Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

NEUROBIOLOGICAL MARKERS FOR REMISSION AND PERSISTENCE OF CHILDHOOD ATTENTION-DEFICIT/HYPERACTIVITY DISORDER

by
Yuyang Luo

Attention-deficit/hyperactivity disorder (ADHD) is one of the most prevalent neurodevelopmental disorders in children. Symptoms of childhood ADHD persist into adulthood in around 65% of patients, which elevates the risk for a number of adverse outcomes, resulting in substantial individual and societal burden. A neurodevelopmental double dissociation model is proposed based on existing studies in which the early onset of childhood ADHD is suggested to associate with dysfunctional subcortical structures that remain static throughout the lifetime; while diminution of symptoms over development could link to optimal development of prefrontal cortex. Current existing studies only assess basic measures including regional brain activation and connectivity, which have limited capacity to characterize the functional brain as a high performance parallel information processing system, the field lacks systems-level investigations of the structural and functional patterns that significantly contribute to the symptom remission and persistence in adults with childhood ADHD. Furthermore, traditional statistical methods estimate group differences only within a voxel or region of interest (ROI) at a time without having the capacity to explore how ROIs interact in linear and/or non-linear ways, as they quickly become overburdened when attempting to combine predictors and their interactions from high-dimensional imaging data set.

This dissertation is the first study to apply ensemble learning techniques (ELT) in multimodal neuroimaging features from a sample of adults with childhood ADHD and

controls, who have been clinically followed up since childhood. A total of 36 adult probands who were diagnosed with ADHD combined-type during childhood and 36 matched normal controls (NCs) are involved in this dissertation research. Thirty-six adult probands are further split into 18 remitters (ADHD-R) and 18 persisters (ADHD-P) based on the symptoms in their adulthood from DSM-IV ADHD criteria. Cued attention task-based fMRI, structural MRI, and diffusion tensor imaging data from each individual are analyzed. The high-dimensional neuroimaging features, including pair-wise regional connectivity and global/nodal topological properties of the functional brain network for cue-evoked attention process, regional cortical thickness and surface area, subcortical volume, volume and fractional anisotropy of major white matter fiber tract for each subject are calculated. In addition, all the currently available optimization strategies for ensemble learning techniques (i.e., voting, bagging, boosting and stacking techniques) are tested in a pool of semi-final classification results generated by seven basic classifiers, including K-Nearest Neighbors, support vector machine (SVM), logistic regression, Naïve Bayes, linear discriminant analysis, random forest, and multilayer perceptron.

As hypothesized, results indicate that the features of nodal efficiency in right inferior frontal gyrus, right middle frontal (MFG)-inferior parietal (IPL) functional connectivity, and right amygdala volume significantly contributed to accurate discrimination between ADHD probands and controls; higher nodal efficiency of right MFG greatly contributed to inattentive and hyperactive/impulsive symptom remission, while higher right MFG-IPL functional connectivity strongly linked to symptom persistence in adults with childhood ADHD. The utilization of ELTs indicates that the bagging-based ELT with the base model of SVM achieves the best results, with the most

significant improvement of the area under the receiver of operating characteristic curve (0.89 for ADHD probands vs. NCs, and 0.9 for ADHD-P vs. ADHD-R). The outcomes of this dissertation research have considerable value for the development of novel interventions that target mechanisms associated with recovery.

NEUROBIOLOGICAL MARKERS FOR REMISSION AND PERSISTENCE OF CHILDHOOD ATTENTION-DEFICIT/HYPERACTIVITY DISORDER

by
Yuyang Luo

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
and Rutgers University Biomedical and Health Sciences – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Biomedical Engineering

Department of Biomedical Engineering

May 2020

APPROVAL PAGE

NEUROBIOLOGICAL MARKERS FOR REMISSION AND PERSISTENCE OF CHILDHOOD ATTENTION-DEFICIT/HYPERACTIVITY DISORDER

Yuyang Luo

Dr. Xiaobo Li, Dissertation Primary Advisor — Date
Associate Professor of Biomedical Engineering, NJIT

Dr. Tara L. Alvarez, Dissertation Co-Advisor — Date
Professor of Biomedical Engineering, NJIT

Dr. Bharat B. Biswal, Committee Member — Date
Distinguished Professor of Biomedical Engineering, NJIT

Dr. Yiyan Liu, Committee Member — Date
Professor of Radiology, Rutgers New Jersey Medical School

Dr. Kevin Pang, Committee Member — Date
Professor of Pharmacology, Physiology & Neuroscience, Rutgers New Jersey Medical School

BIOGRAPHICAL SKETCH

Author: Yuyang Luo

Degree: Doctor of Philosophy

Date: May 2020

Undergraduate and Graduate Education:

- Doctor of Philosophy in Biomedical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2020

- Master of Engineering in Electrical Engineering,
  Stevens Institute of Technology, Hoboken, NJ, 2014

- Bachelor of Science in Electronics and Information Engineering,
  Wuhan University of Science and Technology, Wuhan, China, 2012

Major: Biomedical Engineering

Presentations and Publications:

Luo, Y., Alvarez, T.L., Halperin, J.M., & Li, X. (2020). Multimodal neuroimaging-based prediction of adult outcomes in childhood-onset ADHD using ensemble learning techniques. *Neuroimage Clin* 26**,** 102238.

Luo, Y., Halperin, J.M., & Li, X. (2020). Anatomical substrates of symptom remission and persistence in young adults with childhood attention deficit/hyperactivity disorder. *Eur Neuropsychopharmacol* 33, 117-125.

Wu, Z., Luo, Y., Gao, Y., Han, Y., Wu, K., & Li, X. (2020). The role of prefrontal and occipital cortices in visual sustained attention processing in young adults with attention-deficit/hyperactivity disorder: a functional near-infrared spectroscopy study. *Neurosci Bull*.

Luo, Y., Weibman, D., Halperin, J.M., & Li, X. (2019). A review of heterogeneity in attention deficit/hyperactivity disorder (ADHD). *Front Hum Neurosci* 13**,** 42.

Huang, Y., Wu, T., Gao, Y., Luo, Y., Wu, Z., Fagan, S., Leung, S., & Li, X. (2019). The impact of callous-unemotional traits and externalizing tendencies on neural responsivity to reward and punishment in healthy adolescents. *Front Neurosci* 13**,** 1319.

Luo, Y., Schulz, K.P., Alvarez, T.L., Halperin, J.M., & Li, X. (2018). Distinct topological properties of cue-evoked attention processing network in persisters and remitters of childhood ADHD. *Cortex* 109**,** 234-244.

Abstracts:

Luo, Y. & Li, X. (2019). Disrupted brain network topology in motoric cognitive risk syndrome: a resting-state functional magnetic resonance imaging study. Poster presentation at the *Biomedical Engineering Society (BMES) Annual Meeting*, Philadelphia, Pennsylvania.

Luo, Y. & Li, X. (2019). Disrupted brain network topology in motoric cognitive risk syndrome: a resting-state functional magnetic resonance imaging study. Oral presentation at the *45th Northeast Bioengineering Conference* (NEBEC), New Brunswick, New Jersey.

Luo, Y. & Li, X. (2018). Structural brain abnormalities in young adults remitted and persistent with childhood-onset attention-deficit/hyperactivity disorder. Poster presentation at the *44th Northeast Bioengineering Conference* (NEBEC), Philadelphia, Pennsylvania.

Luo, Y., & Li, X. (2017). Functional brain connectivity patterns for cued alerting processing in young adults remitted and persistent with childhood-onset ADHD. Oral presentation at the *43rd Northeast Bioengineering Conference* (NEBEC), Newark, New Jersey.

Wu, Z., Luo, Y., Baskar, A.K. & Li, X. (2017). Testing cortical activation responding to visual attention in young adults with traumatic brain injury – a functional near-infrared spectroscopy pilot study. Oral presentation at the *43rd Northeast Bioengineering Conference* (NEBEC), Newark, New Jersey.

Give us grace. Give us space. Understand why we do the things we do. And get to know us. You'll find most of us make loyal, dedicated friends. Give us a chance. We will cherish you forever for it.

-Elizabeth Broadbent

To my dear father and mother, Zongming Luo and Rui Liu for being my pillars of Strength during this rollercoaster of a journey with their unconditional love.

To my parents-in-law, Gang Zhao and Yi Yang, for their continuous support and for the encouragement and motivation.

To my wife, Jiangyang Zhao for being always with me whatever it is ups and downs.

To my friends, old and new, for their continuous support and confidence.

To myself, never forget appreciation and cherishing.

ACKNOWLEDGMENT

I would like to express my warmest gratitude to my dissertation advisor, Dr. Xiaobo Li for her academic supervision and remarkable mentoring through all my years in graduate school. Her immense knowledge, motivation and experience have guided and shaped my ability as a researcher and will influence me for the rest of my life.

I am thankful for my co-advisor, Dr. Tara L. Alvarez, for her encouragement and insightful suggestions from the start to the end of this Ph.D. I also extend my sincere thanks to Dr. Bharat B. Biswal from NJIT, Dr. Kevin Pang and Dr. Yiyan Liu from Rutgers NJMS for their time and invaluable feedback to improve the quality of this dissertation. They all have played a major role in polishing my research.

I would like to acknowledge the support from the National Institute of Mental Health (R03MH109791, R15MH117368, and R01MH060698) and the New Jersey Commission on Brain Injury Research (CBIR17PIL012).

I would also like to thank current and former members of the Computational Neuroanatomy and Neuroinformatics Laboratory: Ziyan Wu, Jincheng Li, Cristian Morales, Meng Cao and Dana Weibman for giving me the motivation and emotional support I needed to finish my work.

Last, but not least, I am genuinely grateful to my parents, Zongming Luo and Rui Liu, my parents-in-law, Gang Zhao and Yi Yang, and my wife, Jingyang Zhao, whose love and good wishes are with me wherever I go.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF TABLES
(Continued)

LIST OF FIGURES

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

CHAPTER 1

INTRODUCTION

1.1 Background and Significance

1.1.1 General Introduction to Attention-Deficit/Hyperactivity Disorder

Attention-deficit/hyperactivity disorder (ADHD) is one of the most commonly diagnosed neurodevelopmental disorders in children which has a prevalence of 5.29% worldwide (Polanczyk et al., 2007) and affect approximately 9.5% in school-age children in the United States (Visser et al., 2014; Pastor et al., 2015; Danielson et al., 2018; Xu et al., 2018). Throughout an individual's lifetime, ADHD can significantly increase risk for other psychiatric disorders, educational and occupational failure, accidents, criminality, social disability and addictions. ADHD is associated with widespread functional brain impairments that may result in substantial cognitive deficits, including inattention, hyperactivity and impulsivity, and thus lead to behavioral anomalies in ADHD patients (Sonuga-Barke, 2002). Symptoms of ADHD persist into adulthood in around 65% of patients, and impose enormous impairments and suffering in patients, their families, and society (Faraone et al., 2006). The behavioral abnormalities of ADHD and the characteristic of persistence elevate the risk for a number of adverse outcomes that result in substantial individual, familial and societal burden. This disparity among the different adult outcomes may be explained by the heterogeneity of ADHD.

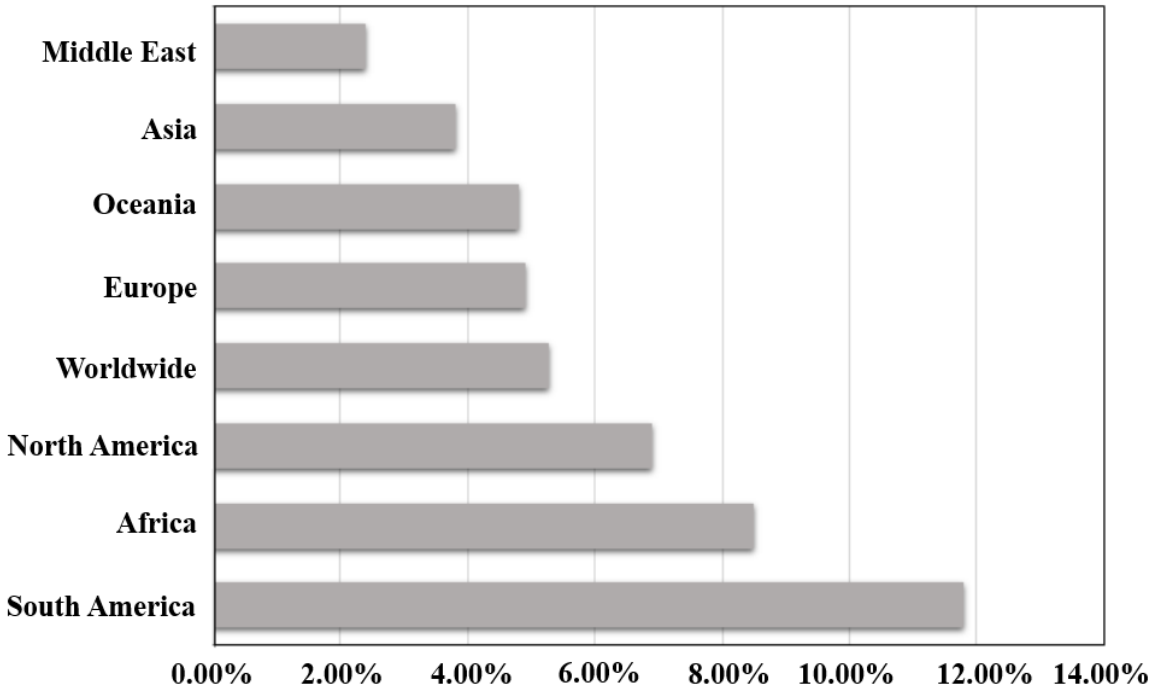

Figure 1.1 Worldwide prevalence of ADHD.

