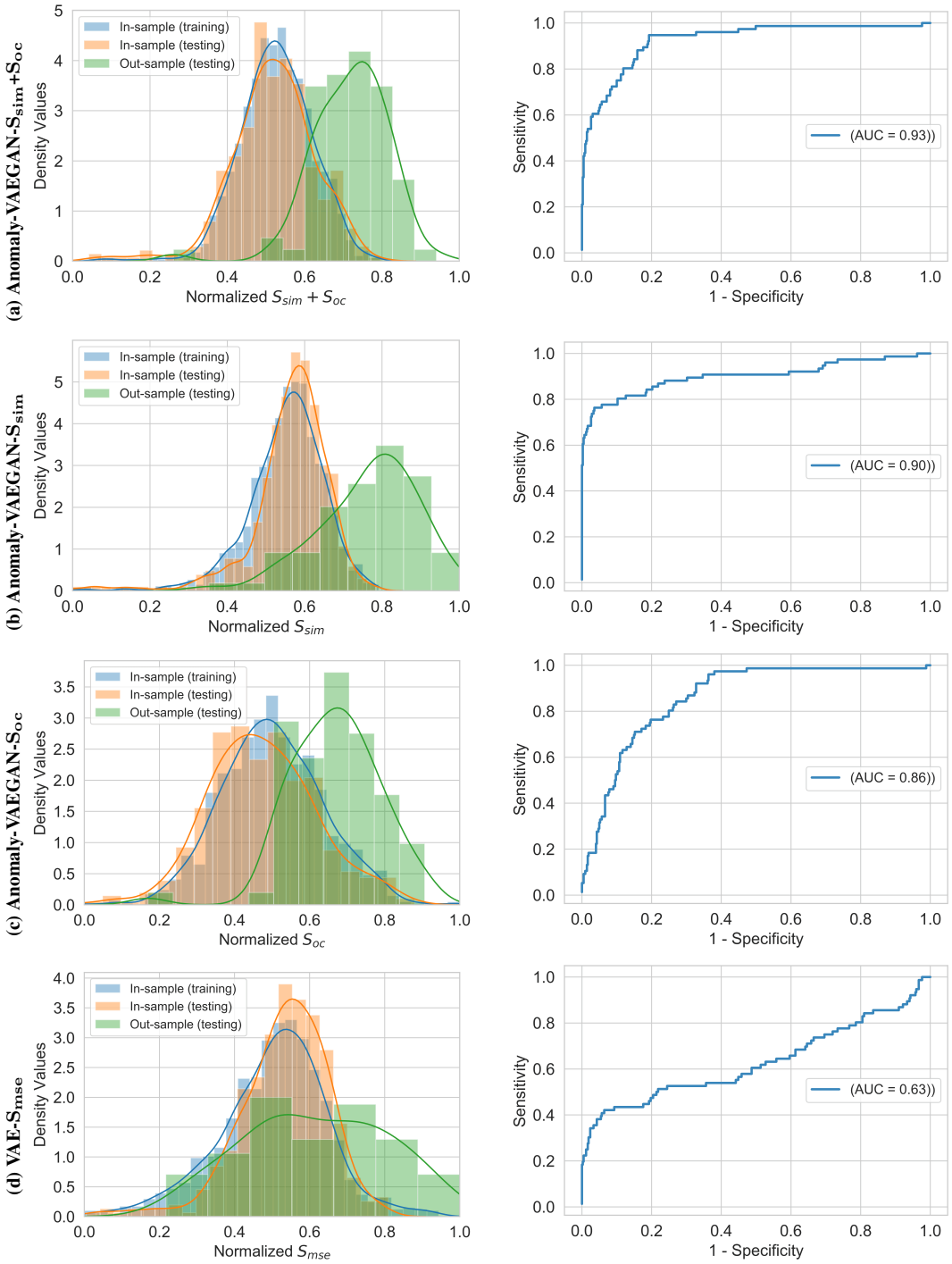
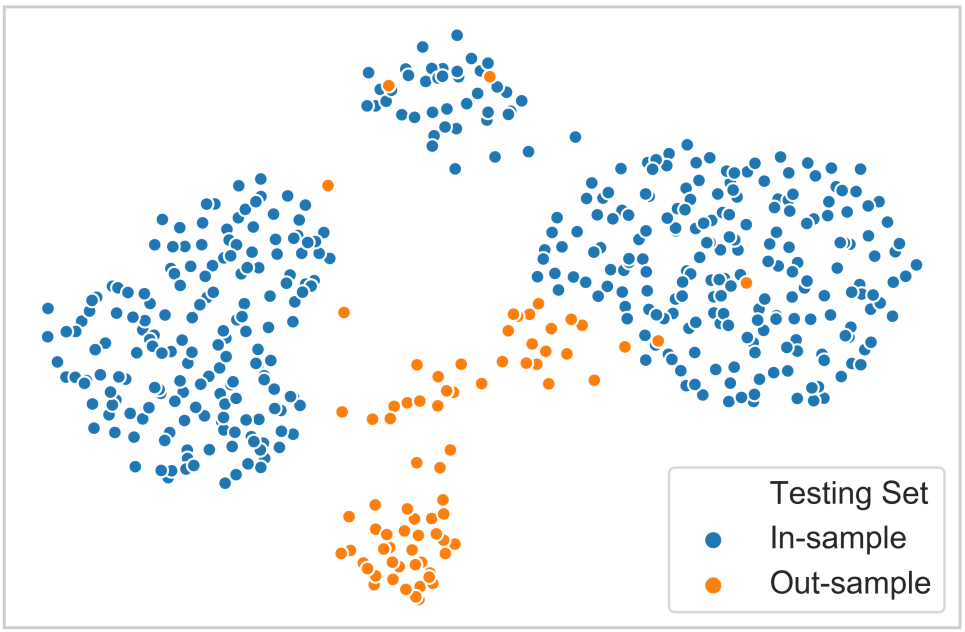
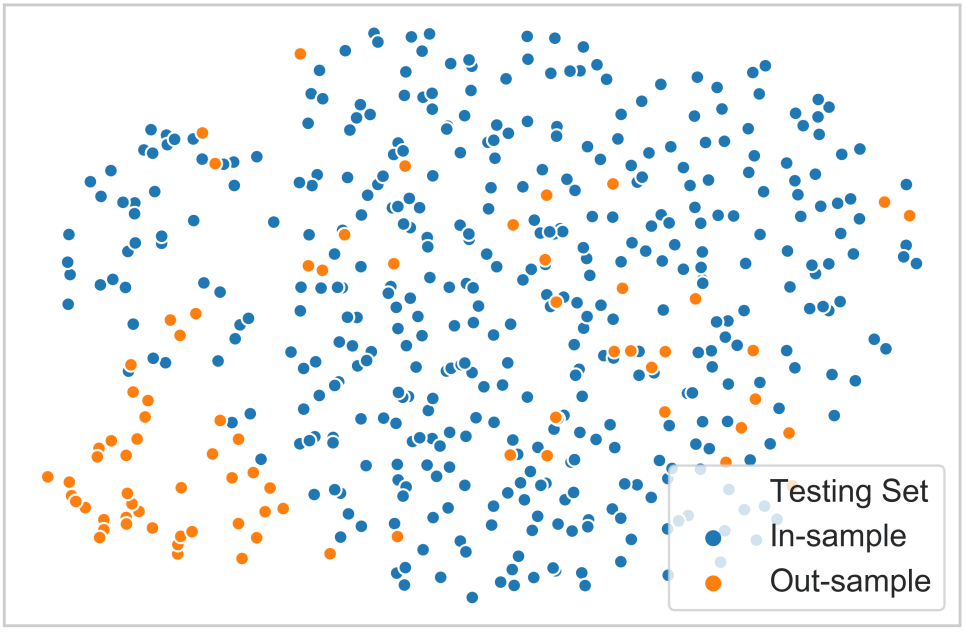


Figure 7.6: Left: normalized histograms of different out-of-sample scores for both training and testing data; right: the corresponding ROC curves computed on the test set. Out-of-sample detection is performed using the trained Anomaly-VAEGAN framework with (a) a combined similarity S_{sim} and one-class classifier S_{oc} scores, (b) using S_{sim} score only, and (c) using S_{oc} score only. Out-of-sample detection is also performed using (d) a VAE framework utilizing an element-wise reconstruction error (S_{mse}) (Lu and Xu, 2018). The best performance was achieved using the proposed Anomaly-VAEGAN framework with a combination of two proposed anomaly scores.



Anomaly-VAEGAN



VAE

Figure 7.7: 2D visualization of the learned latent representations (testing set) for qualitative assessment of the trained model. The dimensionality of the mean latent code vector μ_{enc} was reduced from the original 300 dimensions to a dimension of 2 using a UMAP projection (McInnes et al., 2018). Compared to the latent code learned by the VAE-based approach (Lu and Xu, 2018) (lower figure), the latent space learned by the proposed Anomaly-VAEGAN (lower figure) is better organized, and samples from untrained classes are better separated from the training data.

7.4 Summary

In this chapter I proposed a novel semi-supervised framework based on a 3D VAE-GAN for learning from imbalanced and heterogeneous datasets. The proposed deep generative approach can characterize the distribution of well-represented classes, a step enabling the identification of samples that were improbably generated from that distribution. The proposed framework relies on a high-level similarity metric and invariant representations learned by a semi-supervised discriminator to evaluate the generated images. The encoded latent representations were constrained according to user-defined properties through a jointly trained predictor network. Anomaly samples were detected using learned similarity scores and/or scores from an online one-class neural network.

The high performance of the proposed methods was confirmed via a novel application to the automatic QC of structural MR images. The experimental results have shown that the proposed Anomaly-VAEGAN framework can better assess the reconstruction of a new unseen test image by replacing element-wise distance metrics typically used with a learned similarity measure. Results demonstrated that extending VAE-GAN discriminator to semi-supervised learning enabled the network to capture high-level structure relevant to the data distribution, and it improved the overall network stability. I also showed that the proposed model was able to organize the learned latent space according to user-defined clustering, thereby enabling the effective utilization of non-imaging information. Additionally, I showed that using a one-class classifier that is jointly trained with the VAE-GAN model can influence the learned encoded representation, resulting in better detection of anomaly samples in the latent space. Finally, I pointed out that for optimal anomaly detection, a combined score obtained from assessing a test image in both the original and latent spaces needs to be utilized. The success of the proposed Anomaly-VAEGAN is not limited to the given application of automatic QC of structural MRI. Notably, the flexibility of the proposed methods would allow for a wide range of applications that involve learning from similar datasets. The early and accurate prediction of developmental disorders (e.g., ASD) is an instance where the proposed Anomaly-VAEGAN could be applied in low prevalence and highly heterogeneous settings.