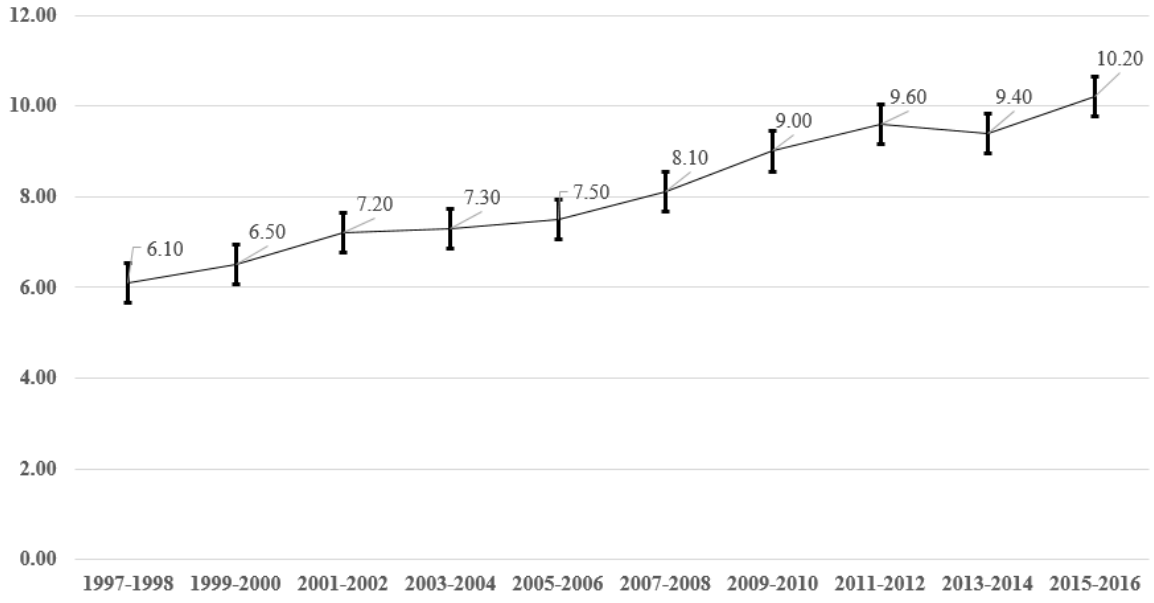

Figure 1.2 Prevalence of diagnosed attention-deficit/hyperactivity disorder in US children and adolescents from 1997 to 2016.

1.1.2 Characteristics of Attention-Deficit/Hyperactivity Disorder

Previous studies have repeatedly emphasized that ADHD is a heterogeneous disorder, in terms of the multifactorial etiological risk factors, diverse expressions of the symptom domains, comorbid disorders, neuropsychological impairments, and long-term trajectories (Luo et al., 2019).

1.1.2.1 Etiological Risk Factors in Attention-Deficit/Hyperactivity Disorder. The etiological heterogeneity in terms of the biological and environmental factors is likely reflected in variation in neural correlates, and results in the diverse cognitive and behavioral profiles and developmental trajectories of the disorder. Existing research suggests that genetic variants, and pre- and peri-natal risk factors relate to the manifestation of ADHD symptoms, and appear to be associated with various neurodevelopmental and psychiatric outcomes (Bonvicini et al., 2018; Uchida et al., 2018). Family-based studies have consistently found higher rates of ADHD in parents and siblings of affected probands, compared to the biological relatives of unaffected controls (Chen et al., 2008). Twins studies showed that monozygotic twin pairs have much higher concordance rates for ADHD than dizygotic twin pairs (Faraone et al., 2005). Adoption studies reported increased rates of ADHD in the biological parents of ADHD adoptees, compared to both the adoptive parents of the probands and parents of controls without ADHD (Sprich et al., 2000). All these studies suggest a strong genetic component of the disorder, with heritability estimates of 60%–80% (Faraone et al., 2005).

In the past three decades, molecular genetic association studies (Faraone et al., 2005; Gizer et al., 2009), linkage studies (Ogdie et al., 2004), meta-analyses (Neale et al., 2010), and recent reviews (Thapar et al., 2013; Klein et al., 2017) have identified a number

of genes that might contribute to the onset of childhood ADHD, including, but not limited to, dopamine receptor genes such as DRD4, DRD5, DRD2, DRD3, dopamine-beta-hydroxylase (DBH), dopamine transporter gene (DAT, SLC6A3), norepinephrine transporter gene (SLC6A2), noradrenergic receptor genes such as ADRA2A, 2C, and 1C, monoamine oxidase-A (MAO-A), catechol-O-methyltransferase (COMT) serotonin receptor and transporter genes including HTR2A, HTR1B, 5-HTT, SLC6A4, and at least one GABA gene, GABRB3. However, results have been inconsistent and many findings have not been consistently replicated. For example, multiple candidate-gene association studies have indicated that polymorphisms in DRD4 and DAT1 are associated with childhood ADHD (Brookes et al., 2006a; Gornick et al., 2007), while results from a case-control study did not find any association between these two genes and ADHD (Johansson et al., 2008). Moreover, studies attempting to replicate the genetic associations have yielded mixed results. Several studies have suggested that DRD4 is more strongly associated with inattentive symptoms than hyperactive-impulsive symptoms of ADHD (McCracken et al., 2000; Gizer and Waldman, 2012), while other studies showed that DRD4 was implicated in both inattentive and hyperactive symptoms (Lasky-Su et al., 2007; Bidwell et al., 2011). Despite evidence of a strong genetic contribution to ADHD, the inconsistent findings from genetic association studies may result from the relatively small effect sizes, with each gene only accounting for a small proportion of the overall ADHD risk (Gizer et al., 2009).

Environmental risk factors, including maternally related prenatal risks, pregnancy and birth complications, traumatic brain injuries and other external factors, have also been linked to ADHD (Lahey et al., 2009; Thapar and Rutter, 2009; Froehlich et al., 2011a;

Thapar et al., 2012; Van Batenburg-Eddes et al., 2013; Adeyemo et al., 2014; Glover, 2014; Silva et al., 2014; Chang et al., 2018; Saez et al., 2018; Schwenke et al., 2018). Prenatal and perinatal factors, including maternal alcohol consumption and smoking (Yoshimasu et al., 2009; Silva et al., 2014; Schwenke et al., 2018), maternal stress (Grizenko et al., 2008; Van Batenburg-Eddes et al., 2013; Glover, 2014), and low birth weight and prematurity (Mick et al., 2002; Thapar et al., 2013) are frequently associated with ADHD. Exposures to other toxins in prenatal and postnatal life have also been considered as increasing the risk of ADHD (Thapar et al., 2013). In particular, organophosphate pesticides, polychlorinated biphenyls, and lead may damage the neural systems implicated in ADHD (Nigg, 2008; Chang et al., 2018). Damage to the brain after birth due to traumatic brain injury has also been considered as a risk factor for ADHD (Pineda et al., 2007; Adeyemo et al., 2014). Multiple indicators of psychosocial adversity, including conflict/parent-child hostility, family adversity and low income, severe early deprivation, have also been found to be associated with ADHD (Pheula et al., 2011). Although there are biologically plausible mechanisms through which these risks could contribute to ADHD, it remains controversial about whether the associations of these environmental risks are directly causal. For example, studies found that children's ADHD symptoms impact mother-son hostility, rather than the hostility having a causal effect on ADHD (Lifford et al., 2009). However, such statement could not be applied on other environmental risk factors, such as perinatal complications or lead exposure.

Genes and environment do not work independently of each other (Nigg et al., 2010). Studies have explored ways in which the inherited genetic factors might interact with environmental risk factors to influence ADHD development and outcomes. The

DAT1 gene has been found to interact with maternal smoking and alcohol use during pregnancy (Brookes et al., 2006b), while other studies have failed to replicate these results (Becker et al., 2008). Interaction between DAT1 and maternal alcohol use during pregnancy has also been reported to decrease risk for ADHD (Brookes et al., 2006b). Using a sample of twin pairs, one study found that the interaction between DAT1 9R-allele or DRD4 7R-allele and prenatal smoke exposure increased risk for combined-type ADHD by nine-fold (Neuman et al., 2007). In addition to increasing susceptibility to prenatal adversities, the DAT1 gene has also been found to increase risk for ADHD in the presence of psychosocial adversity. Specifically, the DAT1 10R-6R allele has been found to moderate the effects of early institutional deprivation (Stevens et al., 2009) and psychosocial adversity (Laucht et al., 2007) on ADHD risk. DRD4 has been reported to significantly interact with high stress level during pregnancy and some chemical toxins, e.g., dimethyl phosphate, in children with ADHD (Grizenko et al., 2012; Chang et al., 2018). Nevertheless, these findings have either not been replicated or inconsistent.

Results from existing studies reviewed above suggest that multiple genetic and environmental risk factors with small individual effect sizes contribute to the heterogeneity of ADHD. These etiological risk factors may interact with each other and the complex developmental neural mechanisms, together result in the diverse clinical profiles and outcomes of ADHD.

1.1.2.2 Clinical Information of Attention-Deficit/Hyperactivity Disorder. Currently, the diagnosis of ADHD is characterized by age-inappropriate symptoms of inattention and/or hyperactivity-impulsivity, categorized into three presentations including predominantly inattentive, predominantly hyperactive/impulsive and combined

presentation, based on the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5; American Psychiatric Association). A diagnosis of ADHD is typically determined by a clinician based on the number, severity, and duration of behavioral symptoms observed by parents/caregivers and teachers. They are not defined based on etiological sources, or any biologically identified markers. It still remains an open question about the relations between the clinical definitions, etiological sources and neurobiological substrates of ADHD. A wide range of comorbid behavioral and psychiatric conditions are associated with ADHD, including learning disabilities, language disorders, mood disorders, anxiety, and conduct/oppositional disorder. These comorbid problems can complicate both diagnosis and treatment of ADHD.

Most diagnoses of ADHD are made in school-age children. ADHD, according to DSM-5, requires symptoms to present in multiple settings before the age of 12 years. However, the course and outcome of childhood ADHD are highly heterogeneous. Cross-sectional studies have found that ADHD-related symptoms have development-specific features, with motor restlessness, aggressive and disruptive behaviors commonly observed in preschoolers, disorganized, impulsive, and inattentive symptoms more typically presented in adolescents and adults (Wilens et al., 2009). Long-term follow-up studies suggest that hyperactive and impulsive behaviors tend to decrease with age, while inattentive symptoms show greater persistence and may stay life-long (Biederman et al., 2000; Molina et al., 2009). Besides age effects, gender-related differences are also observed in ADHD. Boys are more likely to be diagnosed with ADHD than girls. In the 2011 National Health Interview Survey, the estimated prevalence of ADHD in males was 12 percent; by contrast, in females the estimated prevalence was only

4.7 percent (Perou et al., 2013). Meanwhile, symptom profiles differ between males and females, with females more likely to be diagnosed with predominantly inattentive presentation (Biederman et al., 2005).

Recently, an increasing number of adolescents and young adults have presented to clinics with ADHD symptoms started after 12 or even later in life (Moffitt et al., 2015). A recent longitudinal study found that 2.5%–10.7% of subjects with ADHD first emerge in adolescence or adulthood, with the majority of adults with ADHD (67.5%–90.0%) not experiencing symptoms in childhood (Agnew-Blais et al., 2016). It is still debatable about whether adult ADHD is defined as a childhood-onset neurodevelopmental disorder or not (Moffitt et al., 2015; Agnew-Blais et al., 2016). Results from a recent large sample Multimodal Treatment Study of ADHD (MTA) suggest that common sources of impairing late-onset ADHD symptoms in adolescence and young adulthood were heavy substance use and comorbid psychiatric or learning problems (Sibley et al., 2018).

Although DSM-5-based diagnosis of ADHD offers a common language and standard criteria for identification of the disorder and its sub-types (presentations), emotion lability-based sub-types were also suggested. Deficient emotional self-regulation has been suggested to be a core component of ADHD (Shaw et al., 2014). A recent meta-analysis of 77 studies with a total of 32,044 participants observed the associations of ADHD with impaired emotional reactivity/negativity/liability, and empathy/callous-unemotional traits of emotional regulation (Graziano and Garcia, 2016). A study conducted by (Karalunas et al., 2014) attempted to define three distinct subtypes of ADHD based on temperament profiles, including a “Mild” type, whose members are characterized only by deficits in core ADHD symptom domains; a “Surgent” type, characterized by high levels of positive

approach-motivated behaviors and activity level, shorter cardiac pre-ejection period, parasympathetic withdrawal in response to positive emotions, and atypical amygdala connectivity to medial frontal areas; and an "Irritable" type, characterized by high levels of negative emotionality, weak parasympathetic response to negative emotional stimuli, reduced amygdala-insula connectivity, and a doubling of risk for onset of new behavioral or emotional disorders.

Nevertheless, categorical classification system has its shortcomings regarding to what is the best cut-off thresholds of the symptoms characterized as a dimension, the large number of intermediate cases, comorbidities, etc. (Lilienfeld and Treadway, 2016). The NIH Research Domain Criteria (RDoC) represents a new research framework for investigating ADHD by integrating multi-level information (from genomics and circuits to behavior and self-reports) to explore basic dimensions of functioning that span the full range of human behavior from normal to abnormal. The goal of RDoC is to understand the nature of mental health and illness in terms of varying degrees of dysfunctions in general psychological/biological systems (Harkness et al., 2014; Lilienfeld and Treadway, 2016).

Treatment strategies for ADHD symptoms include medication-based, behavior-based, and combined interventions (Antshel et al., 2011; Sibley et al., 2014). Stimulant medications that affect the dopaminergic system, including methylphenidate (Ritalin, Concerta, Metadate, Methylin) and certain amphetamines (Dexedrine, Dextrostat, Adderall), are most commonly prescribed for ADHD. Besides medications, behavior-based treatments, including education and/or behavior therapy, and social skills training have also been implemented in practice for ADHD interventions (Pelham et al., 2016; DuPaul et al., 2017; Anastopoulos et al., 2018; Chacko et al., 2018). Nevertheless,

there is yet no curative treatment for ADHD without thoroughly understanding its heterogeneous and developmental pathophysiological mechanisms.

Psychiatric and behavioral comorbidities, such as depression, anxiety disorder, bipolar disorder, substance use, and personality disorders, often co-occur with ADHD and result in increased difficulties for appropriate diagnoses and treatments (Barkley and Brown, 2008; Rosler et al., 2010; Mao and Findling, 2014; Katzman et al., 2017). Although different pharmacological treatment strategies have been applied to ADHD patients with various comorbidities, evidence from a large body of studies showed that treatment responses from different patients are widely different in terms of the types of pharmacological treatments, dosage requirements, tolerability, response rates, and adverse-event profiles (Spencer et al., 1996; Efron et al., 1997; Pliszka, 2007; Newcorn et al., 2008; Victor et al., 2009; Wilens et al., 2011; Hodgkins et al., 2012). Multiple factors may contribute to the treatment response heterogeneity in ADHD. For instance, the treatment response of methylphenidate was suggested to rely on inter-individual variability in the amount of dopamine released by neurons (Volkow et al., 2002; Berridge et al., 2006; Hannestad et al., 2010). Certain polymorphisms, such as the 40-pb variable number tandem repeat polymorphism, noradrenaline, and serotonin transporter genes, were also suggested to be associated with the treatment response to methylphenidate (Winsberg and Comings, 1999; Yang et al., 2004; Thakur et al., 2010; Froehlich et al., 2011b; Bidwell et al., 2017; da Silva et al., 2018). However, recent meta-analyses did not support this polymorphism association hypothesis (Kambeitz et al., 2014; Bonvicini et al., 2016). Treatment response heterogeneity in ADHD, especially those with other psychiatric and behavioral comorbidities, should be more closed investigated in future.

Neurocognitive impairments are hypothesized as a core part of ADHD symptomatology. Neurocognitive impairments including, but not limited to, domains of sustained attention or vigilance, executive function, working memory, and self-regulation, have been frequently reported in individuals with ADHD, even after controlling the effect of ADHD presentation, age and gender (Nigg et al., 2005; Willcutt et al., 2005). Notably, the nature of neurocognitive deficits is highly variable across individuals and some have no such difficulties. A number of theoretical models of the neurobiological and pathological substrates of ADHD have emerged in the past three decades, aimed at providing systematic guides for more effective strategies in diagnosis and treatment. These models included: (1) cognitive and motivational impairment models (Tannock et al., 1995; Barkley, 1997; Sonuga-Barke, 2003; Martinussen et al., 2005; Martinussen and Tannock, 2006; Rogers et al., 2011; Mawjee et al., 2017; Simone et al., 2018); (2) cognitive-energetic model (CEM) (Sergeant, 2000); and (3) neurodevelopmental model (Halperin and Schulz, 2006).

1.1.3 Neuroimaging Studies of Attention-Deficit/Hyperactivity Disorder

A large number of functional and structural neuroimaging studies have been conducted to identify the neurobiological mechanisms of emergence of ADHD. They suggest that ADHD symptoms in children are associated with widespread functional and neuroanatomical alterations in prefrontal cortex (PFC), parietal lobe, anterior cingulate cortex (ACC), striatum, and thalamus, which are key components in the cortico-striato-thalamo-cortical (CSTC) loops that subserve attention and cognitive processing. Functional aberrations in the fronto-thalamal/fronto-striatal circuitries have also been frequently reported to link with symptom onset in children with ADHD. For instance, significantly reduced task-responsive activation in frontal cortex, parietal areas,

anterior cingulate cortex, thalamus, and striatum, and their functional connectivities were observed in children with ADHD relative to the group-matched controls, when performing behavioral tasks that assess attentional and inhibitory control functions (such as the go/no-go task, stop signal task, continuous performance task, stroop task, etc.) (Durston et al., 2003; Booth et al., 2005; Durston et al., 2006; Pliszka et al., 2006; Smith et al., 2006; Durston et al., 2007; Suskauer et al., 2008; Li et al., 2012a; Li et al., 2013). Significantly reduced activation in these cortical and subcortical regions have also been consistently reported in children with ADHD relative to controls, when performing tasks assessing working memory, decision making, reward processing, and interference control functions (Vaidya et al., 2005; Konrad et al., 2006; Vance et al., 2007; Cao et al., 2008). Additionally, substantial structural magnetic resonance imaging (MRI) and Diffusion Tensor Imaging (DTI) studies have suggested that gray- and/or white-matter (GM/WM) structural underdevelopment in frontal lobe, thalamus, and striatum significantly contribute to the emergence of ADHD during childhood (Ellison-Wright et al., 2008; Xia et al., 2012). Structural neuroimaging studies have found ADHD symptoms in childhood to be associated with decreased regional GM volume in frontal cortex, striatum and cerebellum (Ellison-Wright et al., 2008; Bledsoe et al., 2011; Mahone et al., 2011). Reduced regional cortical GM thickness in frontal and parietal cortices have also been linked with ADHD symptoms (Batty et al., 2010; Almeida Montes et al., 2013). WM structural deficits, especially reduced WM volume and/or fractional anisotropy (FA) in the fronto-parietal, fronto-limbic, corona radiata, cerebellar- and temporo-occipital, and internal capsule fiber tracts have been consistently demonstrated in children with ADHD

(Durston et al., 2004; Nagel et al., 2011; Peterson et al., 2011; Qiu et al., 2011; Xia et al., 2012).

The majority of existing clinical and neuroimaging studies in ADHD have focused on understanding the neural correlates of symptoms in cross-sectional samples of children and young adults. Far fewer studies have examined neural substrates associated with the diverse adult outcomes of childhood ADHD. For instance, altered task-driven or spontaneous neural activities in prefrontal cortex, thalamus, and striatum, and their functional connectives, have been found to significantly associate with increased inattentive and/or impulsive symptoms in children with ADHD (Rubia et al., 1999; Durston, 2003; Bush et al., 2005; Yang et al., 2011; Cubillo et al., 2012; Li et al., 2012a). Increasingly, neuroimaging studies have found that optimal structural/functional development in fronto-subcortical pathways may contribute to symptom reduction and remission of ADHD in adulthood. For instance, a longitudinal study found that persistently decreased GM thickness in dorsolateral prefrontal, middle frontal, and inferior parietal regions, and reduced WM FA in left uncinated and inferior frontal-occipital fasciculi were associated with a greater number of ADHD symptoms persisting into adulthood (Shaw et al., 2013; Shaw et al., 2015). Proal et al. reported that adults with persistent ADHD had thinner cortical thickness relative to the remitted ADHD in prefrontal region (Proal et al., 2011). In addition, greater thalamo-prefrontal functional connectivity during cue-evoked attention process (Clerkin et al., 2013), and greater within-frontal functional connectivity during resting-state (Francx et al., 2015), have been observed in adult ADHD remitters (ADHD-R) relative to the ADHD persisters (ADHD-P). However, neuroimaging findings are widely inconsistent, partially due to the sample biases, differences of the implemented

imaging and analytic techniques, and the limitations of the traditional parametric models for group comparisons. Indeed, traditional statistical methods (e.g., t-tests, analysis of variance (ANOVA), correlation, etc.) estimate group differences only within a voxel or region of interest (ROI) at a time without having the capacity to explore how ROIs interact in linear and/or non-linear ways, as they quickly become overburdened when attempting to combine predictors and their interactions from high dimensional imaging data sets (Sun et al., 2009).

1.1.4 Machine Learning Studies of Attention-Deficit/Hyperactivity Disorder

The current ADHD diagnostic standards are fully clinical symptom-based, and relay on information collected from multiple sources through subjective observations, which often cause biases and inconsistencies of the diagnoses. Although existing neuroimaging studies utilized traditional parametric models to identify relatively objective biomarkers, compared to traditional parametric models, multivariate machine learning techniques are able to leverage high dimensional information simultaneously to understand how variables jointly distinguish between groups (Greenstein et al., 2012). In literature, support vector machine (SVM) is the most frequently applied machine learning classifier in neuroimaging data from children with ADHD, which has been aided by recursive feature elimination (RFE), temporal averaging, principle component analysis (PCA), fast Fourier transform (FFT), independent component analysis (ICA), 10-fold cross-validation (CV), hold-out, and leave-one-out CV (LOOCV) techniques, to distinguish children with ADHD from normal controls (NCs) (Brown et al., 2012; Chang et al., 2012; Cheng et al., 2012; Colby et al., 2012; Fair et al., 2012; Johnston et al., 2014; Iannaccone et al., 2015; Du et al., 2016; Yasumura et al., 2017; Sen et al., 2018). The commonly reported most important features

(according to importance score) that contribute to successful group discrimination included functional connectivity of bilateral thalamus, functional connectivity, surface area, cortical curvature and/or voxel intensity in frontal lobe, cingulate gyrus, temporal lobe, etc. (Brown et al., 2012; Colby et al., 2012; Iannaccone et al., 2015). SVM has also been applied to structural MRI and DTI data collected from adults with ADHD and controls, which reported between-group differences in widespread GM and WM regions in cortices, thalamus, and cerebellum (Chaim-Avancini et al., 2017). Meanwhile, neural network-based techniques, including deep belief network, fully connected cascade artificial neural network, convolutional neural network, extreme learning machine, and hierarchical extreme learning machine, have also been utilized to structural MRI and resting-state functional MRI (fMRI) data in children with ADHD and controls (Peng et al., 2013; Kuang and He, 2014; Deshpande et al., 2015; Qureshi et al., 2016; Qureshi et al., 2017; Zou et al., 2017). The most important group discrimination predictors identified by these neural network studies included functional connectivity within cerebellum, functional connectivity, surface area, cortical thickness and/or folding indices of frontal lobe, temporal lobe, occipital lobe and insula (Peng et al., 2013; Deshpande et al., 2015; Qureshi et al., 2017). In addition, principle component-based Fisher discriminative analysis (PC-FDA) (Zhu et al., 2008), Gaussian process classifiers (GPC) (Lim et al., 2013; Hart et al., 2014), and multiple kernel learning (Dai et al., 2012; Ghiassian et al., 2016) have also been used in functional and structural MRI data to discriminate children with ADHD from controls. More details of existing machine leaning studies in ADHD are provided in Table 1.1. These existing studies have either utilized features representing regional/voxel brain properties collected from only single imaging modality, or the

combination of two modalities (mostly structural MRI and resting-state fMRI) (Brown et al., 2012; Fair et al., 2012; Hart et al., 2014; Johnston et al., 2014; Iannaccone et al., 2015), or reported poor accuracy (Dai et al., 2012; Zou et al., 2017; Sen et al., 2018). Some studies did not conduct the very necessary step of estimating the most important features that contribute to accurate classifications (Chang et al., 2012; Dai et al., 2012; Kuang and He, 2014; Tenev et al., 2014; Qureshi et al., 2016; Zou et al., 2017; Sen et al., 2018). In this field, systems-level functional and structural features, such as global and regional topological properties from functional brain networks during cognitive processes and WM tract properties have not been considered. In addition, relations between the suggested predictors from imaging features and clinical/behavioral symptoms in samples of ADHD patients, which can provide important clinical context, have not been studied.

Table 1.1 Existing Machine Learning Studies in Neuroimaging Data from Children and/or Adults with ADHD and Group-matched Controls **(Continued)**

Author, Year	Models	Validation	Features	Feature Type	Predictors	Accuracy	AUC
Children with ADHD							
(Brown et al., 2012)	SVM	10-fold CV	rs-fMRI	Voxel	FC of bilateral thalamus, bilateral TL, MFG, PCC, Cerebellum	0.71	N/A
(Colby et al., 2012)	SVM	Hold-out	sMRI, rs-fMRI	Voxel	FC, SA, cortical curvature in FL, CT and FC of CG and TL	0.55	N/A
(Dai et al., 2012)	MKL	10-fold CV	sMRI, rs-fMRI	ROI	N/A	0.68	0.71
(Deshpande et al., 2015)	FCCANN	LOOCV	rs-fMRI	ROI	FC of OFC and cerebellum	0.90	N/A
(Du et al., 2016)	SVM	10-fold CV	rs-fMRI	Network	HSIC score of Operculum, Insula, putamen, STG	0.95	0.97
(Eloyan et al., 2012)	Voting	Hold-out	sMRI, rs-fMRI	Voxel	FC between DM and DL in MC	0.78	N/A
(Fair et al., 2012)	SVM	LOOCV	rs-fMRI	ROI	FC in PFC, PL, Cerebellum	0.83	N/A
(Ghiassian et al., 2016)	MHPC	Hold-out	sMRI, rs-fMRI	Voxel	Voxel intensity and functional activation in FL, Cerebellum	0.70	N/A
(Hart et al., 2014)	GPC	LOOCV	tb-fMRI	Voxel	Functional activation in PFC, CG, BG, Thalamus, PL	0.77	0.81
(Iannaccone et al., 2015)	SVM	LOOCV	tb-fMRI	Voxel	Functional activation in SFG, PCC, TL, Brainstem, Cerebellum	0.78	0.82
(Johnston et al., 2014)	SVM	LOOCV	sMRI	Voxel	Volume of Brainstem	0.93	N/A
(Peng et al., 2013)	ELM	LOOCV	sMRI	ROI	SA and FI of FL, SA and FI in TL, SA and volume in OL, FI of Insula	0.90	0.88
(Qureshi et al., 2016)	H-ELM	10-by-10 Nested CV	rs-fMRI	ROI	N/A	0.71	N/A
(Qureshi et al., 2017)	ELM	Random CV	sMRI, rs-fMRI	ROI	CT and FC in SFG and MTG	0.93	N/A
(Sen et al., 2018)	SVM	Hold-out	sMRI, rs-fMRI	Voxel	N/A	0.67	N/A
(Zou et al., 2017)	CNN	Hold-out	sMRI, rs-fMRI	Voxel	N/A	0.69	N/A

Table 1.1 (Continued) Existing Machine Learning Studies in Neuroimaging Data from Children and/or Adults with ADHD and Group-matched Controls

Author, Year	Models	Validation	Features	Feature Type	Predictors	Accuracy	AUC
Children with ADHD							
(Yasumura et al., 2017)	SVM	3-fold CV	fNIRS	ROI	Oxygenated hemoglobin change in PFC	0.86	0.898
(Zhu et al., 2008)	PC-FDA	LOOCV	rs-fMRI	Voxel	ReHo of PFC, ACC, Thalamus	0.85	N/A
(Zhang-James et al., 2019)	ELTs	Hold-out	sMRI	ROI	ICV, SA of FL, volume of Caudate and Thalamus	0.61	0.67
(Kuang and He, 2014)	DBN	Hold-out	rs-fMRI	Voxel	N/A	0.45	N/A
(Chang et al., 2012)	SVM	3-fold CV	sMRI	ROI	N/A	0.70	N/A
(Lim et al., 2013)	GPC	LOOCV	sMRI	Voxel	Voxel intensity in FL, Premotor, TL, Brainstem	0.79	0.83
(Cheng et al., 2012)	SVM	LOOCV	rs-fMRI	ROI, Voxel	FC in FL and Cerebellum	0.76	N/A
Adults with ADHD							
(Tenev et al., 2014)	Voting	10-fold CV	EEG	ROI	N/A	0.82	N/A
(Zhang-James et al., 2019)	ELTs	Hold-Out	sMRI	ROI	ICV, SA of FL, volume of Caudate and Thalamus	0.62	0.66
(Chaim-Avancini et al., 2017)	SVM	10-fold CV	sMRI, DTI	ROI, Voxel	GM and WM intensity across FL, TL, OL, Thalamus, Cerebellum, FA	0.66	0.71

(ICV: intracranial volume; MFG: middle frontal gyrus; MC: motor cortex; SMC: sensorimotor cortex; PFC: prefrontal cortex; FL: frontal lobe; PL: parietal lobe; TL: temporal lobe; OL: occipital lobe; BG: basal ganglia; CG: cingulate gyrus; MTG: middle temporal gyrus; OFC: orbitofrontal cortex; STG: superior temporal gyrus; PFC: prefrontal cortex; MHPC: (f)MRI HOG-feature-based patient classification; GPC: Gaussian process classifiers; ACC: anterior cingulate cortex; PCC: posterior cingulate cortex; H-ELM: hierarchical extreme learning machine; ELM: extreme learning machine; DBN: deep belief network; CNN: convolutional neural network; MKL: multiple kernel learning; PC-FDA: principle component-based Fisher discriminative analysis; ROI: region of interest; Acc: accuracy; CV: cross validation; LOOCV: leave-one-out cross validation; sMRI: structural magnetic resonance imaging; rs-fMRI: resting-state functional magnetic resonance imaging; tb-fMRI: task-based functional magnetic resonance imaging; FC: functional connectivity; SA: surface area; CT: cortical thickness; FI: folding index; ReHo: regional homogeneity; HSIC: Hilbert-Schmidt Independence Criterion; DM: dorsomedial; DL: dorsolateral; GM: gray matter; WM: white matter; N/A: not available)