CHAPTER 8: SUMMARY AND CONCLUDING DISCUSSION

In this last chapter I first review the contributions of my work in Section 8.1. I then discuss the limitations of my work in Section 8.2 followed by a discussion of future work in Section 8.3[1].

8.1 Summary of Contributions

This dissertation proposed methods to improve current deep learning-based neuroimaging pipelines including *(1)* extracting disease-specific, accurate and reliable local features from the cortical surface, *(2)* determining a parsimonious subset of cortical features for optimal disease diagnosis, *(3)* learning directly from a non-Euclidean high-dimensional feature space, *(4)* improving the robustness of multi-task multi-modal deep learning models, and *(5)* identifying anomalies in imbalanced and heterogeneous settings. The contributions of this dissertation are as follows.

[1] A *local measure of extra-axial cerebrospinal fluid were reliably extracted from structural MRI.*

Chapter 3 presented a novel framework for the local extraction of accurate local EA-CSF measurements along the brain cortex. The proposed framework is the first to address the problem of extracting local EA-CSF measurements in a way that is suitable for localized surface-based analysis. Hence, the proposed methods would allow neuroimaging labs to investigate the use of local EA-CSF in the early diagnostics of various NDDs.

The proposed processing combines probabilistic brain tissue segmentation, cortical surface reconstruction, and streamline-based local EA-CSF quantification. For streamline computation, I employ the vector field generated by solving a Laplacian PDE between the cortical surface and the outer CSF hull. An RK4 integration method was used to create the streamlines while

[1]The acronyms used in this chapter are defined in the list of abbreviations provided at the beginning of this dissertation.

achieving sub-voxel accuracy levels. Finally, the local EA-CSF was computed by accumulating the CSF probability along the generated streamlines. The proposed methods achieved an accurate and reliable automatic local EA-CSF extraction, as illustrated from the low CV scores computed in a scan/rescan human dataset. The quantitative analysis of local EA-CSF was used to study the early postnatal brain development in low-risk infants with no family history of ASD. The consistency in the local EA-CSF patterns across the examined time points further confirmed the stability of the proposed methods. Statistical analysis performed based on a linear mixed model revealed several novel findings that could not be observed if a global EA-CSF measure was utilized. In Chapter 4 the diagnostic capabilities of the extracted local EA-CSF measurements were investigated in the context of the early prediction of ASD diagnosis.

[2] An *early ASD diagnosis was predicted using cortical shape and cerebrospinal fluid measures.*

Chapter 4 described a complete framework to generate and combine measures of cortical surface anatomy from a structural MRI scan at 6 months to accurately predict ASD diagnosis at 24 months in HR infants. The obtained results improve upon previous approaches by downwardly extending the prediction to an earlier age (from 12 months to 6 months) using a conventional sMRI scan. The current results suggest the feasibility of pre-symptomatic detection of ASD in HR infants, affording the possibility of intervention in the first year of life, approximately two years before treatment occurs in the community for most children (Baio, 2012).

The proposed Deep-ASD framework utilizes a 6-month sMRI scan to generate four measures of infant brain anatomy, namely local EA-CSF, SCI, SA, and CT. Each of these measures was generated at more than 160,000 points across the cortical surface. Then, for feature reduction purposes and anatomical localization, the points were parcellated into predefined ROIs based on different anatomically- and functionally-based cortical surface atlases. The four measures were then entered into a fully cross-validated deep learning prediction algorithm to classify ASD, with particular attention paid to solving the class imbalance problem during the training phase. The prediction process was performed iteratively: first, each of the four anatomical measures

was entered independently as a single measure to predict ASD diagnosis. Then, the measures were entered in various combinations to find the optimal parsimonious combination of MRI brain features at 6 months that would improve the accuracy of ASD prediction at 24 months of age. Applied to a dataset collected from a prospective study of HR infants, the proposed Deep-ASD framework utilizing cortical shape and CSF measures was able to provide high prediction performance as reflected from the obtained PPV results, confirming the potential clinical relevance of the proposed methods for infants born into a high-risk family.

[3] *Convolutional neural networks were extended to learn from non-Euclidian cortical features.*

Chapter 5 presented a novel data-driven generalization of CNNs on non-Euclidean cortical surfaces. The proposed Surface-CNN can directly learn from high-dimensional cortical features, thus avoiding the need for employing a dimensionality reduction using a prior cortical parcellation. The flexibility of the proposed methods allows for various kinds of applications that involve learning from limited datasets of high-dimensional non-Euclidean features.

The proposed CNN extension relied on adaptive surface kernels that were defined using a locally constructed geodesic grid, which in turn allowed for extracting corresponding surface patches. A surface convolution block was then designed using novel surface convolution, pooling, and bottleneck layers. A Surface-CNN classifier was then implemented using such surface convolution blocks, allowing for efficient learning of discriminative surface features while the model complexity was still controlled. The high performance of the proposed Surface-CNN was demonstrated in accurately discriminating AD patients from CN patients based on CT and SA cortical measurements. The proposed Surface-CNN framework outperformed competing methods employing a variety of parcellation atlases or using traditional machine learning methods for dimensionality reduction. Further, Surface-CNN additionally outperformed a hybrid approach where image-based CNN was applied to 2D geometry-image representations, thus confirming the need for learning methods that avoid non-isometric transformations. In addition to its ability to provide accurate classification results, Surface-CNN demonstrated an

ability to highlight critical surface regions affected by the investigated disease. Surface-CNN's generalizability was shown in its ability to learn from thickness and curvature measures derived from different subcortical brain surfaces in the context of early prediction of ASD.

[4] *Multi-modal deep networks were robustly trained in a joint multi-task learning framework.*

Chapter 6 proposed a multi-modal block design as a generalization of convolutional or fully connected layers for learning from paired images and meta-data information. Such design allowed for the development of robust DNNs able to produce accurate results on a wide range of acquisition protocols and scanner models. The robustness of the proposed deep learning models would help increase the confidence of the clinical decisions of neuroimaging pipelines.

The proposed multi-modal block was defined using convolutional layer, fully connected layer, adapted instance normalization layer, and non-linear layer, together allowing for bi-directional information flow between images and meta-data. Using multi-modal blocks, an M-Net architecture was designed with meta-data available at different levels, permitting the network to encode deep shift-covariant and shift-invariant features in an efficient way. The proposed M-Net can provide both scalar and image outputs; thus, it was also utilized for multi-task learning using the proposed MM-Net architecture. MM-Net can also be adapted to ignore any of its outputs to obtain a network with a similar API to current classification or image-to-image networks. The robustness of the proposed methods was demonstrated in an sMRI tissue segmentation task where multi-modal 3D U-Net-based architectures were introduced with sequence parameters, geometrical information, and task-specific prior information defined as meta-data. When applied to a large and highly heterogeneous dataset, the proposed M-Net generated highly accurate tissue segmentation results within seconds and outperformed the competing baseline segmentation networks. The high performance of the proposed MM-Net was displayed in jointly classifying AD patients from healthy controls, predicting associated MMSE cognitive scores, and segmenting subcortical structures from T1w MRI and corresponding scan and patient demographic meta-data. The obtained results of the proposed MM-Net trained with

different multi-task loss strategies confirmed its ability to exploit the underlying similarities between correlated tasks to improve the generalization and accuracy of prediction models.

[5] A *deep generative model was designed for out-of-sample detection in imbalanced datasets.*

Chapter 7 presented a generative approach for semi-supervised out-of-sample detection in imbalanced and heterogeneous datasets. The proposed Anomaly-VAEGAN was able to characterize the distribution of well-represented classes; this learned distribution was then utilized to enable the identification of anomalous samples that were improbably generated from this distribution. The success of the proposed methods is not limited to the given application of automatic QC of sMRI, as the model flexibility allows for applications that involve learning in low prevalence and highly heterogeneous settings (e.g., NDDs neuroimaging datasets).

The proposed generative framework relied on a VAE-GAN architecture utilizing learned high-level similarity metric and invariant representations to asses the quality of the generated images. Hence, by avoiding the use of current pixel-wise reconstruction scores, the proposed Anomaly-VAEGAN was able to provide invariance to spatial variability or intensity variations expected when learning from noisy MRI datasets. The proposed Anomaly-VAEGAN was also adapted using a semi-supervised discriminator, which stabilized the network training and ensured a meaningful learned similarity metric. The encoded latent representations were organized according to user-defined properties through a jointly trained predictor network. Moreover, a one-class neural network was together trained with the VAE-GAN, which also influenced the learned representations. Anomalous samples were detected using high-level similarity scores and/or scores from the online one-class classifier. The high performance of the proposed Anomaly-VAEGAN was confirmed via a novel application to the automatic QC of T1w MRI. The proposed methods were able to robustly detect images with insufficient quality in different settings, including scanner models, acquisition parameters, and subject age. In particular, the proposed Anomaly-VAEGAN framework with a combination of two introduced anomaly scores provided the best accuracy results as compared to the competing baseline methods. The

availability of such a fast within-session automatic QC tool would allow MR technicians to assess whether an acquired T1w image is appropriate for quantitative neuroimaging workflows (structural, functional, or diffusion connectivity) or whether a repeat T1w scan is required.