1.2 Significance, Objective and Specific Aims of Ph.D. Dissertation Research

A neurodevelopmental double dissociation model was proposed based on existing studies in which the early onset of childhood ADHD is suggested to associate with dysfunctional subcortical structures that remain static throughout the lifetime; while diminution of symptoms over development could link to optimal development of prefrontal cortex (Halperin and Schulz, 2006). The current existing studies only assessed basic measures including regional brain activation and connectivity, which have limited capacity to characterize the functional brain as a high performance parallel information processing system. The field lacks systems-level investigations of the structural and functional patterns that significantly contribute to the symptom remission and persistence in adults with childhood ADHD. Moreover, neuroimaging findings are widely inconsistent, partially due to the sample biases, differences of the implemented imaging and analytic techniques, and the limitations of the traditional parametric models for group comparisons. Traditional statistical methods (e.g., t-test, ANOVA, correlation, etc.) estimate group differences only within a voxel or ROI at a time without having the capacity to explore how ROIs interact in linear and/or non-linear ways, as they quickly become overburdened when attempting to combine predictors and their interactions from high dimensional imaging data sets (Sun et al., 2009). It is urgent to fill this gap, which could aid clinical prediction and the development of individualized pharmacological and neurobehavioral interventions that yield enduring benefits and improve long-term outcomes. The overarching goal of this dissertation research is to assess functional and structural neurobiological substrates associated with variability of clinical adult outcomes in childhood ADHD by using machine learning techniques. The central hypothesis of this dissertation research is that the

remission of ADHD is associated with the optimal functional and structural frontal and associated circuits. To fill this gap, this dissertation research proposes three specific aims:

Specific Aim 1: Identify the topological features of cue-evoked attention processing network, which associates with remission and persistence in adults with childhood ADHD.

Hypothesis: Relative to ADHD-P, the ADHD-R will show optimal topological organization of the functional brain network for cue-evoked attention processing may be associated with symptom remission in adults with childhood ADHD, including significantly higher nodal efficiency in frontal lobe; improved functioning of fronto-parietal, subcortico-frontal circuits and the circuit-associated nodes, such as parietal lobe and thalamus.

Specific Aim 2: Delineate the GM and WM structural correlates of remission and persistence of childhood ADHD.

Hypothesis: Compared to ADHD-P, ADHD-R will present optimal structural development associated with the frontal and parietal lobes, such as greater regional GM thickness, higher FA of the WM tracts that connect subcortical structures (i.e., thalamus, caudate) and frontal/parietal cortices.

Specific Aim 3: Construct prediction models by using ensemble learning techniques, and detected features in Aims 1 and 2, to identify the most important features that determine the diverse adult outcomes of childhood ADHD.

Hypothesis: Structural and functional alterations in frontal, parietal and subcortical areas and their interactions would significantly contribute to accurate discrimination of ADHD probands (adults diagnosed with ADHD in childhood) from controls; while abnormal fronto-parietal hyper-communications in right hemisphere would play an important role in

inattentive and hyperactive/impulsive symptom persistence in adults with childhood ADHD. We also hypothesized that classification performance parameters (accuracy, area under the curve (AUC) of the receiver operating characteristics (ROC), etc.) derived from ensemble learning technique (ELT)-based procedures would be superior to those of basic model-based procedures.

This Ph.D. dissertation research is significant because 1) understanding the neurobiological basis determining the diverse adult outcomes of childhood ADHD is urgently needed and vitally important for public health that can inform and ultimately guide individualized strategies for long-term treatment and interventions; 2) the utilization of ELT will greatly enhance the classification performance for identifying reliable neurobiological markers for remission and persistence of childhood ADHD.

CHAPTER 2

GENERAL METHODOLOGY

2.1 Participants

This study was approved by the institutional review board (IRB) of New Jersey Institute of Technology (NJIT), City University of New York (CUNY), and Icahn School of Medicine at Mount Sinai. Participants provided signed informed consent and were reimbursed for their time and travel expenses. Seventy-two young adults [mean (standard deviation (SD)) age 24.4 (2.1) years] who provided good quality data from multimodal neuroimaging and clinical assessments, participated in this study. There were 36 ADHD probands diagnosed with ADHD combined-type (ADHD-C) in childhood and 36 group-matched comparison subjects with no history of ADHD. Among the 36 ADHD probands, 18 were classified as ADHD-R, who were endorsed no more than 3 inattentive or 3 hyperactive/impulsive symptoms in adulthood and had no more than 5 symptoms in total. The other 18 probands were classified as ADHD-P, endorsing at least five inattentive and/or hyperactive/impulsive symptoms in their adulthood and at least 3 symptoms in each domain.

ADHD patients were recruited from a childhood follow-up study, which consisted of three time points evaluations for childhood ADHD patients. The initial sample consisted of 106 young adults who had been clinically followed since childhood, including 60 probands who were diagnosed with ADHD-C at their 7-11 years of age, were screened during their young adulthood. The comparison group was recruited during an adolescent follow-up study with two time points evaluations. A total of 46 young adults, who were evaluated as non-ADHD in childhood, were re-evaluated for the current study. Thirty-six

of them met the inclusion criteria for controls (had no history of childhood ADHD and no more than three inattentive or hyperactive/impulsive symptoms) and provided usable data.

Those with ADHD were recruited when they were 7-11 years-old and subsequently clinically followed. Childhood diagnoses were based on teacher rating using the IOWA Conners' Teachers Rating Scale (Loney and Milich, 1982) and parent interview using the Diagnostic Interview Schedule for Children version 2 (DISC-2) (Shaffer et al., 1989). Exclusion criteria in childhood were chronic medical illness; neurological disorder; diagnosis of schizophrenia, autism spectrum disorder, or chronic tic disorder; Full Scale IQ<70; and not speaking English. The never ADHD comparison group was recruited in adolescence, as part of an adolescent follow-up of the ADHD sample, and history of ADHD was ruled out using the ADHD module of the DISC-2, the IOWA Conners, and the Schedule for Affective Disorders and Schizophrenia for School-Age Children (K-SADS) (Kaufman et al., 1997), which was administered to both the parent and adolescent. Adult psychiatric status was assessed using the Structured Clinical Interview for DSM-IV Axis I Disorders (First et al., 2002), supplemented by a semi-structured interview for ADHD that was adapted from the K-SADS and the Conners' Adult ADHD Diagnostic Interview for DSM-IV (Epstein et al., 2006). Raw scores of inattentive and hyperactive/impulsive symptoms from the Conner's Adult ADHD Self-Rating Scale (CAARS) were normalized into T-scores based on DSM-IV standard, and were used as dimensional measures for inattentive and hyperactive/impulsive behaviors. Exclusion criteria in adulthood were psychotropic medication that could not be discontinued and conditions that would preclude MRI (e.g., metal in body, pregnancy, too obese to fit in scanner).

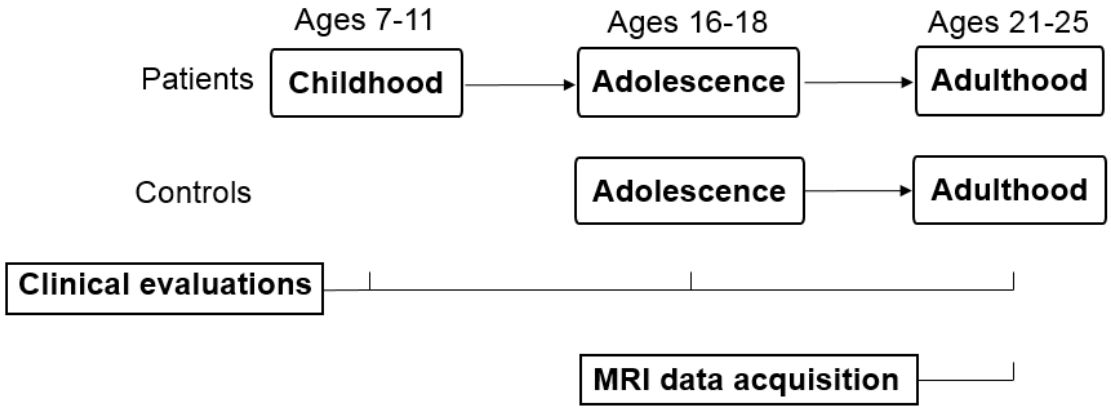

Figure 2.1 Clinical evaluations and MRI data acquisition for participants.

Thirty-six probands had been treated with short-acting psychostimulants. Mean duration of treatment was 2.5 years (SD=3.85) for ADHD-R and 4.8 years (SD=4.19) for ADHD-P (t=-1.65, p=0.108). Two subjects with persistent ADHD were taking psychostimulants at the time of this study, and underwent a 48-hour medication wash-out period before MRI scan. Therefore, 36 NCs and 36 ADHD probands were involved in this dissertation research. Thirty-six ADHD probands were further split into 18 ADHD-P and 18 ADHD-R. More specifically, thirty-three NCs and thirty-five ADHD probands, including 17 ADHD-P and 18 ADHD-R provided usable data in Specific Aim 1; thirty-five NCs and thirty-two ADHD probands, including 16 ADHD-P and 16 ADHD-R provided usable data in Specific Aim 2; and 36 NCs and 36 ADHD probands including 18 ADHD-P and 18 ADHD-R provided usable data in Specific Aim 3.

The initial exclusion criteria for participants were: Chronic medical illness or was taking systemic medication; Diagnosed neurological disorder; Diagnosis of schizophrenia, autism, pervasive developmental disorder or a chronic tic disorder; Full Scale IQ<70; Not attending school; Not English speaking. In addition, adults were excluded from the

imaging study if they: Currently have a chronic medical illness or are taking systemic medication; Have had a traumatic head injury or been diagnosed with neurological disorder; Have taken a psychotropic medication (including stimulants) within the past 3 months; Present with a urine screen indicating recent illicit drug use on the day of assessment; Have history of surgery involving metal implants, possible metal fragments in the eyes, braces, or a pacemaker; Have a history of claustrophobia; Are pregnant (females only). Summary of the demographic information of 72 participants involved in this study were shown in Table 2.1.

Table 2.1 Demographic Information of 72 Participants Involved in this Dissertation **(Continued)**

Subject ID	Group2	Group3	Age	Gender	T-score/IN	T-score/HI
6	NC	NC	23.99	M	36	46
9	NC	NC	26.15	M	36	44
17	NC	NC	25.90	M	51	46
21	NC	NC	28.53	M	36	39
23	NC	NC	27.51	M	40	39
24	NC	NC	23.88	M	61	44
26	NC	NC	24.20	F	45	45
30	NC	NC	26.10	M	40	41
34	NC	NC	25.78	M	43	41
36	NC	NC	31.10	M	48	54
58	NC	NC	24.81	M	46	54
61	NC	NC	26.30	M	46	46
63	NC	NC	24.62	M	59	59
75	NC	NC	25.88	M	36	35
80	NC	NC	22.96	M	53	44
90	NC	NC	24.93	F	54	43
95	NC	NC	23.22	M	36	35
97	NC	NC	26.31	M	48	54
100	NC	NC	24.48	M	61	54
113	NC	NC	25.02	M	53	39
120	NC	NC	23.28	M	39	44
128	NC	NC	23.37	M	39	35
154	NC	NC	22.03	F	35	39
158	NC	NC	23.74	M	40	39
159	NC	NC	22.55	M	46	39
162	NC	NC	22.36	M	36	33
165	NC	NC	22.65	M	36	39
169	NC	NC	22.44	M	59	44

Table 2.1 (Continued) Demographic Information of 72 Participants Involved in this Dissertation

Subject ID	Group2	Group3	Age	Gender	T-score/IN	T-score/HI
171	NC	NC	22.38	M	43	39
172	NC	NC	22.21	M	43	46
176	NC	NC	22.64	M	48	39
179	NC	NC	21.15	F	45	45
185	NC	NC	23.14	M	66	46
187	NC	NC	20.72	F	37	33
188	NC	NC	20.81	M	59	44
203	NC	NC	27.56	M	48	41
1	ADHD	ADHD-R	25.88	M	61	61
5	ADHD	ADHD-R	24.96	M	69	71
7	ADHD	ADHD-R	25.56	M	48	39
8	ADHD	ADHD-R	25.98	M	36	35
18	ADHD	ADHD-R	25.32	M	40	44
28	ADHD	ADHD-R	24.29	M	46	44
29	ADHD	ADHD-R	25.05	M	61	46
31	ADHD	ADHD-R	27.76	M	48	41
78	ADHD	ADHD-R	25.71	M	46	54
112	ADHD	ADHD-R	23.64	M	53	46
131	ADHD	ADHD-R	22.62	M	59	46
140	ADHD	ADHD-R	22.20	M	36	39
163	ADHD	ADHD-R	21.91	M	49	47
175	ADHD	ADHD-R	25.07	M	48	44
202	ADHD	ADHD-R	24.13	F	35	34
207	ADHD	ADHD-R	21.36	F	72	52
208	ADHD	ADHD-R	30.44	M	41	41
209	ADHD	ADHD-R	24.35	M	49	47
2	ADHD	ADHD-P	26.17	M	90	84
10	ADHD	ADHD-P	23.91	M	56	56
22	ADHD	ADHD-P	23.84	M	53	49
42	ADHD	ADHD-P	27.32	M	51	56
43	ADHD	ADHD-P	27.77	F	80	68
51	ADHD	ADHD-P	24.95	F	63	70
55	ADHD	ADHD-P	23.81	M	74	69
67	ADHD	ADHD-P	23.42	M	74	86
68	ADHD	ADHD-P	25.76	M	61	61
81	ADHD	ADHD-P	26.22	M	69	59
106	ADHD	ADHD-P	24.79	M	66	69
133	ADHD	ADHD-P	22.83	M	64	44
139	ADHD	ADHD-P	22.51	M	51	59
144	ADHD	ADHD-P	22.16	M	74	54
174	ADHD	ADHD-P	22.95	F	47	41
182	ADHD	ADHD-P	21.05	M	64	64
189	ADHD	ADHD-P	27.54	M	51	59
210	ADHD	ADHD-P	24.35	F	49	52

2.2 General Techniques for Magnetic Resonance Imaging

2.2.1 The Components of a Magnetic Resonance Imaging Scanner

MRI is one of the most commonly utilized instrumentations for disease detection, diagnosis, and treatment monitoring. MRI is a non-invasive imaging technology that produces three dimensional detailed anatomical/functional images. It is based on sophisticated technology that excites and detects the change in the direction of the rotational axis of protons found in the water that makes up living tissues. The three main components of an MRI scanner are the static magnetic field, radiofrequency coils, and gradient coils, which together allow maintaining the imaging collection. In addition, shimming coils, which ensure the homogeneity of the static magnetic field; specialized computer systems for controlling the scanner; the experimental task; and physiological monitoring equipment are other important components for functional MRI (fMRI).