[6] *Publicly available source codes and trained deep learning models.*

An initial version of the source codes and the trained models of the proposed methods in this dissertation was made publicly available at `https://github.com/MahmoudMostapha/PhD-Dissertation/` (see Figure 8.1 for the current repository organization).

```
PhD-Dissertation
       |--> Chapter3/
              |--> Code/
                     |--> CreateOuterSurface/
                            |--> scr/
                                   |--> CreateOuterSurface.cxx
                                   |--> CMakeLists.txt

                     .
                     .
                     |--> SolveLaplacePDE/
                     |--> EstimateLocalEACSF/
       |--> Chapter4/
              |--> Code/
                     |--> Combining_All_predictors/
                            |--> Fit_Deep-ASD.py
              |--> Models/
                     |--> Combining_All_predictors/
                            |--> CT_0.h5
                            |--> EACSF_0.h5
                            .
                            .
                            |--> Best_weights.pkl
       .
       .
       |--> Chapter7/
              |--> Code/
                     |--> Anomaly-VAEGAN/
                            |--> AugmentData.py
                            |--> Train.py
                            .
                            .
              |--> Models/
                     |--> Anomaly-VAEGAN/
                            |--> Trained_Models_Link.md
```

Figure 8.1: The current organization of my PhD Dissertation Github repository.

8.2 Limitations

This section provides a summary of the main limitations of each of the proposed methods while Section 8.3 discusses future directions to ameliorate such limitations.

[1] *Local extra-axial cerebrospinal fluid extraction*

The quantification of local EA-CSF measurements relies on streamlines that were generated based on solving an isotropic Laplace PDE between inner and outer surfaces on a created voxel grid. The isotropic PDE solution is purely based on the boundaries implied by the inner and outer surfaces. Hence, the generated streamlines were not constrained to areas containing CSF and also failed to account for partial volume effects that lie between the two boundaries.

[2] *Cortical measures combination for ASD prediction*

The proposed Deep-ASD framework generated, summarized, and optimally combined 6 months cortical measurements to predict ASD diagnosis at 24 months in HR infants accurately. Although the prediction algorithm was rigorously tested using 10-fold cross-validation, the current study did not include an independent HR sample for replication. Moreover, there are still open questions on the usefulness of the proposed Deep-ASD framework when applied to the general population as currently available ASD datasets do not provide a representative sample.

[3] *Convolutional learning on cortical brain surfaces*

The proposed Surface-CNN relies on surface convolution operations that utilize a geodesic coordinate system computed at each surface point. However, the current spatial implementation of surface convolution operation requires updating local neighborhood information while traversing surface points, which significantly slows down the execution of the training process.

[4] *Robust multi-modal multi-task deep representation learning*

Due to the lack of publicly available manually segmented brain MRI datasets, the proposed M-Nets relied on pseudo-ground-truth annotations generated by current neuroimaging processing

pipelines (e.g., FSL or FreeSurfer) to train the models and also to evaluate their performance. The proposed methods attempted to handle learning from noisy labels by modeling the heteroscedastic aleatoric uncertainty. Nevertheless, it is still not clear how to evaluate the success of such a model with currently available imperfect ground truth labels.

[5] *Deep generative anomaly detection in heterogeneous datasets*

The proposed Anomaly-VAEGAN architecture relied on learned similarity scores and scores from an online one-class neural network to robustly detect anomalous samples. However, the stability and performance of the proposed extended VAE-GAN model highly depends on a costly hyperparameter tuning procedure to balance the losses of various network components during training as well as the effect of different anomaly scores during testing. Hence, there is a need to simplify the current model and automate hyperparameter optimization to guarantee high performance of the proposed Anomaly-VAEGAN when deployed to new applications.

8.3 Future Work

There are several future research directions for improving current methods and for employing them in new clinical applications. Some of these research directions are reviewed in this section.

8.3.1 Computational and Design Issues

[1] The proposed local EA-CSF quantification presented in Chapter 3 can be improved by replacing the isotropic Laplace PDE with an anisotropic version (Joshi et al., 2018) in which the diffusion coefficient is varied spatially in proportion to the fraction of CSF in each voxel. This allows for generating streamlines that follows a realistic path through areas containing CSF, leading to improvements in the accuracy and reliability of the extracted local EA-CSF measurements.

[2] The proposed Deep-ASD prediction framework presented in Chapter 4 can be extended to dynamically learn to fuse complementary information provided by different cortical measurements describing the infant brain anatomy. This would enable the inclusion of additional

cortical measures while allowing the network to learn useful interactions between various cortical measures and maintaining a controlled model complexity needed for learning from limited datasets. Moreover, learning a spatially varying fusion map that quantifies the relative importance of each cortical features across different spatial locations would allow for a better understanding of the underlying processes related to the investigated disease (e.g., ASD).

[3] The speed of the Surface-CNN architecture proposed in Chapter 5 can be improved on GPUs by switching to frequency-domain surface convolution implementation based on the algorithm. Future work will include investigating learning from cortical features on alternative shape representations to mesh surfaces such as point clouds (Qi et al., 2017), graphs (Wang et al., 2019), and implicit surfaces (Park et al., 2019) to alleviate some of the current complications.

[4] The segmentation M-Nets presented in Chapter 6 relied on a large dataset with pseudo-ground-truth labels to improve the model robustness. However, when manually segmented data becomes available, the proposed M-Nets can be extended in a two-stream approach where attention models can be implemented to reduce the influence of back-propagated gradients produced by inaccurate labels, resulting in a network that is robust to unclean data (Min et al., 2019). This would allow for further improvements in the performance of the trained M-Nets while maintaining their robustness in a wide range of MRI acquisition protocols.

[5] The proposed Anomaly-VAEGAN performance can be optimized by implementing an automatic Bayesian hyperparameter tuning (Dewancker et al., 2015) and by conducting a full model analysis to simplify the network architecture. Directions to further improve the network performance include using a more flexible prior (e.g., Gaussian mixture model) (Dilokthanakul et al., 2016), constraining latent code using Siamese networks (Koch et al., 2015) and using a more expressive decoder (e.g., autoregressive model) (Van den Oord et al., 2016). Such proposed model extensions would increase the model flexibility, allowing for more accurate characterization of the training data and better subsequent anomaly detection.

8.3.2 Clinical Applications

[1] An altered flow of CSF has been shown to impair the clearance of pathogenic inflammatory proteins (e.g.,amyloid-β) involved in neurodegenerative conditions such as Alzheimer's Disease (Tarasoff-Conway et al., 2015). Hence, the local EA-CSF extraction framework presented in Chapter 3 can be used to provide useful biomarkers for the early detection of such diseases.

[2] The proposed Deep-ASD framework presented in Chapter 4 can be extended to investigate other cortical measures. For example, the predictive ability of quantitative cortical folding descriptors such as local GI (Lyu et al., 2018) was not investigated yet in the context of early prediction of ASD. The generalizability of the proposed techniques would allow for investigating their ability to detect other NDDs in the HR population provided by the IBIS dataset.

[3] The proposed techniques presented in Chapter 5 for convolutional learning on surfaces can be utilized to introduce an extension to deep convolutional GANs. Hence, such a GAN extension can produce synthetic features on cortical surfaces, which can help deal with the class imbalance problem that usually exists when learning from neurodevelopmental MRI datasets (e.g., ASD).

[4] The MM-Net proposed in Chapter 6 can be extended to predict the different stages of AD progression (e.g., healthy aging, early MCI, late MCI, mild AD, moderate AD, severe AD) (Sperling et al., 2011). The proposed MM-Net can also be employed in more challenging clinical applications, including the early prediction of ASD. In particular, complementary data provided by multi-modal MRI (e.g., structural, functional, diffusion connectivity) and non-imaging information (e.g., scan parameters, behavioral assessments, demographic information) can be utilized by the proposed MM-Net to predict ASD categorical outcome as well as continuous dimensions of disease-related scores (e.g., ADOS, ADI-R, Mullen). The ability to predict later ASD-related scores of cognition and behavior at an individual level would facilitate the development and implementation of targeted, individualized, presymptomatic interventions.

[5] The generalizability of the proposed Anomaly-VAEGAN in Chapter 7 can be further demonstrated in detecting anomalies in imbalanced and heterogeneous datasets. Notably, learning from the low prevalence and highly varied NDDs datasets is an instance where the proposed Anomaly-VAEGAN can be useful to parse the heterogeneity of such complex disorders. Moreover, efforts have been initiated to utilize the proposed Anomaly-VAEGAN to detect rare fetal abnormalities (e.g., skull, brain, face, neck, spine) (Van den Veyver, 2019) using first-trimester ultrasound produced by a single scan of a portable probe during a screening session.

BIBLIOGRAPHY

Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1):63–72.

Ad-Dabbagh, Y., Lyttelton, O., Muehlboeck, J., Lepage, C., Einarson, D., Mok, K., Ivanov, O., Vincent, R., Lerch, J., Fombonne, E., et al. (2006). The civet image-processing environment: a fully automated comprehensive pipeline for anatomical neuroimaging research. In *Proceedings of the 12th annual meeting of the organization for human brain mapping*, volume 2266. Florence, Italy.

Aganj, I., Sapiro, G., Parikshak, N., Madsen, S. K., and Thompson, P. M. (2009). Measurement of cortical thickness from mri by minimum line integrals on soft-classified tissue. *Human brain mapping*, 30(10):3188–3199.

Ajayi-Obe, M., Saeed, N., Cowan, F., Rutherford, M., and Edwards, A. (2000). Reduced development of cerebral cortex in extremely preterm infants. *The Lancet*, 356(9236):1162–1163.

Akkus, Z., Ali, I., Sedlar, J., Kline, T. L., Agrawal, J. P., Parney, I. F., Giannini, C., and Erickson, B. J. (2016). Predicting 1p19q chromosomal deletion of low-grade gliomas from mr images using deep learning. *arXiv preprint arXiv:1611.06939*.

Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., and Erickson, B. J. (2017). Deep learning for brain mri segmentation: state of the art and future directions. *Journal of digital imaging*, 30(4):449–459.

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein gan. *arXiv preprint arXiv:1701.07875*.

Armstrong, E., Schleicher, A., Omran, H., Curtis, M., and Zilles, K. (1995). The ontogeny of human gyrification. *Cerebral cortex*, 5(1):56–63.

Ashburner, J. and Friston, K. J. (2000). Voxel-based morphometrythe methods. *Neuroimage*, 11(6):805–821.

Association, A. P. (2000). *Diagnostic criteria from dsM-iV-tr*. American Psychiatric Pub.

Association, A. P. et al. (2010). American psychiatric association dsm-5 development. *Proposed Revisions/Somatic Symptom Disorders/J*, 2.

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011). A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage*, 54(3):2033–2044.

Awate, S. P., Yushkevich, P. A., Song, Z., Licht, D. J., and Gee, J. C. (2010). Cerebral cortical folding analysis with multivariate modeling and testing: studies on gender differences and neonatal development. *Neuroimage*, 53(2):450–459.

Aylward, E. H., Liu, D., Nopoulos, P. C., Ross, C. A., Pierson, R. K., Mills, J. A., Long, J. D., Paulsen, J. S., Investigators, P.-H., of the Huntington Study Group, C., et al. (2012). Striatal volume contributes to the prediction of onset of huntington disease in incident cases. *Biological psychiatry*, 71(9):822–828.

Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization. *arXiv preprint arXiv:1607.06450*.

Baio, J. (2012). Prevalence of autism spectrum disorders: Autism and developmental disabilities monitoring network, 14 sites, united states, 2008. morbidity and mortality weekly report. surveillance summaries. volume 61, number 3. *Centers for Disease Control and Prevention*.

Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V. (2018). An unsupervised learning model for deformable medical image registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9252–9260.

Balasubramanian, M., Polimeni, J. R., and Schwartz, E. L. (2008). Exact geodesics and shortest paths on polyhedral surfaces. *IEEE transactions on pattern analysis and machine intelligence*, 31(6):1006–1016.

Baltrušaitis, T., Ahuja, C., and Morency, L.-P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):423–443.

Baur, C., Wiestler, B., Albarqouni, S., and Navab, N. (2018). Deep autoencoding models for unsupervised anomaly segmentation in brain mr images. *arXiv preprint arXiv:1804.04488*.

Baxter, J. (1997). A bayesian/information theoretic model of learning to learn via multiple task sampling. *Machine learning*, 28(1):7–39.

Bengio, Y., Laufer, E., Alain, G., and Yosinski, J. (2014). Deep generative stochastic networks trainable by backprop. In *International Conference on Machine Learning*, pages 226–234.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Bermudez, C., Plassard, A. J., Davis, L. T., Newton, A. T., Resnick, S. M., and Landman, B. A. (2018). Learning implicit brain mri manifolds with deep learning. In *Medical Imaging 2018: Image Processing*, volume 10574, page 105741L. International Society for Optics and Photonics.

Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., and West, M. (2007). Generative or discriminative? getting the best of both worlds. *Bayesian statistics*, 8(3):3–24.

Blumenthal, J. D., Zijdenbos, A., Molloy, E., and Giedd, J. N. (2002). Motion artifact in magnetic resonance imaging: implications for automated analysis. *Neuroimage*, 16(1):89–92.

Boscaini, D., Masci, J., Melzi, S., Bronstein, M. M., Castellani, U., and Vandergheynst, P. (2015). Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks. In *Computer Graphics Forum*, volume 34, pages 13–23. Wiley Online Library.

Brown, C. J., Miller, S. P., Booth, B. G., Poskitt, K. J., Chau, V., Synnes, A. R., Zwicker, J. G., Grunau, R. E., and Hamarneh, G. (2015). 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*, pages 69–76. Springer.

Bryson, S. E., Zwaigenbaum, L., McDermott, C., Rombough, V., and Brian, J. (2008). The autism observation scale for infants: scale development and reliability data. *Journal of autism and developmental disorders*, 38(4):731–738.

Bussu, G., Jones, E. J., Charman, T., Johnson, M. H., Buitelaar, J., Team, B., et al. (2018). Prediction of autism at 3 years from behavioural and developmental measures in high-risk infants: a longitudinal cross-domain classifier analysis. *Journal of autism and developmental disorders*, 48(7):2418–2433.

Caruna, R. (1993). Multitask learning: A knowledge-based source of inductive bias. In *Machine Learning: Proceedings of the Tenth International Conference*, pages 41–48.

Causon, D. and Mingham, C. (2010). *Introductory finite difference methods for PDEs*. Bookboon.

Chalapathy, R., Menon, A. K., and Chawla, S. (2018). Anomaly detection using one-class neural networks. *arXiv preprint arXiv:1802.06360*.

Chawarska, K., Shic, F., Macari, S., Campbell, D. J., Brian, J., Landa, R., Hutman, T., Nelson, C. A., Ozonoff, S., Tager-Flusberg, H., et al. (2014). 18-month predictors of later outcomes in younger siblings of children with autism spectrum disorder: a baby siblings research consortium study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(12):1317–1327.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.

Chawla, N. V., Lazarevic, A., Hall, L. O., and Bowyer, K. W. (2003). Smoteboost: Improving prediction of the minority class in boosting. In *European conference on principles of data mining and knowledge discovery*, pages 107–119. Springer.

Chen, H., Duan, X., Liu, F., Lu, F., Ma, X., Zhang, Y., Uddin, L. Q., and Chen, H. (2016). Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivitya multi-center study. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 64:1–9.

Chen, X. and Konukoglu, E. (2018). Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders. *arXiv preprint arXiv:1806.04972*.

Chen, Z., Badrinarayanan, V., Lee, C.-Y., and Rabinovich, A. (2017). Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. *arXiv preprint arXiv:1711.02257*.

Chung, M. K., Robbins, S. M., Dalton, K. M., Davidson, R. J., Alexander, A. L., and Evans, A. C. (2005). Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage*, 25(4):1256–1265.

Chung, M. K., Worsley, K. J., Nacewicz, B. M., Dalton, K. M., and Davidson, R. J. (2010). General multivariate linear modeling of surface shapes using surfstat. *Neuroimage*, 53(2):491–505.

Cidav, Z., Munson, J., Estes, A., Dawson, G., Rogers, S., and Mandell, D. (2017). Cost offset associated with early start denver model for children with autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, 56(9):777–783.

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

Clark, M. L. E., Vinen, Z., Barbaro, J., and Dissanayake, C. (2018). School age outcomes of children diagnosed early and later with autism spectrum disorder. *Journal of autism and developmental disorders*, 48(1):92–102.

Clevert, D.-A., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*.

Cohen, O., Zhu, B., and Rosen, M. S. (2018). Mr fingerprinting deep reconstruction network (drone). *Magnetic resonance in medicine*, 80(3):885–894.

Dai, Y., Shi, F., Wang, L., Wu, G., and Shen, D. (2013). ibeat: a toolbox for infant brain magnetic resonance image processing. *Neuroinformatics*, 11(2):211–225.

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., Donaldson, A., and Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: the early start denver model. *Pediatrics*, 125(1):e17–e23.

de Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M., and Išgum, I. (2019). A deep learning framework for unsupervised affine and deformable image registration. *Medical image analysis*, 52:128–143.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

Deo, R. C. (2015). Machine learning in medicine. *Circulation*, 132(20):1920–1930.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980.

Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 53(1):1–15.

Dewancker, I., McCourt, M., and Clark, S. (2015). Bayesian optimization primer.

Dickerson, B. C., Bakkour, A., Salat, D. H., Feczko, E., Pacheco, J., Greve, D. N., Grodstein, F., Wright, C. I., Blacker, D., Rosas, H. D., et al. (2008). The cortical signature of alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild ad dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral cortex*, 19(3):497–510.

Dijkstra, E. W. et al. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271.

Dilokthanakul, N., Mediano, P. A., Garnelo, M., Lee, M. C., Salimbeni, H., Arulkumaran, K., and Shanahan, M. (2016). Deep unsupervised clustering with gaussian mixture variational autoencoders. *arXiv preprint arXiv:1611.02648*.

Dolz, J., Desrosiers, C., and Ayed, I. B. (2018). 3d fully convolutional networks for subcortical segmentation in mri: A large-scale study. *NeuroImage*, 170:456–470.

Dora, L., Agrawal, S., Panda, R., and Abraham, A. (2017). State-of-the-art methods for brain tissue segmentation: A review. *IEEE reviews in biomedical engineering*, 10:235–249.

Doshi-Velez, F. and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Dosovitskiy, A. and Brox, T. (2016). Generating images with perceptual similarity metrics based on deep networks. In *Advances in Neural Information Processing Systems*, pages 658–666.

Du, A., Schuff, N., Amend, D., Laakso, M., Hsu, Y., Jagust, W., Yaffe, K., Kramer, J., Reed, B., Norman, D., et al. (2001). Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 71(4):441–447.

Duong, L., Cohn, T., Bird, S., and Cook, P. (2015). Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, volume 2, pages 845–850.

El Azami, M., Hammers, A., Jung, J., Costes, N., Bouet, R., and Lartizien, C. (2016). Detection of lesions underlying intractable epilepsy on t1-weighted mri as an outlier detection problem. *PloS one*, 11(9):e0161498.

Elison, J. T., Paterson, S. J., Wolff, J. J., Reznick, J. S., Sasson, N. J., Gu, H., Botteron, K. N., Dager, S. R., Estes, A. M., Evans, A. C., et al. (2013). White matter microstructure and atypical visual orienting in 7-month-olds at risk for autism. *American Journal of Psychiatry*, 170(8):899–908.

Elsayed, G., Shankar, S., Cheung, B., Papernot, N., Kurakin, A., Goodfellow, I., and Sohl-Dickstein, J. (2018). Adversarial examples that fool both computer vision and time-limited humans. In *Advances in Neural Information Processing Systems*, pages 3914–3924.

Emerson, R. W., Adams, C., Nishino, T., Hazlett, H. C., Wolff, J. J., Zwaigenbaum, L., Constantino, J. N., Shen, M. D., Swanson, M. R., Elison, J. T., et al. (2017). Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. *Science translational medicine*, 9(393):eaag2882.

Erhan, D., Bengio, Y., Courville, A., and Vincent, P. (2009). Visualizing higher-layer features of a deep network. *University of Montreal*, 1341(3):1.

Estes, A., Zwaigenbaum, L., Gu, H., John, T. S., Paterson, S., Elison, J. T., Hazlett, H., Botteron, K., Dager, S. R., Schultz, R. T., et al. (2015). Behavioral, cognitive, and adaptive development in infants with autism spectrum disorder in the first 2 years of life. *Journal of neurodevelopmental disorders*, 7(1):24.

Fischl, B. (2012). Freesurfer. *Neuroimage*, 62(2):774–781.

Fischl, B. and Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, 97(20):11050–11055.