2.2.2 Basic Principles of Magnetic Resonance Signal Generation

MRI employ powerful static magnets which produce a strong magnetic field that forces protons in the body to align with that field. An equilibrium state exists when the human body is placed in any magnetic field, such that the net magnetization of atomic nuclei (e.g., hydrogen) within the body becomes aligned with the magnetic field. The radiofrequency coils send electromagnetic waves that resonate at a particular frequency, as determined by the strength of the magnetic field, into the body, perturbing this equilibrium state. This process is known as excitation. When atomic nuclei are excited, they absorb the energy of the radiofrequency pulse. But, when the radiofrequency pulse ends, the hydrogen nuclei return to the equilibrium state and release the energy that was absorbed during excitation. The resulting release of energy can be detected by the radiofrequency coils, in a process

known as reception. The combination of a static magnetic field and a radiofrequency coil allows detection of MR signal, but MR signal alone cannot be used to create an image. The fundamental measurement in MRI is merely the amount of current through a coil, which in itself has no spatial information. By introducing magnetic gradients superimposed upon the strong static magnetic field, gradient coils provide the final component necessary for imaging. The purpose of a gradient coil is to cause the MR signal to become spatially dependent in a controlled fashion, so that different locations in space contribute differently to the measured signal over time. Similar to the radiofrequency coil, the gradient coils are only used during image acquisition, as they are typically turned on briefly after the excitation process to provide spatial encoding needed to resolve an image. To make the recovery of spatial information as simple as possible, gradient coils are used to generate a magnetic field that increases in strength along one spatial direction. The spatial directions used are relative to the main magnetic field, with z going parallel to the main field and x and y going perpendicularly to the main field. In combination with three different coils, the detected electromagnetic pulse defines the raw MR signal. The electromagnetic signal is then acquired in the frequency space and transformed using inverse Fourier transform to generate 3-D images.

Figure 2.2 Diagram of traditional MRI scanner.

The basis of MRI is the directional magnetic field, or moment, associated with charged particles in motion. Nuclei containing an odd number of protons and/or neutrons have a characteristic motion or precession. Because nuclei are charged particles, this precession produces a small magnetic moment. When a human body is placed in a large magnetic field, many of the free hydrogen nuclei align themselves with the direction of the magnetic field. The nuclei precess about the magnetic field direction like gyroscopes. This behavior is termed Larmor precession. The frequency of Larmor precession is proportional to the applied magnetic field strength as defined by the Larmor frequency, ω_0:

$$\omega_0 = \gamma B_0 \tag{2.1}$$

where γ is the gyromagnetic ratio and B_0 is the strength of the applied magnetic field. The gyromagnetic ratio is a nuclei specific constant. For hydrogen, $\gamma = 42.6\ MHz/Tesla$. To obtain an MR image of an object, the object is placed in a uniform magnetic field, B_0. As a result, the object's hydrogen nuclei align with the magnetic field and create a net magnetic

moment, M, parallel to B_0. This behavior is illustrated in Figure 2.3. Next, a radio-frequency (RF) pulse, B_{rf}, is applied perpendicular to B_0. This pulse, with a frequency equal to the Larmor frequency, causes M to tilt away from B_0 as in Figure 2.4(a). Once the RF signal is removed, the nuclei realign themselves such that their net magnetic moment, M, is again parallel with B_0. This return to equilibrium is referred to as relaxation. During relaxation, the nuclei lose energy by emitting their own RF signal (see Figure 2.4(b)). This signal is referred to as the free-induction decay (FID) response signal. The FID response signal is measured by a conductive field coil placed around the object being imaged. This measurement is processed or reconstructed to obtain 3D gray-scale MR images. To produce a 3D image, the FID resonance signal must be encoded for each dimension. The encoding in the axial direction, the direction of B_0, is accomplished by adding a gradient magnetic field to B_0. This gradient causes the Larmor frequency to change linearly in the axial direction. Thus, an axial slice can be selected by choosing the frequency of B_{rf} to correspond to the Larmor frequency of that slice. The 2D spatial reconstruction in each axial slice is accomplished using frequency and phase encoding. A "preparation" gradient, G_y, is applied causing the resonant frequencies of the nuclei to vary according to their position in the y-direction. G_y is then removed and another gradient, G_x, is applied perpendicular to G_y. As a result, the resonant frequencies of the nuclei vary in the x-direction due to G_x and have a phase variation in the y-direction due to the previously applied G_y. Thus, x-direction samples are encoded by frequency and y-direction samples are encoded by phase. A 2D Fourier Transform is then used to transform the encoded image to the spatial domain.

The voxel intensity of a given tissue type (i.e. white matter vs grey matter) depends on the proton density of the tissue; the higher the proton density, the stronger the FID response signal. MR image contrast also depends on two other tissue-specific parameters: the longitudinal relaxation time, T_1, and the transverse relaxation time, T_2. T_1 measures the time required for the magnetic moment of the displaced nuclei to return to equilibrium (ie. realign itself with B_0). T_2 indicates the time required for the FID response signal from a given tissue type to decay.

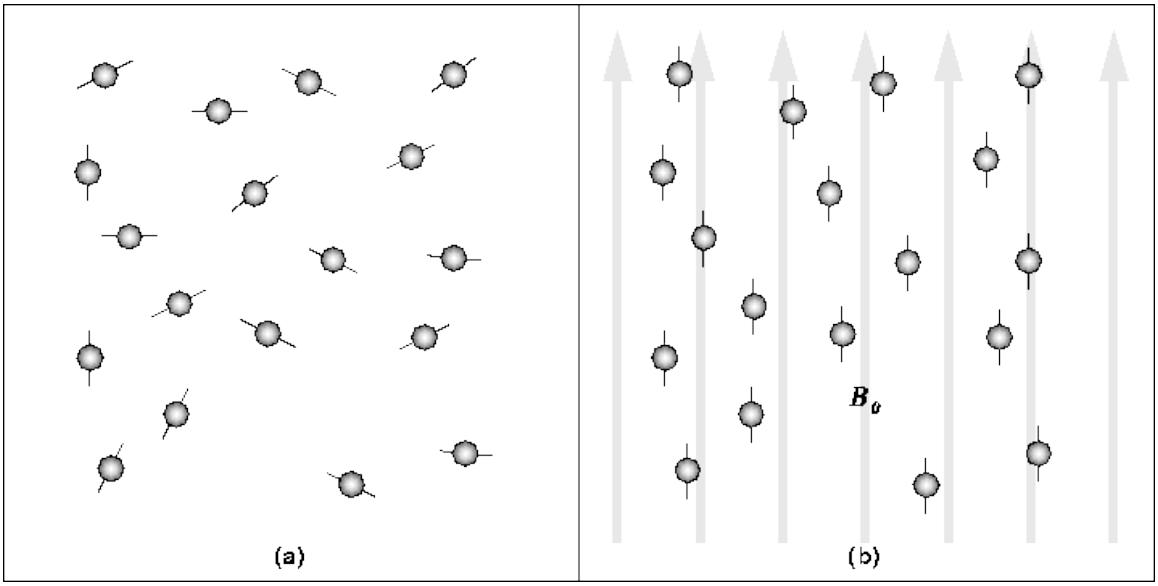

Figure 2.3 Hydrogen nuclei precession direction in the absence of a strong magnetic field and in the strong magnetic field.
Source: https://www2.cs.sfu.ca/~stella/papers/blairthesis/main/node11.html. Accessed on April 10, 2020.

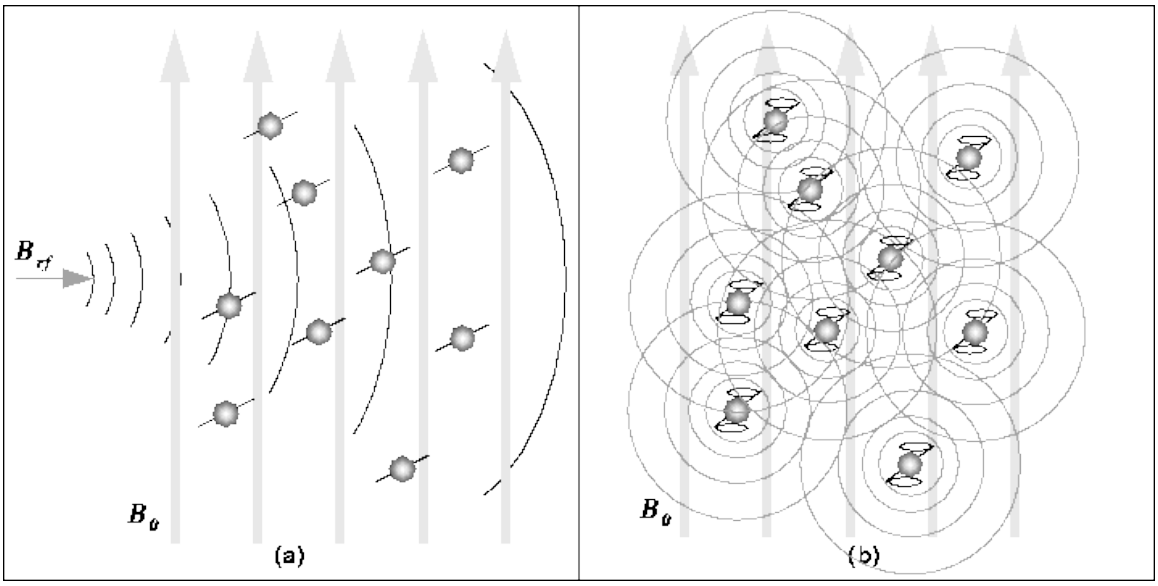

Figure 2.4 Excitation and relaxation conditions when the radiofrequency pulse is on and off.
Source: https://www2.cs.sfu.ca/~stella/papers/blairthesis/main/node11.html. Accessed on April 10, 2020.

When MR images are acquired, the RF pulse, B_{rf}, is repeated at a predetermined rate. The period of the RF pulse sequence is the repetition time (TR). The FID response signals can be measured at various times within the TR interval. The time between which the RF pulse is applied and the response signal is measured is the echo delay time (TE). By adjusting TR and TE the acquired MR image can be made to contrast different tissue types.

2.2.3 Functional Magnetic Resonance Imaging Technique

FMRI is based on MRI, which in turn uses magnetic resonance coupled with gradients in magnetic field to create images that can incorporate brain activity by detecting alterations associated with blood flow. FMRI creates images of physiological activity that is correlated with neuronal activity. The information processing activity of neurons increase their metabolic requirements. To meet these requirements, energy must be provided. The vascular system supplies cells with two fuel sources, glucose and oxygen, the latter bound

to hemoglobin molecules. The differential magnetic properties of oxygenated and deoxygenated hemoglobin are utilized to construct images based upon blood oxygen level dependent (BOLD) contrast. BOLD contrast is used in virtually all conventional fMRI experiments. BOLD contrast results from the change in magnetic field surrounding the red blood cells depending on the oxygen state of the hemoglobin. When fully oxygenated hemoglobin is diamagnetic and is magnetically indistinguishable from brain tissue. However, fully deoxygenated hemoglobin has four unpaired electrons and is highly paramagnetic (Thulborn et al., 1982). This paramagnetism results in local gradients in magnetic field whose strength depends on the oxygenated hemoglobin concentration. These endogenous gradients in turn modulate the intra- and extra-vascular blood's T2 and $T2^{*}$ relaxation times through diffusion and intravoxel dephasing, respectively.

2.2.4 Structural Magnetic Resonance Imaging Technique

As one of the most frequently utilized technique, structural MRI is currently widely utilized in clinical diagnosis, especially T1 and T2 contrasts. As for T1-weighted imaging, if the relative signal intensity of voxels within the image depends upon the T1 value of the tissue, at very TR, there is no time for longitudinal magnetization to recover and thus no MR signal is recorded for either tissue. Conversely, at very long TR, all longitudinal magnetization recovers for both tissues. So, at short and long TR values, the amount of longitudinal magnetization will be similar between the tissues. At intermediate TR, however, there are clear differences between them. The tissue that has a shorter T1 value recovers more rapidly and thus has greater MR signal. For any two tissues that differ in T1, there is an optimal TR value that maximally differentiates between them. To generate images sensitive to T1 contrast, a pulse sequence with intermediate TR and short TE is

needed. As for T2-weighted images, the amount of signal loss depends upon the time between excitation and data acquisition, or echo time. An optimal combination of TR and TE exists for any two tissues to maximize the T2 contrast between them. If an image is acquired immediately after excitation, such that the TE is very short, then little transverse magnetization will be lost regardless of T2 and thus there will be no T2 contrast. If the TE is too long, then nearly all transverse magnetization will be lost and still the image will have no T2 contrast. But at an intermediate TE, the difference in transverse magnetization can be maximized. To have exclusive T2 contrast, a long TR and intermediate TE is needed.

2.2.5 Diffusion Tensor Imaging Technique

Diffusion tensor imaging can quantify the relative diffusivity among directional components. For example, white matter, which is composed mostly of nerve fibers, shows prominent anisotropy, such that water molecules diffuse most quickly along the length of the fiber and most slowly across the width of the fiber. A scalar quantity known as FA can be computed for each voxel to express the preference of water to diffuse in an isotropic or anisotropic manner. FA values are bounded by 0 and 1 and are calculated using Equation (2.1), where D_x, D_y, and D_z represent the three principal axes of the diffusion tensor:

$$FA = \frac{\sqrt{(D_x - D_y)^2 + (D_y - D_z)^2 + (D_z - D_x)^2}}{\sqrt{2({D_x}^2 + {D_y}^2 + {D_z}^2)}} \tag{2.2}$$

FA values approaching 1 indicate that nearly all of the water molecules in the voxel are diffusing along the same preferred axis, while FA values approaching 0 indicated that the

water molecules are equally likely to diffuse in any direction. Fractional anisotropy provides important information about the composition of tissue within a voxel. Notably, some neurological diseases, such as multiple sclerosis and vascular dementia, are characterized by potentially severe WM pathology. The resulting axonal damage can be identified as decreased FA values in affected voxels. More complex forms of diffusion tensor imaging can track nerve fibers as they travel between functionally associated brain regions.