Fischl, B., Liu, A., and Dale, A. M. (2001). Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE transactions on medical imaging*, 20(1):70–80.

Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B. T., Mohlberg, H., Amunts, K., and Zilles, K. (2007). Cortical folding patterns and predicting cytoarchitecture. *Cerebral cortex*, 18(8):1973–1980.

Fishbaugh, J., Durrleman, S., Piven, J., and Gerig, G. (2012). A framework for longitudinal data analysis via shape regression. In *Proceedings of SPIE*, volume 8314. NIH Public Access.

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., Collins, D. L., Group, B. D. C., et al. (2011). Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage*, 54(1):313–327.

Fonov, V. S., Janke, A., Caramanos, Z., Arnold, D. L., Narayanan, S., Pike, G. B., and Collins, D. L. (2010). Improved precision in the measurement of longitudinal global and regional volumetric changes via a novel mri gradient distortion characterization and correction technique. In *International Workshop on Medical Imaging and Virtual Reality*, pages 324–333. Springer.

Free, S., Sisodiya, S., Cook, M., Fish, D., and Shorvon, S. (1996). Three-dimensional fractal analysis of the white matter surface from magnetic resonance images of the human brain. *Cerebral Cortex*, 6(6):830–836.

Gardner, M. W. and Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)a review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14-15):2627–2636.

Garland, M. and Heckbert, P. S. (1997). Surface simplification using quadric error metrics. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 209–216. ACM Press/Addison-Wesley Publishing Co.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., and Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal mri study. *Nature neuroscience*, 2(10):861–863.

Gilmore, J. H., Schmitt, J. E., Knickmeyer, R. C., Smith, J. K., Lin, W., Styner, M., Gerig, G., and Neale, M. C. (2010). Genetic and environmental contributions to neonatal brain structure: a twin study. *Human brain mapping*, 31(8):1174–1182.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., and Petersen, S. E. (2014). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral cortex*, 26(1):288–303.

Gouttard, S., Styner, M., Prastawa, M., Piven, J., and Gerig, G. (2008). Assessment of reliability of multi-site neuroimaging via traveling phantom study. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 263–270. Springer.

Green, J., Charman, T., Pickles, A., Wan, M. W., Elsabbagh, M., Slonims, V., Taylor, C., McNally, J., Booth, R., Gliga, T., et al. (2015). Parent-mediated intervention versus no intervention for infants at high risk of autism: a parallel, single-blind, randomised trial. *The Lancet Psychiatry*, 2(2):133–140.

Gu, X., Gortler, S. J., and Hoppe, H. (2002). Geometry images. *ACM Transactions on Graphics (TOG)*, 21(3):355–361.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved training of wasserstein gans. In *Advances in neural information processing systems*, pages 5767–5777.

Guo, M., Haque, A., Huang, D.-A., Yeung, S., and Fei-Fei, L. (2018). Dynamic task prioritization for multitask learning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 270–287.

Haidar, H. and Soul, J. S. (2006). Measurement of cortical thickness in 3d brain mri data: validation of the laplacian method. *Journal of Neuroimaging*, 16(2):146–153.

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., and Bing, G. (2017). Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73:220–239.

Hallahan, B., Daly, E., McAlonan, G., Loth, E., Toal, F., O'brien, F., Robertson, D., Hales, S., Murphy, C., Murphy, K., et al. (2009). Brain morphometry volume in autistic spectrum disorder: a magnetic resonance imaging study of adults. *Psychological Medicine*, 39(2):337–346.

Hartberg, C. B., Lange, E. H., Lagerberg, T. V., Haukvik, U. K., Andreassen, O. A., Melle, I., and Agartz, I. (2018). Cortical thickness, cortical surface area and subcortical volumes in schizophrenia and bipolar disorder patients with cannabis use. *European Neuropsychopharmacology*, 28(1):37–47.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., and Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31.

Hazlett, H. C., Gu, H., McKinstry, R. C., Shaw, D. W., Botteron, K. N., Dager, S. R., Styner, M., Vachet, C., Gerig, G., Paterson, S. J., et al. (2012). Brain volume findings in 6-month-old infants at high familial risk for autism. *American Journal of Psychiatry*, 169(6):601–608.

Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., Elison, J. T., Swanson, M. R., Zhu, H., Botteron, K. N., et al. (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature*, 542(7641):348.

Hazlett, H. C., Poe, M. D., Gerig, G., Styner, M., Chappell, C., Smith, R. G., Vachet, C., and Piven, J. (2011). Early brain overgrowth in autism associated with an increase in cortical surface area before age 2 years. *Archives of general psychiatry*, 68(5):467–476.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507.

Hoo-Chang, S., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5):1285.

Hoppe, H. (1999). New quadric metric for simplifiying meshes with appearance attributes. In *Proceedings of the conference on Visualization'99: celebrating ten years*, pages 59–66. IEEE Computer Society Press.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257.

Howell, B. R., Styner, M. A., Gao, W., Yap, P.-T., Wang, L., Baluyot, K., Yacoub, E., Chen, G., Potts, T., Salzwedel, A., et al. (2019). The unc/umn baby connectome project (bcp): an overview of the study design and protocol development. *NeuroImage*, 185:891–905.

Hu, J., Shen, L., and Sun, G. (2017). Squeeze-and-excitation networks. *arXiv preprint arXiv:1709.01507*, 7.

Huang, C., Li, Y., Change Loy, C., and Tang, X. (2016). Learning deep representation for imbalanced classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5375–5384.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *CVPR*, volume 1, page 3.

Iliff, J. J., Wang, M., Liao, Y., Plogg, B. A., Peng, W., Gundersen, G. A., Benveniste, H., Vates, G. E., Deane, R., Goldman, S. A., et al. (2012). A paravascular pathway facilitates csf flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β. *Science translational medicine*, 4(147):147ra111–147ra111.

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

Ismail, M. M., Keynton, R. S., Mostapha, M. M., ElTanboly, A. H., Casanova, M. F., Gimel'farb, G. L., and El-Baz, A. (2016). Studying autism spectrum disorder with structural and diffusion magnetic resonance imaging: a survey. *Frontiers in human neuroscience*, 10:211.

Janowczyk, A. and Madabhushi, A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of pathology informatics*, 7.

Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., and Bengio, Y. (2017). The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 11–19.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage*, 62(2):782–790.

Jenkinson, M., Pechaud, M., Smith, S., et al. (2005). Bet2: Mr-based estimation of brain, skull and scalp surfaces. In *Eleventh annual meeting of the organization for human brain mapping*, volume 17, page 167. Toronto.

Jha, S. C., Xia, K., Schmitt, J. E., Ahn, M., Girault, J. B., Murphy, V. A., Li, G., Wang, L., Shen, D., Zou, F., et al. (2018). Genetic influences on neonatal cortical thickness and surface area. *Human brain mapping*, 39(12):4998–5013.

Jog, A., Hoopes, A., Greve, D. N., Van Leemput, K., and Fischl, B. (2019). Psacnn: Pulse sequence adaptive fast whole brain segmentation. *NeuroImage*, 199:553–569.

Jones, E. J., Dawson, G., Kelly, J., Estes, A., and Webb, S. J. (2017). Parent-delivered early intervention in infants at risk for asd: Effects on electrophysiological and habituation measures of social attention. *Autism Research*, 10(5):961–972.

Joshi, A. A., Bhushan, C., Salloum, R., Wisnowski, J. L., Shattuck, D. W., and Leahy, R. M. (2018). Using the anisotropic laplace equation to compute cortical thickness. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 549–556. Springer.

Kandiah, N., Zainal, N. H., Narasimhalu, K., Chander, R. J., Ng, A., Mak, E., Au, W. L., Sitoh, Y. Y., Nadkarni, N., and Tan, L. C. (2014). Hippocampal volume and white matter disease in the prediction of dementia in parkinson's disease. *Parkinsonism & related disorders*, 20(11):1203–1208.

Karani, N., Chaitanya, K., Baumgartner, C., and Konukoglu, E. (2018). A lifelong learning approach to brain mr segmentation across scanners and protocols. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 476–484. Springer.

Karras, T., Laine, S., and Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4401–4410.

Kendall, A. and Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in neural information processing systems*, pages 5574–5584.

Kendall, A., Gal, Y., and Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7482–7491.

Kim, H., Lepage, C., Maheshwary, R., Jeon, S., Evans, A. C., Hess, C. P., Barkovich, A. J., and Xu, D. (2016a). Neocivet: Towards accurate morphometry of neonatal gyrification and clinical applications in preterm newborns. *Neuroimage*, 138:28–42.

Kim, J. S., Singh, V., Lee, J. K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., Lee, J. M., Kim, S. I., and Evans, A. C. (2005). Automated 3-d extraction and evaluation of the inner and outer cortical surfaces using a laplacian map and partial volume effect classification. *Neuroimage*, 27(1):210–221.

Kim, S. H., Fonov, V. S., Dietrich, C., Vachet, C., Hazlett, H. C., Smith, R. G., Graves, M. M., Piven, J., Gilmore, J. H., Dager, S. R., et al. (2013). Adaptive prior probability and spatial temporal intensity change estimation for segmentation of the one-year-old human brain. *Journal of neuroscience methods*, 212(1):43–55.

Kim, S. H., Lyu, I., Fonov, V. S., Vachet, C., Hazlett, H. C., Smith, R. G., Piven, J., Dager, S. R., Mckinstry, R. C., Pruett, J. R., et al. (2016b). Development of cortical shape in the human brain from 6 to 24months of age via a novel measure of shape complexity. *NeuroImage*, 135:163–176.

Kimmel, R. and Sethian, J. (1998a). Computing geodesic paths on manifolds. In *Procedings of the National Academy of Sciences*, volume 95, pages 8431–8435.

Kimmel, R. and Sethian, J. A. (1998b). Computing geodesic paths on manifolds. *Proceedings of the national academy of Sciences*, 95(15):8431–8435.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Knickmeyer, R. C., Gouttard, S., Kang, C., Evans, D., Wilber, K., Smith, J. K., Hamer, R. M., Lin, W., Gerig, G., and Gilmore, J. H. (2008). A structural mri study of human brain development from birth to 2 years. *Journal of Neuroscience*, 28(47):12176–12182.

Knickmeyer, R. C., Xia, K., Lu, Z., Ahn, M., Jha, S. C., Zou, F., Zhu, H., Styner, M., and Gilmore, J. H. (2016). Impact of demographic and obstetric factors on infant brain volumes: a population neuroscience study. *Cerebral Cortex*, 27(12):5616–5625.

Koch, G., Zemel, R., and Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2.

Koenderink, J. J. and Van Doorn, A. J. (1992). Surface shape and curvature scales. *Image and vision computing*, 10(8):557–564.

Kooi, T., Litjens, G., van Ginneken, B., Gubern-Mérida, A., Sánchez, C. I., Mann, R., den Heeten, A., and Karssemeijer, N. (2017). Large scale deep learning for computer aided detection of mammographic lesions. *Medical image analysis*, 35:303–312.

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5(4):221–232.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Kumar, D., Wong, A., and Taylor, G. W. (2017). Explaining the unexplained: A class-enhanced attentive response (clear) approach to understanding deep neural networks. In *IEEE Computer Vision and Pattern Recognition (CVPR) Workshop*.

Küstner, T., Liebgott, A., Mauch, L., Martirosian, P., Bamberg, F., Nikolaou, K., Yang, B., Schick, F., and Gatidis, S. (2018). Automated reference-free detection of motion artifacts in magnetic resonance images. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 31(2):243–256.

Landa, R. J., Gross, A. L., Stuart, E. A., and Faherty, A. (2013). Developmental trajectories in children with and without autism spectrum disorders: the first 3 years. *Child development*, 84(2):429–442.

Larsen, A. B. L., Sønderby, S. K., Larochelle, H., and Winther, O. (2015). Autoencoding beyond pixels using a learned similarity metric. *arXiv preprint arXiv:1512.09300*.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

Lee, H., Grosse, R., Ranganath, R., and Ng, A. Y. (2009). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th annual international conference on machine learning*, pages 609–616. ACM.

Lee, J., Kim, S. H., Oguz, I., and Styner, M. (2016). Enhanced cortical thickness measurements for rodent brains via lagrangian-based rk4 streamline computation. In *Proceedings of SPIE–the International Society for Optical Engineering*, volume 9784. NIH Public Access.

LERCH, J., PRUESSNER, J., ZIJDENBOS, A., and EVANS, A. C. (2003). A validation study of cortical thickness analysis. In *Proceedings of the 9th Annual Meeting of the Organization for Human Brain Mapping*.

Levin, A. R., Varcin, K. J., OLeary, H. M., Tager-Flusberg, H., and Nelson, C. A. (2017). Eeg power at 3 months in infants at high familial risk for autism. *Journal of neurodevelopmental disorders*, 9(1):34.

Lewis, J. D., Evans, A., Pruett, J., Botteron, K., Zwaigenbaum, L., Estes, A., Gerig, G., Collins, L., Kostopoulos, P., McKinstry, R., et al. (2014). Network inefficiencies in autism spectrum disorder at 24 months. *Translational psychiatry*, 4(5):e388.

Li, G., Wang, L., Shi, F., Gilmore, J. H., Lin, W., and Shen, D. (2015). Construction of 4d high-definition cortical surface atlases of infants: Methods and applications. *Medical image analysis*, 25(1):22–36.

Li, G., Wang, L., Shi, F., Lyall, A. E., Lin, W., Gilmore, J. H., and Shen, D. (2014). Mapping longitudinal development of local cortical gyrification in infants from birth to 2 years of age. *Journal of Neuroscience*, 34(12):4228–4238.

Lipton, Z. C. (2016). The mythos of model interpretability. *arXiv preprint arXiv:1606.03490*.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88.

Liu, M., Zhang, J., Adeli, E., and Shen, D. (2018). Landmark-based deep multi-instance learning for brain disease diagnosis. *Medical image analysis*, 43:157–168.

Liu, S., Johns, E., and Davison, A. J. (2019). End-to-end multi-task learning with attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1871–1880.

Lombardo, M. V., Pierce, K., Eyler, L. T., Barnes, C. C., Ahrens-Barbeau, C., Solso, S., Campbell, K., and Courchesne, E. (2015). Different functional neural substrates for good and poor language outcome in autism. *Neuron*, 86(2):567–577.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., and Rutter, M. (2000). The autism diagnostic observation schedulegeneric: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*, 30(3):205–223.

Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of autism and developmental disorders*, 24(5):659–685.

Lorensen, W. E. and Cline, H. E. (1987). Marching cubes: A high resolution 3d surface construction algorithm. In *ACM siggraph computer graphics*, volume 21, pages 163–169. ACM.

Lu, D., Popuri, K., Ding, G. W., Balachandar, R., and Beg, M. F. (2018). Multimodal and multiscale deep neural networks for the early diagnosis of alzheimers disease using structural mr and fdg-pet images. *Scientific reports*, 8(1):5697.

Lu, Y. and Xu, P. (2018). Anomaly detection for skin disease images using variational autoencoder. *arXiv preprint arXiv:1807.01349*.

Lun, M. P., Monuki, E. S., and Lehtinen, M. K. (2015). Development and functions of the choroid plexus–cerebrospinal fluid system. *Nature Reviews Neuroscience*, 16(8):445–457.

Lyttelton, O., Boucher, M., Robbins, S., and Evans, A. (2007). An unbiased iterative group registration template for cortical surface analysis. *Neuroimage*, 34(4):1535–1544.

Lyu, I., Kim, S. H., Girault, J. B., Gilmore, J. H., and Styner, M. A. (2018). A cortical shape-adaptive approach to local gyrification index. *Medical image analysis*, 48:244–258.

MacDonald, D., Avis, D., and Evans, A. C. (1994). Multiple surface identification and matching in magnetic resonance images. In *Visualization in Biomedical Computing 1994*, volume 2359, pages 160–169. International Society for Optics and Photonics.

MacDonald, D., Kabani, N., Avis, D., and Evans, A. C. (2000). Automated 3-d extraction of inner and outer surfaces of cerebral cortex from mri. *NeuroImage*, 12(3):340–356.

MacDonald, J. D. (1997). *A method for identifying geometrically simple surfaces from three dimensional images*. PhD thesis, McGill University Libraries.

Magnotta, V. A., Andreasen, N. C., Schultz, S. K., Harris, G., Cizadlo, T., Heckel, D., Nopoulos, P., and Flaum, M. (1999). Quantitative in vivo measurement of gyrification in the human brain: changes associated with aging. *Cerebral Cortex*, 9(2):151–160.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2009). Online dictionary learning for sparse coding. In *Proceedings of the 26th annual international conference on machine learning*, pages 689–696. ACM.

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., and Frey, B. (2015). Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*.

Maloof, M. A. (2003). Learning when data sets are imbalanced and when costs are unequal and unknown. In *ICML-2003 workshop on learning from imbalanced data sets II*, volume 2, pages 2–1.

Masci, J., Boscaini, D., Bronstein, M., and Vandergheynst, P. (2015). Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 37–45.

Mazurowski, M. A., Habas, P. A., Zurada, J. M., Lo, J. Y., Baker, J. A., and Tourassi, G. D. (2008). Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural networks*, 21(2-3):427–436.

McAlonan, G. M., Cheung, V., Cheung, C., Suckling, J., Lam, G. Y., Tai, K., Yip, L., Murphy, D. G., and Chua, S. E. (2004). Mapping the brain in autism. a voxel-based mri study of volumetric differences and intercorrelations in autism. *Brain*, 128(2):268–276.

McAlonan, G. M., Daly, E., Kumari, V., Critchley, H. D., Amelsvoort, T. v., Suckling, J., Simmons, A., Sigmundsson, T., Greenwood, K., Russell, A., et al. (2002). Brain anatomy and sensorimotor gating in aspergers syndrome. *Brain*, 125(7):1594–1606.

McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Mensen, V. T., Wierenga, L. M., van Dijk, S., Rijks, Y., Oranje, B., Mandl, R. C., and Durston, S. (2017). Development of cortical thickness and surface area in autism spectrum disorder. *NeuroImage: Clinical*, 13:215–222.

Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., Church, D. M., Crolla, J. A., Eichler, E. E., Epstein, C. J., et al. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *The American Journal of Human Genetics*, 86(5):749–764.

Miller, M., Iosif, A.-M., Hill, M., Young, G. S., Schwichtenberg, A., and Ozonoff, S. (2017). Response to name in infants developing autism spectrum disorder: a prospective study. *The Journal of pediatrics*, 183:141–146.

Miller, M. I., Massie, A. B., Ratnanather, J. T., Botteron, K. N., and Csernansky, J. G. (2000). Bayesian construction of geometrically based cortical thickness metrics. *NeuroImage*, 12(6):676–687.

Min, S., Chen, X., Zha, Z.-J., Wu, F., and Zhang, Y. (2019). A two-stream mutual attention network for semi-supervised biomedical segmentation with noisy labels.

Mitchell, J. S., Mount, D. M., and Papadimitriou, C. H. (1987). The discrete geodesic problem. *SIAM Journal on Computing*, 16(4):647–668.

Moeskops, P. and Pluim, J. P. (2017). Isointense infant brain mri segmentation with a dilated convolutional neural network. *arXiv preprint arXiv:1708.02757*.

Moeskops, P., Viergever, M. A., Mendrik, A. M., de Vries, L. S., Benders, M. J., and Išgum, I. (2016a). Automatic segmentation of mr brain images with a convolutional neural network. *IEEE transactions on medical imaging*, 35(5):1252–1261.

Moeskops, P., Wolterink, J. M., van der Velden, B. H., Gilhuijs, K. G., Leiner, T., Viergever, M. A., and Išgum, I. (2016b). Deep learning for multi-task medical image segmentation in multiple modalities. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 478–486. Springer.

Montavon, G., Samek, W., and Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., Initiative, A. D. N., et al. (2015). Machine learning framework for early mri-based alzheimer's conversion prediction in mci subjects. *Neuroimage*, 104:398–412.

Mostapha, M., Casanova, M. F., Gimelfarb, G., and El-Baz, A. (2015). Towards non-invasive image-based early diagnosis of autism. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 160–168. Springer.