2.3 Multimodal Magnetic Resonance Imaging Data Acquisition Protocols for Projects 1 and 2

2.3.1 Functional Magnetic Resonance Imaging Data Acquisition Protocol

All participants were scanned using the same 3.0T Siemens Allegra (Siemens, Erlangen, Germany) head-dedicated MRI scanner. For each of the four runs, a total of 120 T2*-weighted volumes were acquired in the axial plane with a gradient-echo echo-planar sequence with TR=2,500 ms, TE=27 ms, flip angle=82°, matrix=64×64, slice thickness=4 mm, 40 slices, in-plane resolution=3.75 mm^2). Images were acquired with slices positioned parallel to the anterior commissure-posterior commissure line. Stimuli were projected onto a rear projection screen mounted at the head of the magnet bore that was viewed through a mirror on the head coil. In order to implement the coregistration procedure, a high resolution T2-weighted anatomical volume of the whole brain was also acquired at the same 40 slice locations with a turbo spin-echo pulse sequence with TR=4050 ms, TE=99 ms, flip angle=170°, field of view (FOV)=240 mm, matrix=512×336, slice thickness=4 mm, in-plane resolution=0.41 mm^2.

2.3.2 Structural Magnetic Resonance Imaging Data Acquisition Protocol

High resolution 3-dimensional T1-weighted structural MRI and DTI data were acquired using the same MRI scanner. T1-weighted data was acquired using magnetization prepared rapid gradient echo pulse sequence with TR=2,500 ms, TE=4.38 ms, the inversion time (TI)=1.1 s, flip angle=8°, voxel size=0.94 mm×0.94 mm×1 mm, FOV=256 mm×256 mm×256 mm.

2.3.3 Diffusion Tensor Imaging Data Acquisition Protocol

DTI data were acquired using an echo planar imaging pulse sequence with a b-value=1250 s/mm^2 along 12 independent orientations with TR=5.2 s, TE=80 ms, flip angle=90^o, voxel size=1.875 mm×1.875 mm×4 mm , FOV=128 mm×128 mm, imaging matrix=128×96, number of slices=63. One additional b0 image was collected for eddy current and head motion corrections in DTI data.

2.4 Statistical Methods Utilized in Projects 1, 2 and 3

2.4.1 Chi-square Test

A chi-square statistic is one way to show a relationship between two categorical variables. There are two types of chi-square tests using the chi-square statistic and distribution for different purposes, including chi-square goodness of fit test and chi-square test for independence. The chi-square is based on the differences between the observed values and those that would be expected if the variables were independent.

$$\chi^2 = \sum_{outcomes} \frac{(f_o - f_e)^2}{f_e} \tag{2.3}$$

where f_o is observed frequency of an outcome; f_e is expected frequency of that outcome, if null hypothesis is true. If these differences are small, there is little dependence between the variables; large differences indicate dependence. The actual chi-square statistic is the sum of squares of these differences in ratio to the expected value. A small chi-square statistic arises if the observed values are close to the values we would expect if the two variables were unrelated. A large chi-square statistic arises if the observed values are rather different from those we would expect from unrelated variables. If the chi-square statistic is large enough that it is unlikely to have occurred by chance, we conclude that it is significant and that the rows variable is not totally independent of the columns variable.

2.4.2 One Sample T-test

The one sample t-test is a statistical procedure used to determine whether a sample of observations could have been generated by a process with a specific mean.

$$t = \frac{M - \mu}{s_M} \tag{2.4}$$

where M is sample mean; μ is population mean; s_M is estimated standard error of mean. There are two kinds of hypotheses for a one sample t-test, the null hypothesis and the alternative hypothesis. The alternative hypothesis assumes that some difference exists between the true mean and the comparison value, whereas the null hypothesis assumes that no difference exists. The purpose of the one sample t-test is to determine if the null hypothesis should be rejected, given the sample data. The alternative hypothesis can assume

one of three forms depending on the question being asked. If the goal is to measure any difference, regardless of direction, a two-tailed hypothesis is used. If the direction of the difference between the sample mean and the comparison value matters, either an upper-tailed or lower-tailed hypothesis is used. The null hypothesis remains the same for each type of one sample t-test.

2.4.3 Independent Samples T-test

The independent samples t-test, also called two samples t-test, is an inferential statistical test that determines whether there is a statistically significant difference between the means in two unrelated groups.

$$t = \frac{M_A - M_B}{\sqrt{\frac{(n_A - 1){s_{M_A}}^2 + (n_B - 1){s_{M_B}}^2}{n_A + n_B - 2}(\frac{1}{n_A} + \frac{1}{n_B})}} \tag{2.5}$$

where M_A, M_B are sample mean of groups A and B; n_A, n_B are the sample size of groups A and B; s_{M_A}, s_{M_B} are the standard error of mean of groups A and B. The null hypothesis for the independent t-test is that the population means from the two unrelated groups are equal; while the alternative hypothesis is that the population means are not equal.

2.4.4 One-way Analysis of Variance and One-way Analysis of Covariance

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of three or more independent unrelated groups. The one-way ANOVA compares the means between the groups you are interested in and determines whether any of those means are statistically significantly

different from each other. Specifically, it tests the null hypothesis: $\mu_1 = \mu_2 = \cdots = \mu_k$, where μ is group mean and k is number of groups. F-value is calculated in this statistic analysis:

$$F = \frac{variance\ between\ groups}{variance\ due\ to\ sampling\ error} = \frac{MS_{between-groups}}{MS_{within-groups}} \tag{2.6}$$

$$MS_{between-groups} = \frac{SS_{between-groups}}{df_{between-groups}} \tag{2.7}$$

$$MS_{within-groups} = \frac{SS_{within-groups}}{df_{within-groups}} \tag{2.8}$$

where SS stands for sum of squares; df is degree of freedom. If the one-way ANOVA returns a statistically significant result, we accept the alternative hypothesis, which is that there are at least two group means that are statistically significantly different from each other. At this point, it is important to realize that the one-way ANOVA is an omnibus test statistic and cannot tell you which specific groups were statistically significantly different from each other, only that at least two groups were. To determine which specific groups differed from each other, you need to use a post hoc test, which attempts to control the experiment-wise error rate in the same manner that the one-way ANOVA is used instead of multiple t-tests.

Similar with one-way ANOVA, one-way ANCOVA can be thought of as an extension of the one-way ANCOVA to incorporate a covariate. Like the one-way ANOVA,

the one-way ANCOVA is used to determine whether there are any significant differences between two or more independent unrelated groups on a dependent variable. However, whereas the ANOVA looks for differences in the group means, the ANCOVA looks for differences in adjusted means (i.e., adjusted for the covariate). As such, compared to the one-way ANOVA, the one-way ANCOVA has the additional benefit of allowing you to statistically control for a third variable (sometimes known as a confounding variable), which you believe will affect your results. This third variable that could be confounding your results is called the covariate and you include it in one-way ANCOVA analysis.

2.4.5 Pearson's Correlation Analysis

The Pearson's correlation coefficient is a measure of the linear correlation between two variables with a value between -1 and 1, where -1 indicates a total negative linear correlation, 0 is no linear correlation, and 1 is stands for a total positive linear correlation. Given a pair of random variables X and Y, the Pearson's correlation coefficient is defined as

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{2.9}$$

where μ_X, μ_Y are the mean of X and Y; σ_X, σ_Y are the standard deviation of X and Y.

CHAPTER 3

TOPOLOGICAL FEATURES OF CUE-EVOKED ATTENTION PROCESSING NETWORK ASSOCIATED WITH REMISSION AND PERSISTENCE IN ADULTS WIHT CHILDHOOD ATTENTION-DEFICIT/HYPERACTIVITY DISORDER

3.1 Introduction

3.1.1 Background

CSTC loops that support attention and cognitive processing have been implicated in both the etiology (Casey et al., 2007) and developmental remission of ADHD (Halperin and Schulz, 2006). Abnormal functional activation and connectivity in thalamic, striatal, and prefrontal areas have been reported in children and adults with ADHD, suggesting that they reflect a core pathophysiology that remains static over the lifespan (Cortese et al., 2012; Hart et al., 2013). Abnormalities in prefrontal activation, and connectivity associated with ADHD have been found to vary as a function of adult outcome (Shaw et al., 2006; Li et al., 2007; Shaw et al., 2013; Mattfeld et al., 2014; Francx et al., 2015; Shaw et al., 2015; Schulz et al., 2017). Persistence of ADHD into adulthood has been linked to reduced prefrontal activation and connectivity within the caudal and subcortical regions (Mattfeld et al., 2014; Francx et al., 2015; Schulz et al., 2017).

Our existing studies in young adults with childhood ADHD suggested a double dissociation of functional anomalies linked to childhood onset and adult outcome. Childhood ADHD was associated with reduced cue-related thalamic activation and

connectivity with the brainstem regardless of adult outcome, while symptom remission was linked to enhanced cue-related thalamo-prefrontal connectivity (Clerkin et al., 2013). However, these basic measures of regional activation and connectivity have limited capacity to characterize the functional human brain as a high performance parallel information processing system (Figure 3.1). The field lacks systems-level investigations of the functional patterns that significantly contribute to symptom remission and persistence in adults with childhood ADHD. To fill this gap, graph theoretic techniques (GTT) was conducted to measure the global-, local- and nodal-efficiency, and network hubs for each subject to evaluate the performance of cue-evoked attention network. The objective of this aim is to determine the topological features of functional brain network for cue-evoked attention processing, which associate with the diverse adult outcomes of childhood ADHD.

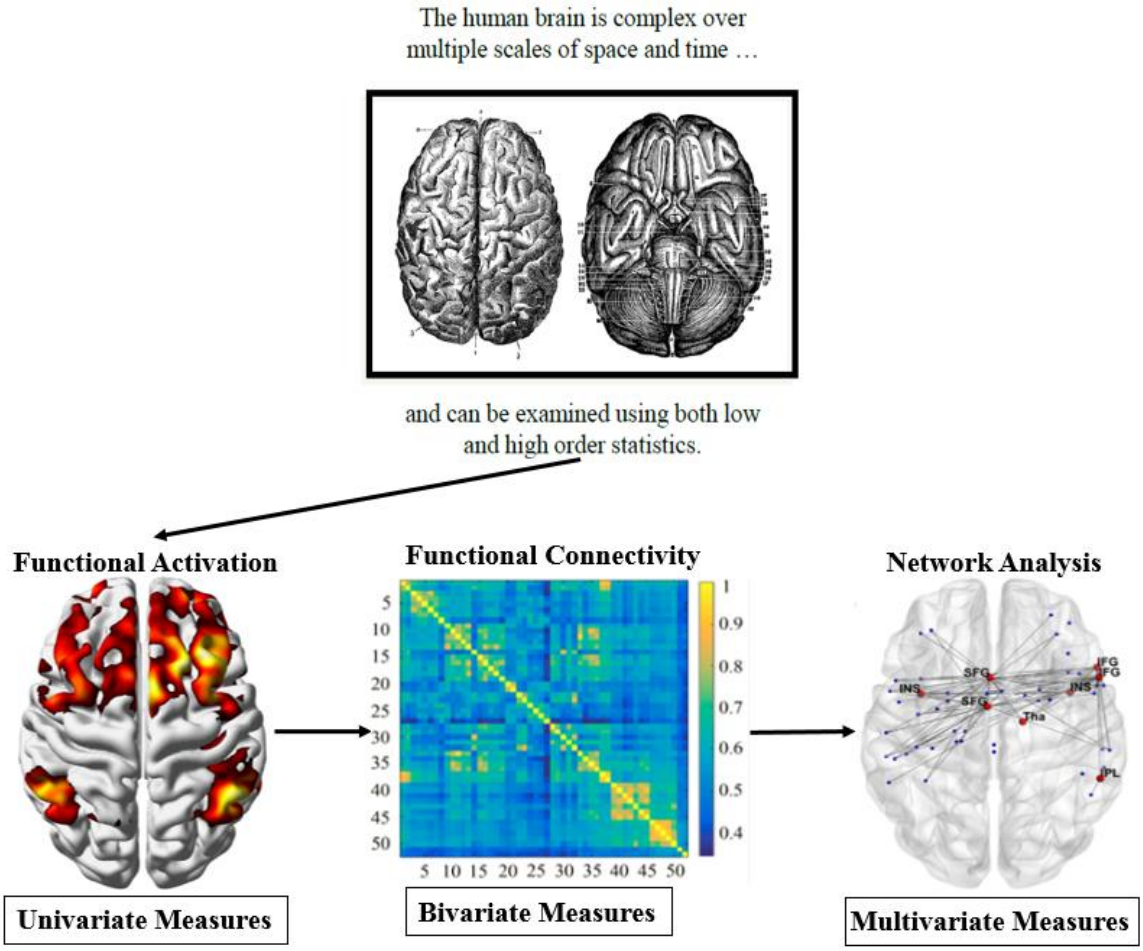

Figure 3.1 General information of univariate, bivariate and network measures.

3.1.2 Graph Theoretic Technique

The human brain comprises about 86 billion neurons connected through approximate 150 trillion synapses that allow neurons to transmit electrical or chemical signals to other neurons (Pakkenberg et al., 2003; Azevedo et al., 2009), which serve the human brain as a complex system to maintain the high efficient information transferring underlying cognition, behavior, and perception (Craddock et al., 2013; Park and Friston, 2013;

Farahani et al., 2019). Graph-based network analysis is a method that could be utilized to reveal the topological properties of human brain functional and structural networks, such as small-worldness, modular organization, high global, local and nodal efficiencies and highly connected or centralized hubs (Bullmore and Sporns, 2012; van den Heuvel and Sporns, 2013).

The first application of graph theory and network analysis can be traced back to 1741 when Leonhard Euler solved the Konigsberg Bridge Problem (Euler, 1741). A graph consists of a finite set of nodes and edges, where the edges represent the connections between the nodes. A human brain network can be classified as one of the four different networks, including binary undirected network, weighted undirected network, binary directed network and weighted directed network (Figure 3.2). Such classification is based on whether the edges between nodes carry directional information (e.g., causal interaction) or not, and whether the edges between nodes are categorized as weighted or binary. Most exiting task-based functional neuroimaging studies have been devoted to the binary undirected networks because of the technical constraints surrounding the inference of directional networks (Liao et al., 2017); while the white matter anatomical network taken by DTI usually utilizes various fiber information, such as fiber number, fiber length, and fractional anisotropy to generate a weighted network (Fornito et al., 2013; Zhong et al., 2015).