Mostapha, M., Kim, S., Wu, G., Zsembik, L., Pizer, S., and Styner, M. (2018a). Non-euclidean, convolutional learning on cortical brain surfaces. In *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, pages 527–530. IEEE.

Mostapha, M., Prieto, J., Murphy, V., Girault, J., Foster, M., Rumple, A., Blocher, J., Lin, W., Elison, J., Gilmore, J., et al. (2019). Semi-supervised vae-gan for out-of-sample detection applied to mri quality control. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 127–136. Springer.

Mostapha, M., Shen, M. D., Kim, S., Swanson, M., Collins, D. L., Fonov, V., Gerig, G., Piven, J., and Styner, M. A. (2018b). A novel framework for the local extraction of extra-axial cerebrospinal fluid from mr brain images. In *Medical Imaging 2018: Image Processing*, volume 10574, page 105740V. International Society for Optics and Photonics.

Mostapha, M. and Styner, M. (2019). Role of deep learning in infant brain mri analysis. *Magnetic resonance imaging*.

Ng, A. Y. and Jordan, M. I. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Advances in neural information processing systems*, pages 841–848.

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., and Ng, A. Y. (2011). Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 689–696.

Nie, D., Wang, L., Adeli, E., Lao, C., Lin, W., and Shen, D. (2018). 3-d fully convolutional networks for multimodal isointense infant brain image segmentation. *IEEE Transactions on Cybernetics*.

Nie, D., Wang, L., Gao, Y., and Sken, D. (2016). Fully convolutional networks for multi-modality isointense infant brain image segmentation. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 1342–1345. IEEE.

Novotni, M. and Klein, R. (2002). Gomputing geodesic distances on triangular meshes.

Ozonoff, S., Iosif, A.-M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., Rogers, S. J., Rozga, A., Sangha, S., Sigman, M., et al. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(3):256–266.

Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., Bryson, S., Carver, L. J., Constantino, J. N., Dobkins, K., et al. (2011). Recurrence risk for autism spectrum disorders: a baby siblings research consortium study. *Pediatrics*, 128(3):e488–e495.

Ozonoff, S., Young, G. S., Steinfeld, M. B., Hill, M. M., Cook, I., Hutman, T., Macari, S., Rogers, S. J., and Sigman, M. (2009). How early do parent concerns predict later autism diagnosis? *Journal of developmental and behavioral pediatrics: JDBP*, 30(5):367.

Paniagua, B., Lyall, A., Berger, J.-B., Vachet, C., Hamer, R. M., Woolson, S., Lin, W., Gilmore, J., and Styner, M. (2013). Lateral ventricle morphology analysis via mean latitude axis. In *Medical Imaging 2013: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 8672, page 86720M. International Society for Optics and Photonics.

Park, J. J., Florence, P., Straub, J., Newcombe, R., and Lovegrove, S. (2019). Deepsdf: Learning continuous signed distance functions for shape representation. *arXiv preprint arXiv:1901.05103*.

Paulsen, J. S., Nopoulos, P. C., Aylward, E., Ross, C. A., Johnson, H., Magnotta, V. A., Juhl, A., Pierson, R. K., Mills, J., Langbehn, D., et al. (2010). Striatal and white matter predictors of estimated diagnosis for huntington disease. *Brain research bulletin*, 82(3-4):201–207.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., and Nichols, T. E. (2011). *Statistical parametric mapping: the analysis of functional brain images*. Elsevier.

Pereira, S., Pinto, A., Alves, V., and Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in mri images. *IEEE transactions on medical imaging*, 35(5):1240–1251.

Perez, L. and Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.

Peyre, G. and Cohen, L. D. (2006). Geodesic remeshing using front propagation. *International Journal of Computer Vision*, pages 145–156.

Pizarro, R., Assemlal, H.-E., De Nigris, D., Elliott, C., Antel, S., Arnold, D., and Shmuel, A. (2019). Using deep learning algorithms to automatically identify the brain mri contrast: implications for managing large databases. *Neuroinformatics*, 17(1):115–130.

Plitt, M., Barnes, K. A., Wallace, G. L., Kenworthy, L., and Martin, A. (2015). Resting-state functional connectivity predicts longitudinal change in autistic traits and adaptive functioning in autism. *Proceedings of the National Academy of Sciences*, 112(48):E6699–E6706.

Qayyum, A., Anwar, S. M., Awais, M., and Majid, M. (2017). Medical image retrieval using deep convolutional neural network. *Neurocomputing*, 266:8–20.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 652–660.

Racine, A. M., Brickhouse, M., Wolk, D. A., Dickerson, B. C., Initiative, A. D. N., et al. (2018). The personalized alzheimer's disease cortical thickness index predicts likely pathology and clinical progression in mild cognitive impairment. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10:301–310.

Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Rajchl, M., Lee, M. C., Oktay, O., Kamnitsas, K., Passerat-Palmbach, J., Bai, W., Damodaram, M., Rutherford, M. A., Hajnal, J. V., Kainz, B., et al. (2017). Deepcut: Object segmentation from bounding box annotations using convolutional neural networks. *IEEE transactions on medical imaging*, 36(2):674–683.

Rasmussen, K. G. B., Kristensen, M. J., Blendal, R. G., Ostergaard, L. R., Plocharski, M., O'Brien, K., Langkammer, C., Janke, A., Barth, M., and Bollmann, S. (2018). Deepqsm-using deep learning to solve the dipole inversion for mri susceptibility mapping. *Biorxiv*, page 278036.

Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., and Lee, H. (2016). Generative adversarial text to image synthesis. *arXiv preprint arXiv:1605.05396*.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144. ACM.

Rice, C. (2009). Prevalence of autism spectrum disorders–autism and developmental disabilities monitoring network, united states, 2006.

Robbins, S., Evans, A. C., Collins, D. L., and Whitesides, S. (2003). Tuning and comparing spatial normalization methods. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 910–917. Springer.

Rodriguez-Carranza, C. E., Mukherjee, P., Vigneron, D., Barkovich, J., and Studholme, C. (2008). A framework for in vivo quantification of regional brain folding in premature neonates. *Neuroimage*, 41(2):462–478.

Rogers, S. J. (2009). What are infant siblings teaching us about autism in infancy? *Autism Research*, 2(3):125–137.

Rogers, S. J., Vismara, L., Wagner, A., McCormick, C., Young, G., and Ozonoff, S. (2014). Autism treatment in the first year of life: a pilot study of infant start, a parent-implemented intervention for symptomatic infants. *Journal of autism and developmental disorders*, 44(12):2981–2995.

Rogers, S. J. and Vismara, L. A. (2008). Evidence-based comprehensive treatments for early autism. *Journal of Clinical Child & Adolescent Psychology*, 37(1):8–38.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

Ruggeri, B., Sarkans, U., Schumann, G., and Persico, A. M. (2014). Biomarkers in autism spectrum disorder: the old and the new. *Psychopharmacology*, 231(6):1201–1216.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252.

Salimans, T. and Kingma, D. P. (2016). Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *Advances in Neural Information Processing Systems*, pages 901–909.

Sarraf, S., Tofighi, G., et al. (2016). Deepad: Alzheimer s disease classification via deep convolutional neural networks using mri and fmri. *BioRxiv*, page 070441.

Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., and Langs, G. (2017). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International Conference on Information Processing in Medical Imaging*, pages 146–157. Springer.

Schlemper, J., Caballero, J., Hajnal, J. V., Price, A. N., and Rueckert, D. (2017). A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE transactions on Medical Imaging*, 37(2):491–503.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61:85–117.

Scholköpf, B. and Smola, A. (2002). Support vector machines, regularization, optimization, and beyond. *Learning with Kernels*.

Schwarz, C. G., Gunter, J. L., Wiste, H. J., Przybelski, S. A., Weigand, S. D., Ward, C. P., Senjem, M. L., Vemuri, P., Murray, M. E., Dickson, D. W., et al. (2016). A large-scale comparison of cortical thickness and volume methods for measuring alzheimer's disease severity. *NeuroImage: Clinical*, 11:802–812.

Sener, O. and Koltun, V. (2018). Multi-task learning as multi-objective optimization. In *Advances in Neural Information Processing Systems*, pages 527–538.

Sethian, J. A. (1998). Fast marching methods. *SIAM Review*, 41:199–235.

Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., Greenstein, D., Clasen, L., Evans, A., Rapoport, J. L., et al. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *Journal of Neuroscience*, 28(14):3586–3594.

Shaw, P., Malek, M., Watson, B., Sharp, W., Evans, A., and Greenstein, D. (2012). Development of cortical surface area and gyrification in attention-deficit/hyperactivity disorder. *Biological psychiatry*, 72(3):191–197.

Shen, M. D. (2018). Cerebrospinal fluid and the early brain development of autism. *Journal of neurodevelopmental disorders*, 10(1):39.

Shen, M. D., Kim, S. H., McKinstry, R. C., Gu, H., Hazlett, H. C., Nordahl, C. W., Emerson, R. W., Shaw, D., Elison, J. T., Swanson, M. R., et al. (2017). Increased extra-axial cerebrospinal fluid in high-risk infants who later develop autism. *Biological psychiatry*, 82(3):186–193.

Shen, M. D., Nordahl, C. W., Li, D. D., Lee, A., Angkustsiri, K., Emerson, R. W., Rogers, S. J., Ozonoff, S., and Amaral, D. G. (2018). Extra-axial cerebrospinal fluid in high-risk and normal-risk children with autism aged 2–4 years: a case-control study. *The Lancet Psychiatry*, 5(11):895–904.

Shen, M. D., Nordahl, C. W., Young, G. S., Wootton-Gorges, S. L., Lee, A., Liston, S. E., Harrington, K. R., Ozonoff, S., and Amaral, D. G. (2013). Early brain enlargement and elevated extra-axial fluid in infants who develop autism spectrum disorder. *Brain*, 136(9):2825–2835.

Shimony, J. S., Smyser, C. D., Wideman, G., Alexopoulos, D., Hill, J., Harwell, J., Dierker, D., Van Essen, D. C., Inder, T. E., and Neil, J. J. (2016). Comparison of cortical folding measures for evaluation of developing human brain. *Neuroimage*, 125:780–790.

Shin, H.-C., Orton, M. R., Collins, D. J., Doran, S. J., and Leach, M. O. (2013). Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1930–1943.