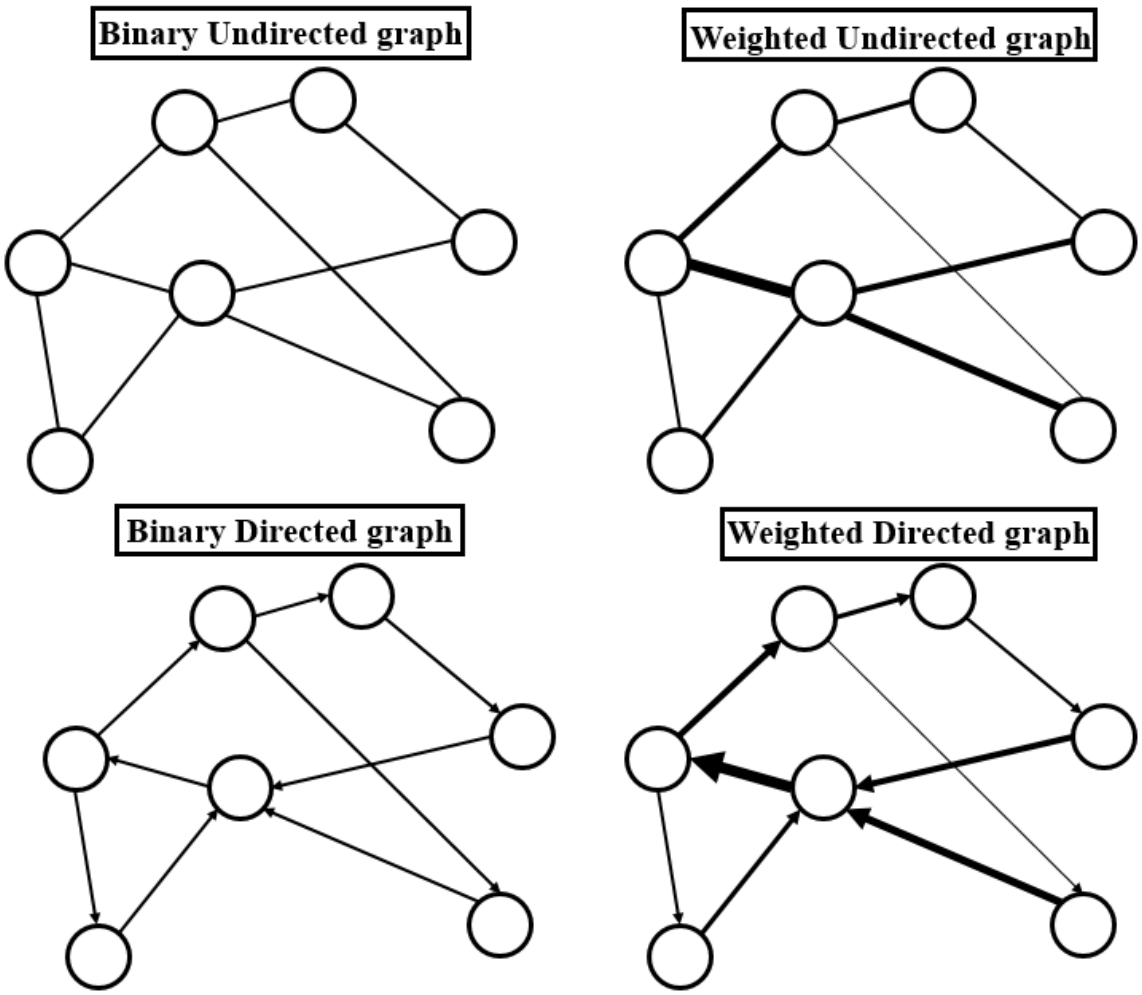

Figure 3.2 Diagrams of four different types of mathematical graphs.

In 1998, the characteristic of small-world was observed by Watts and Strogatz in many social, biological, and geoscience-based networks (Watts and Strogatz, 1998). A small-world network is a type of mathematical graph in which most nodes are likely to be neighbors of each other and most nodes can be reached from every other node by a small number of hops or steps. For the application in human brain, a small-world network has the ability for specialized processing to occur within densely interconnected groups of brain

regions (highly segregated), and also has the ability to combine specialized information from distributed brain regions (highly integrated) (Rubinov and Sporns, 2010). It represents the shortest path between each pair of nodes in the network using the minimum number of edges with a large clustering coefficient and small average path length. These two metrics are the result of a natural process to satisfy the balance between minimizing the resource cost and maximizing the flow of information among the network components (Bassett and Bullmore, 2006; Meunier et al., 2010; Bullmore and Sporns, 2012; Chen et al., 2013; Samu et al., 2014). The wiring costs in connections among anatomically adjacent brain areas are lower than those among distant brain regions (Bullmore and Sporns, 2012). Theoretical examinations have pointed out that the brain regions are more likely to interact with their neighboring areas to reduce the whole metabolic costs, while at the same time they need to have a small number of long-distance connections among themselves to accelerate data transmission (Bullmore and Sporns, 2012; Vertes et al., 2012; Chen et al., 2013). In agreement with theoretical studies, empirical investigations have also proved the dispersion of a few long connections among a plethora of short connections in the human brain network (Salvador et al., 2005; Hagmann et al., 2007; He et al., 2007). Thus, the small-world property of human brain is important for the synchronization of cortical regions to maintain the robustness to perturbations (Yu et al., 2008).

The main capability of graph theory in neuroscience studies is usually unveiled after the construction of a functional brain network. Several measures can be used to assess

the topological patterns of different networks such as global, local, nodal efficiency and network hub measurements, including degree and betweenness centrality, which have been described in detail (Sporns et al., 2004; van den Heuvel et al., 2008). Typically, it is difficult to claim which measures are more suitable for studying the brain network (Bullmore and Sporns, 2009), but given the complex structure of the human brain, measures that can represent the small-world properties of the brain network are of great importance (He and Evans, 2010; Liao et al., 2017). This critical property arises with the help of hubs (i.e., highly connected nodes in a network), causing the creation of local clusters (Bullmore and Sporns, 2009). Efficiency is another biologically relevant metric to describe brain networks from the perspective of information flow, which can deal with the disconnected graphs, nonsparse graphs or both (Latora and Marchiori, 2001; Bassett and Bullmore, 2006). Global efficiency and local efficiency measure the ability of a network to transmit information at the global and local level, respectively (Latora and Marchiori, 2001).

3.1.3 Current Applications of Graph Theoretic Technique in ADHD Neuroimaging Studies

From multiple lines of evidence, it is known that the human brain exhibits both locally specialized processing units within relatively circumscribed regions or clusters (Bartels and Zeki, 2000), as well as distributed interregional interactions [e.g., the front-parietal attention network (Dosenbach et al., 2008). The functional brain networks are small-world

and economical in the sense of providing high global and local efficiencies of parallel information processing with low rewiring cost (Achard and Bullmore, 2007). The development of GTT provided powerful methods to characterize the nodal (regional) and global topological properties of such functional brain network, which has been implemented in ADHD studies. Multiple resting-state fMRI studies have reported abnormal global and regional topological properties in children with ADHD when compared to typically developing children (TDC), such as lower global efficiency, reduced nodal efficiency in frontal areas (Fair et al., 2009; Supekar et al., 2009; Wang et al., 2009; Cao et al., 2014b); delayed maturation of the default mode network (DMN) (Fair et al., 2009; Fair et al., 2010); and different connectivity patterns among nodes of the entire resting-state functional brain network in children with ADHD-inattentive, -hyperactive/impulsive, and -combined subtypes (Fair et al., 2012). Resting-state GTT studies in adults with ADHD reported increased coherence in the DMN when comparing to children with ADHD (Fair et al., 2008); no significant alterations in global network properties, but significantly lower nodal path length of right medial frontal and right superior occipital cortices, and significantly higher nodal clustering coefficient in left orbitofrontal and right superior temporal cortices, relative to group-matched control adults (Cocchi et al., 2012). A few task based GTT studies have also been conducted in children with ADHD. A multi-source interference task-based electroencephalography (EEG) study showed increased local characteristics combined with decreased global characteristics in

children with ADHD compared to the group-matched TDC (Liu et al., 2015). A visual sustained attention task-based fMRI study observed significantly reduced nodal efficiency in frontal and occipital regions in children with ADHD, and a hyperactive network hub in anterior cingulate cortex (Xia et al., 2014). These existing GTT studies in resting-state and task-based functional neuroimaging data demonstrated altered global and regional topological patterns of the functional brain networks in children and adults with ADHD (Cao et al., 2014a). However, systems-level functional brain characteristics associated with the diverse adult outcomes of childhood ADHD have not yet been investigated. Aim 1 utilized GTT to examine the systems-level neurophysiological mechanisms associated with symptom persistence and remission of childhood ADHD in young adults diagnosed with the disorder during childhood and matched comparison subjects (Luo et al., 2018). Within the 69 participants of this study, 67 overlapped with those in a previous study which reported that greater thalamo-frontal functional connectivity during response preparation stage of attention processing significantly linked to ADHD symptom remission (Clerkin et al., 2013). Both the current and the previous studies used fMRI data acquired during the same cued attention task, but focused on different components of attention that were represented by different contrasts of the task. The current study focused on the cue-evoked attention stage defined by the cue vs. baseline contrast, while the previous study focused on the response preparation stage defined by the cue minus non-cue contrast. During cued attention processing, the cue-evoked attention component occurs prior to the response

preparation component. Existing clinical studies have found a delayed pattern of attention development in children with ADHD, especially on the cue-evoked attention component (Suades-Gonzalez et al., 2017). We hypothesized that optimal topological organization of the functional brain network for cue-evoked attention processing may be associated with symptom remission in adults with childhood ADHD.

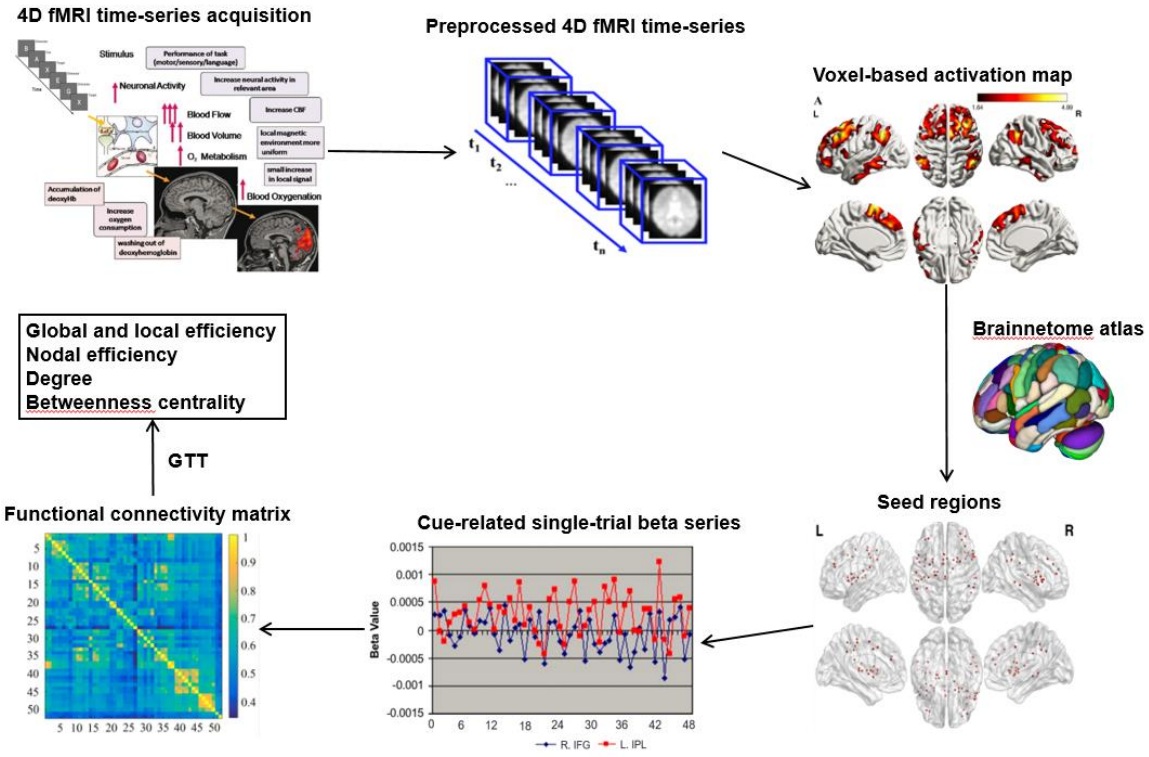

Figure 3.3 General workflow of individual-level analysis.

3.2 Experimental Strategy

3.2.1 Participants

Neuroimaging and clinical data from 69 young adults were included in analyses of this project, including 36 probands and 33 group-matched NCs. Of these 36 adults with childhood ADHD, 17 were classified as ADHD-P and the other 19 were classified as ADHD-R. In these 17 ADHD-P, 14 were diagnosed with ADHD-C, 1 met criteria for ADHD-inattentive subtype, and 2 met criteria for ADHD-hyperactive/impulsive subtype, based on DSM-IV criteria for adult ADHD.

3.2.2 Experimental Task

During fMRI data acquisition in Project 1, each participant performed a cued attention task (CAT), which have an ability to evoke alerting state of readiness that suppresses ongoing activity and lowers motor thresholds to prepare for a rapid response. The CAT was developed and described in detail in (Clerkin et al., 2009; Clerkin et al., 2013; Luo et al., 2018). The event-related task consisted of four 300 seconds runs that each began and ended with a 30-second fixation cross. As shown in Figure 3.4, each run consisted of a series of 120 letters containing 24 targets (“X”), half preceded by a cue (“A”), and the other half preceded by a non-cue letter (“B” through “H”). Participants were told that the letter “A” (the cue) was always followed by the target letter (“X”); but not all targets (i.e., “X”) were preceded by a cue letter (“A”). The stimuli were presented for 200 ms with a

pseudorandom inter-stimulus interval which ranged from 1550 to 2050 ms (mean=1800 ms/run). Participants were instructed to respond to each target as rapidly as possible using their right index finger. Before entering the scanner, detailed instructions and practice trials of the task were provided to each participant to ensure satisfactory performance.