Shrikumar, A., Greenside, P., and Kundaje, A. (2017). Learning important features through propagating activation differences. *arXiv preprint arXiv:1704.02685*.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al. (2017). Mastering the game of go without human knowledge. *Nature*, 550(7676):354.

Simon, M. J. and Iliff, J. J. (2016). Regulation of cerebrospinal fluid (csf) flow in neurodegenerative, neurovascular and neuroinflammatory disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1862(3):442–451.

Simonyan, K., Vedaldi, A., and Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*.

Sled, J. G., Zijdenbos, A. P., and Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE transactions on medical imaging*, 17(1):87–97.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping*, 17(3):143–155.

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack Jr, C. R., Kaye, J., Montine, T. J., et al. (2011). Toward defining the preclinical stages of alzheimers disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):280–292.

Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. (2014). Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.

Styner, M., Oguz, I., Xu, S., Brechbühler, C., Pantazis, D., Levitt, J. J., Shenton, M. E., and Gerig, G. (2006). Framework for the statistical shape analysis of brain structures using spharm-pdm. *The insight journal*, (1071):242.

Su, S., White, T., Schmidt, M., Kao, C.-Y., and Sapiro, G. (2013). Geometric computation of human gyrification indexes from magnetic resonance images. *Human brain mapping*, 34(5):1230–1244.

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

Sun, J., Li, H., Xu, Z., et al. (2016). Deep admm-net for compressive sensing mri. In *Advances in neural information processing systems*, pages 10–18.

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., and Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). A survey on deep transfer learning. In *International conference on artificial neural networks*, pages 270–279. Springer.

Tarasoff-Conway, J. M., Carare, R. O., Osorio, R. S., Glodzik, L., Butler, T., Fieremans, E., Axel, L., Rusinek, H., Nicholson, C., Zlokovic, B. V., et al. (2015). Clearance systems in the brainimplications for alzheimer disease. *Nature reviews neurology*, 11(8):457.

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., et al. (2014). The enigma consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain imaging and behavior*, 8(2):153–182.

Thung, K.-H., Yap, P.-T., and Shen, D. (2017). Multi-stage diagnosis of alzheimers disease with incomplete multimodal data via multi-task deep learning. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 160–168. Springer.

Tohka, J., Zijdenbos, A., and Evans, A. (2004). Fast and robust parameter estimation for statistical partial volume models in brain mri. *Neuroimage*, 23(1):84–97.

Toro, R. and Burnod, Y. (2005). A morphogenetic model for the development of cortical convolutions. *Cerebral cortex*, 15(12):1900–1913.

Tsagkrasoulis, D., Hysi, P., Spector, T., and Montana, G. (2017). Heritability maps of human face morphology through large-scale automated three-dimensional phenotyping. *Scientific reports*, 7:45885.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. (2010). N4itk: improved n3 bias correction. *IEEE transactions on medical imaging*, 29(6):1310–1320.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289.

Ulyanov, D., Vedaldi, A., and Lempitsky, V. (2016). Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*.

Van den Oord, A., Kalchbrenner, N., Espeholt, L., Vinyals, O., Graves, A., et al. (2016). Conditional image generation with pixelcnn decoders. In *Advances in neural information processing systems*, pages 4790–4798.

Van den Veyver, I. B. (2019). Prenatally diagnosed developmental abnormalities of the central nervous system and genetic syndromes: A practical review. *Prenatal diagnosis*.

Van Der Maaten, L., Postma, E., and Van den Herik, J. (2009). Dimensionality reduction: a comparative. *J Mach Learn Res*, 10:66–71.

Van Essen, D. and Drury, H. (1997). Structural and functional analyses of human cerebral cortex using a surface-based atlas. *Journal of Neuroscience*, 17(18):7079–7102.

Van Essen, D. C. (2005). A population-average, landmark-and surface-based (pals) atlas of human cerebral cortex. *Neuroimage*, 28(3):635–662.

Van Essen, D. C., Dierker, D., Snyder, A., Raichle, M. E., Reiss, A. L., and Korenberg, J. (2006). Symmetry of cortical folding abnormalities in williams syndrome revealed by surface-based analyses. *Journal of Neuroscience*, 26(20):5470–5483.

Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. (1999). Automated model-based tissue classification of mr images of the brain. *IEEE transactions on medical imaging*, 18(10):897–908.

Vasilev, A., Golkov, V., Lipp, I., Sgarlata, E., Tomassini, V., Jones, D. K., and Cremers, D. (2018). q-space novelty detection with variational autoencoders. *arXiv preprint arXiv:1806.02997*.

Vidal, C. N., Nicolson, R., Boire, J.-Y., Barra, V., DeVito, T. J., Hayashi, K. M., Geaga, J. A., Drost, D. J., Williamson, P. C., Rajakumar, N., et al. (2008). Three-dimensional mapping of the lateral ventricles in autism. *Psychiatry Research: Neuroimaging*, 163(2):106–115.

Vivanti, G., Dissanayake, C., Team, V. A., et al. (2016). Outcome for children receiving the early start denver model before and after 48 months. *Journal of autism and developmental disorders*, 46(7):2441–2449.

Wagstyl, K. and Lerch, J. P. (2018). Cortical thickness. In *Brain Morphometry*, pages 35–49. Springer.

Wang, J., Vachet, C., Rumple, A., Gouttard, S., Ouziel, C., Perrot, E., Du, G., Huang, X., Gerig, G., and Styner, M. (2014). Multi-atlas segmentation of subcortical brain structures via the autoseg software pipeline. *Frontiers in neuroinformatics*, 8.

Wang, S., Minku, L. L., and Yao, X. (2015). Resampling-based ensemble methods for online class imbalance learning. *IEEE Transactions on Knowledge and Data Engineering*, 27(5):1356–1368.

Wang, S. and Summers, R. M. (2012). Machine learning and radiology. *Medical image analysis*, 16(5):933–951.

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., and Solomon, J. M. (2019). Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (TOG)*, 38(5):146.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack Jr, C. R., Jagust, W., Morris, J. C., et al. (2017). The alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's & Dementia*, 13(5):561–571.

Wernick, M. N., Yang, Y., Brankov, J. G., Yourganov, G., and Strother, S. C. (2010). Machine learning in medical imaging. *IEEE signal processing magazine*, 27(4):25–38.

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., Duggirala, R., and Glahn, D. C. (2010). Cortical thickness or grey matter volume? the importance of selecting the phenotype for imaging genetics studies. *Neuroimage*, 53(3):1135–1146.

Winkler, A. M., Sabuncu, M. R., Yeo, B. T., Fischl, B., Greve, D. N., Kochunov, P., Nichols, T. E., Blangero, J., and Glahn, D. C. (2012). Measuring and comparing brain cortical surface area and other areal quantities. *Neuroimage*, 61(4):1428–1443.

Wolff, J. J., Gerig, G., Lewis, J. D., Soda, T., Styner, M. A., Vachet, C., Botteron, K. N., Elison, J. T., Dager, S. R., Estes, A. M., et al. (2015). Altered corpus callosum morphology associated with autism over the first 2 years of life. *Brain*, 138(7):2046–2058.

Wolff, J. J., Gu, H., Gerig, G., Elison, J. T., Styner, M., Gouttard, S., Botteron, K. N., Dager, S. R., Dawson, G., Estes, A. M., et al. (2012). Differences in white matter fiber tract development present from 6 to 24 months in infants with autism. *American Journal of Psychiatry*, 169(6):589–600.

Wolterink, J. M., Dinkla, A. M., Savenije, M. H., Seevinck, P. R., van den Berg, C. A., and Išgum, I. (2017). Deep mr to ct synthesis using unpaired data. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 14–23. Springer.

Wu, Y. and He, K. (2018). Group normalization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–19.

Xu, Y., Géraud, T., and Bloch, I. (2017). From neonatal to adult brain mr image segmentation in a few seconds using 3d-like fully convolutional network and transfer learning. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 4417–4421. IEEE.

Yaakub, A. and Evans, D. J. (1999). A fourth order runge–kutta rk (4, 4) method with error control. *International journal of computer mathematics*, 71(3):383–411.

Yahata, N., Morimoto, J., Hashimoto, R., Lisi, G., Shibata, K., Kawakubo, Y., Kuwabara, H., Kuroda, M., Yamada, T., Megumi, F., et al. (2016). A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nature communications*, 7:11254.

Yang, X., Kwitt, R., Styner, M., and Niethammer, M. (2017). Quicksilver: Fast predictive image registration–a deep learning approach. *NeuroImage*, 158:378–396.

Yang, Y. and Hospedales, T. M. (2016). Trace norm regularised deep multi-task learning. *arXiv preprint arXiv:1606.04038*.

Yezzi, A. J. and Prince, J. L. (2003). An eulerian pde approach for computing tissue thickness. *IEEE transactions on medical imaging*, 22(10):1332–1339.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., and Lipson, H. (2015). Understanding neural networks through deep visualization. *arXiv preprint arXiv:1506.06579*.

Zeiler, M. D., Taylor, G. W., and Fergus, R. (2011). Adaptive deconvolutional networks for mid and high level feature learning. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2018–2025. IEEE.

Zeng, G. and Zheng, G. (2018). Multi-stream 3d fcn with multi-scale deep supervision for multi-modality isointense infant brain mr image segmentation. In *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, pages 136–140. IEEE.

Zeng, K., Zheng, H., Cai, C., Yang, Y., Zhang, K., and Chen, Z. (2018). Simultaneous single-and multi-contrast super-resolution for brain mri images based on a convolutional neural network. *Computers in biology and medicine*, 99:133–141.

Zhang, C. and Ma, Y. (2012). *Ensemble machine learning: methods and applications*. Springer.

Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., and Shen, D. (2015). Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage*, 108:214–224.

Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, 20(1):45–57.

Zintgraf, L. M., Cohen, T. S., Adel, T., and Welling, M. (2017). Visualizing deep neural network decisions: Prediction difference analysis. *arXiv preprint arXiv:1702.04595*.

Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M., Kaus, M. R., Haker, S. J., Wells III, W. M., Jolesz, F. A., and Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. *Academic radiology*, 11(2):178–189.

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., and Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International journal of developmental neuroscience*, 23(2-3):143–152.