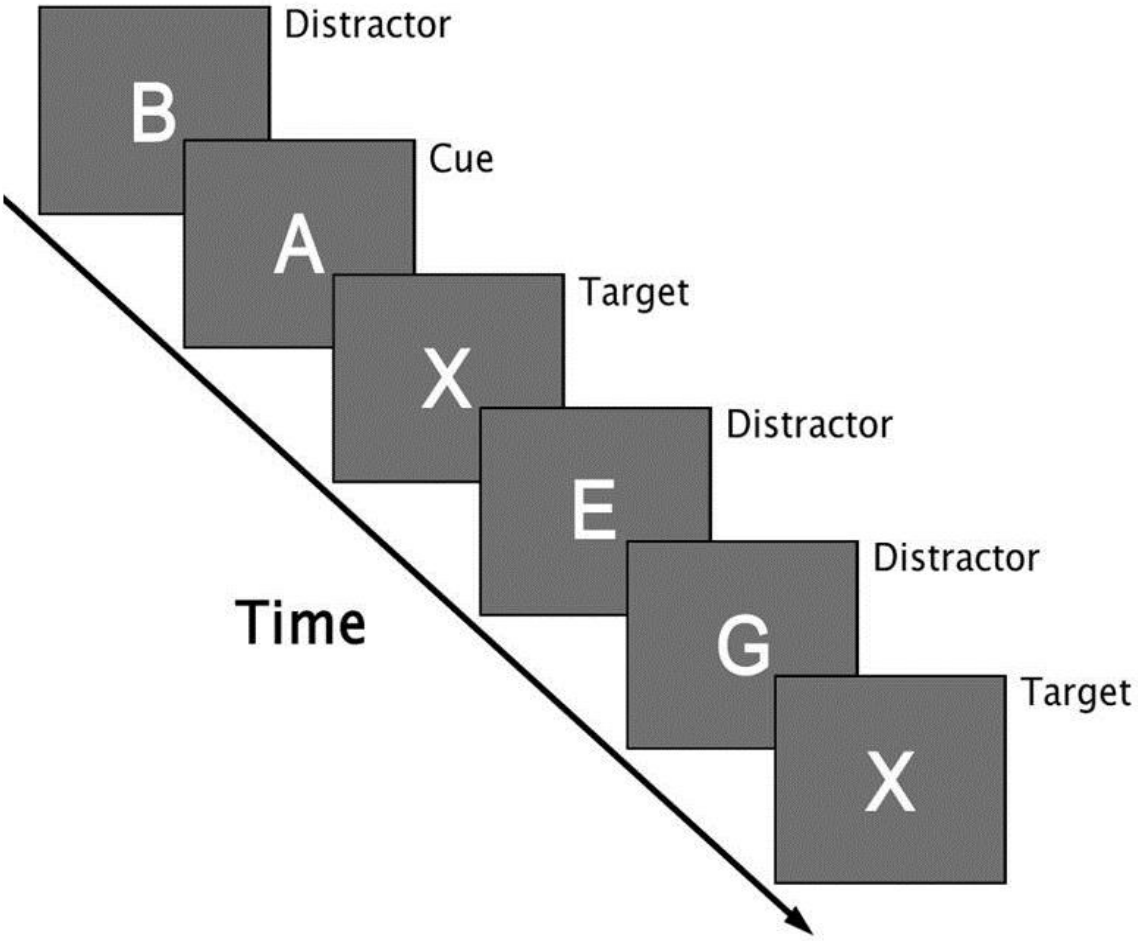

Figure 3.4 Schematic representation of the cued attention task.

3.2.3 Individual-level Functional Magnetic Resonance Imaging Data Preprocessing and Seed Regions Detection

The fMRI data from each subject was preprocessed and analyzed using SPM version 8 (SPM8, Wellcome Trust Centre for Neuroimaging, London, United Kingdom; http://www.fil.ion.ucl.ac.uk/spm/) implemented on a MATLAB platform. The fMRI raw data was slice timing corrected, realigned to the first volume, coregistered to the T2 structural image, segmented, normalized to the Montreal Neurological Institute (MNI) template with a voxel size of $2\times2\times2$ mm^3, and spatially smoothed with an 8-mm full-width at half maximum (FWHM) Gaussian kernel. Additional frame-wise head motion analyses were conducted by calculating the mean absolute displacement of each brain volume compared with the previous volume from the translation parameters (Power et al., 2012). No participant was excluded for excessive head motion detected by frame-wise analyses (Mean Motion<0.2 mm in all data). The six basic motion parameters created during realignment and three frame-based motion parameters were regressed out from each fMRI data. The residual time series were analyzed to generate the cues ("A"), targets ("X"), and non-cues (letters other than "A" and "X")-related activation maps by FMRIB improved linear regression model (FILM), which uses a nonparametric estimation of time series autocorrelation to prewritten each voxel's time series (Smith et al., 2004). The average activation map responding to the cues was generated for each group. Multiple comparisons were controlled in both individual-level and group-level analyses by applying family-wise

error rate (FWER), i.e., Gaussian random-field theory at a cluster corrected significance threshold of $p<0.05$.

As shown in Table 3.1, a total of 52 cortical and subcortical seed regions (nodes of the functional brain network for cue-evoked attention processing) were determined based on the combination of the activation maps of the groups of ADHD probands and controls. The combined activation map was parceled according to the structural and functional connectivity-based Brainnetome atlas (Fan et al., 2016). The seed regions were spheres (radius=4 mm) identified from the coordinates of the local activation peaks containing at least 50 contiguous voxels surrounding the peak voxel (See from Figure 3.5 and 3.6, and Table 3.1 for details of seed region determination). The size of the seed regions was determined based on the estimation of average cortical thickness of adult human brain (Power et al., 2011; Wig et al., 2014).

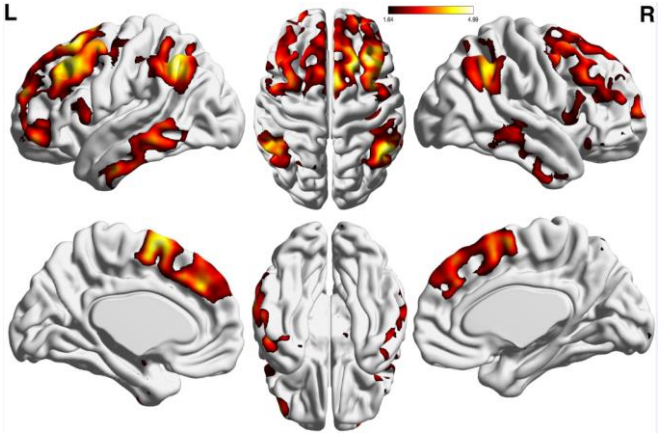

Figure 3.5 Diagram of combination of the cue-evoked attention processing-related activation maps from the groups of ADHD probands and controls.

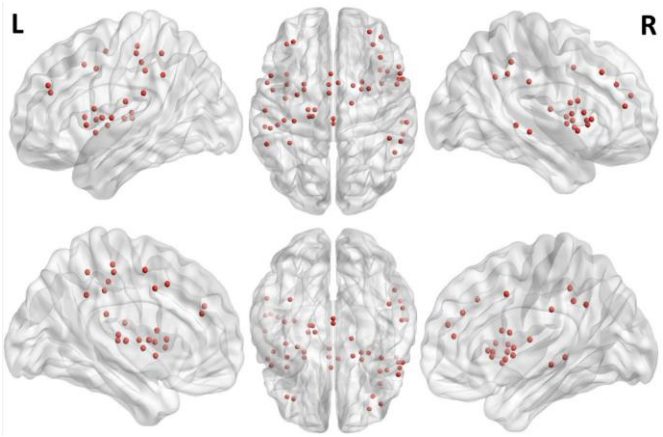

Figure 3.6 Locations of the seeds determined from the combination of the activation maps in controls and ADHD probands.

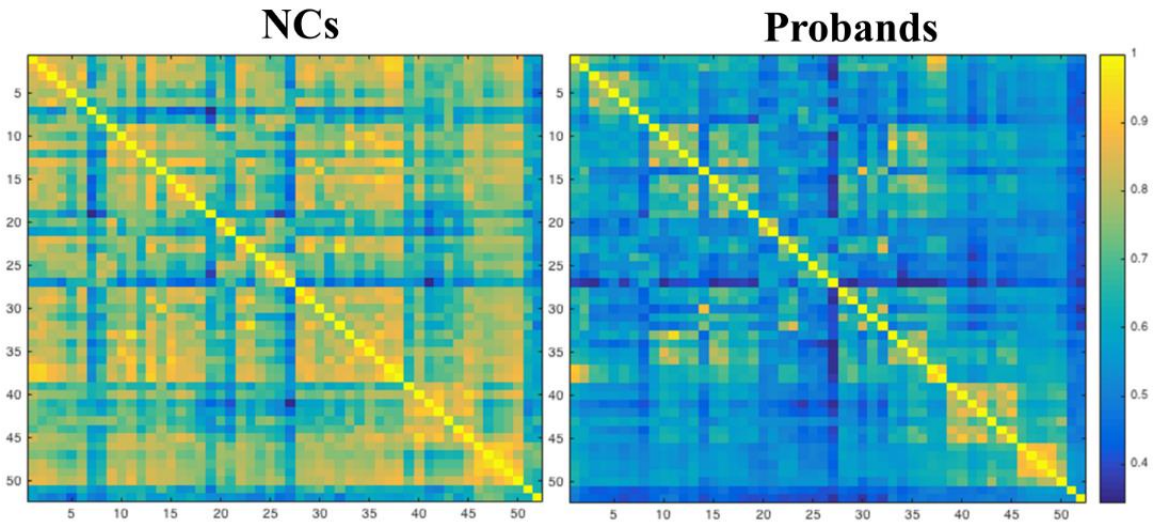

Figure 3.7 The 52×52 functional connectivity matrices in both controls and probands.

Table 3.1 Definition of Selected 52 Nodes **(Continued)**

Node ID	Anatomical Regions	BA	Abbreviation	COG MNI coordinates [x y z]		
1	L. Superior frontal gyrus medial	8	SFG_L_7_1	-2	14	43
2	L. Superior frontal gyrus medial	6	SFG_L_7_5	-3	-3	52
3	L. Middle frontal gyrus, dorsal	9/46	MFG_L_7_1	-32	40	29
4	R. Middle frontal gyrus, dorsal	9/46	MFG_R_7_1	32	38	28
5	R. Middle frontal gyrus	46	MFG_R_7_3	29	49	21
6	L. Middle frontal gyrus, ventral	9/46	MFG_L_7_4	-38	39	24
7	R. Middle frontal gyrus, ventral	9/46	MFG_R_7_4	39	44	13
8	R. Middle frontal gyrus, ventrolateral	8	MFG_R_7_5	38	27	37
9	R. Inferior frontal gyrus, caudal	45	IFG_R_6_3	53	20	4
10	L. Inferior frontal gyrus, opercular	44	IFG_L_6_5	-38	13	7
11	R. Inferior frontal gyrus, opercular	44	IFG_R_6_5	44	16	3
12	L. Inferior frontal gyrus, ventral	44	IFG_L_6_6	-50	12	2
13	R. Inferior frontal gyrus, ventral	44	IFG_R_6_6	54	14	9
14	L. Precentral gyrus (upper limb region)	4	PrG_L_6_3	-32	-26	56
15	L. Precentral gyrus (tongue and larynx region)	4	PrG_L_6_5	-49	0	7
16	R. Precentral gyrus (tongue and larynx region)	4	PrG_R_6_5	52	5	8
17	L. Precentral gyrus, caudal ventrolateral	6	PrG_L_6_6	-54	6	13
18	R. Precentral gyrus, caudal ventrolateral	6	PrG_R_6_6	56	9	17
19	R. Superior temporal gyrus	41	STG_R_6_3	53	9	-4
20	R. Middle temporal gyrus, anterior superior temporal sulcus		MTG_R_4_4	56	-27	-5
21	R. Posterior superior temporal sulcus, rostral		pSTS_R_2_1	53	-37	0
22	L. Superior parietal lobule, postcentral	7	SPL_L_5_3	-35	-44	50
23	L. Inferior parietal lobule, rostrodorsal	40	IPL_L_6_3	-51	-32	39
24	R. Inferior parietal lobule, rostrodorsal	40	IPL_R_6_3	46	-40	44
25	L. Inferior parietal lobule, caudal	40	IPL_L_6_4	-54	-45	36
26	R. Inferior parietal lobule, caudal	40	IPL_R_6_4	55	-43	36
27	R. Inferior parietal lobule, rostroventral	39	IPL_R_6_5	50	-52	34
28	L. Inferior parietal lobule, rostroventral	40	IPL_L_6_6	-55	-31	24
29	R. Inferior parietal lobule, rostroventral	40	IPL_R_6_6	60	-27	30
30	L. Postcentral gyrus (upper limb, head and face region)	1/2/3	PoG_L_4_1	-40	-25	51

Table 3.1 (Continued) Definition of Selected 52 Nodes

Node ID	Anatomical Regions	BA	Abbreviation	COG MNI coordinates [x y z]		
31	L. Postcentral gyrus (tongue and larynx region)	1/2/3	PoG_L_4_2	-55	-17	18
32	L. Postcentral gyrus	2	PoG_L_4_3	-45	-29	45
33	R. Insular gyrus, dorsal agranular insular		INS_R_6_3	38	17	0
34	L. Insular gyrus, dorsal granular insular		INS_L_6_5	-39	-7	7
35	L. Insular gyrus, dorsal dysgranular insular		INS_L_6_6	-38	5	6
36	R. Insular gyrus, dorsal dysgranular insular		INS_R_6_6	39	5	3
37	L. Cingulate gyrus, caudodorsal	24	CG_L_7_5	-4	5	40
38	R. Cingulate gyrus, caudodorsal	24	CG_R_7_5	4	6	40
39	L. Basal ganglia, globus pallidus		BG_L_6_2	-24	0	7
40	R. Basal ganglia, globus pallidus		BG_R_6_2	23	1	6
41	L. Basal ganglia, ventromedial putamen		BG_L_6_4	-24	5	-3
42	R. Basal ganglia, ventromedial putamen		BG_R_6_4	23	7	-2
43	R. Basal ganglia, dorsal caudate		BG_R_6_5	16	4	16
44	L. Basal ganglia, dorsolateral putamen		BG_L_6_6	-29	-4	2
45	R. Basal ganglia, dorsolateral putamen		BG_R_6_6	29	1	1
46	L. Thalamus, pre-motor thalamus		Tha_L_8_2	-20	-16	7
47	L. Thalamus, sensory thalamus		Tha_L_8_3	-19	-22	7
48	L. Thalamus, posterior parietal thalamus		Tha_L_8_5	-17	-22	10
49	L. Thalamus, lateral Pre-frontal thalamus		Tha_L_8_8	-15	-17	6
50	R. Thalamus, lateral Pre-frontal thalamus		Tha_R_8_8	15	-11	11
51	Brainstem, pons		BS_P	0	-28	-34
52	Brainstem, midbrain		BS_MB	0	-24	-12

3.2.4 Network Analyses

To construct the cue-evoked attention processing network, the single-trial beta value maps from the total of 48 cue-related events in the preprocessed data from each subject was first extracted (technical details see from (Rissman et al., 2004)). These 48 maps were sequentially combined to form the whole brain single-trial beta value series. The average beta value series in each of the 52 seed regions was then calculated. Pearson correlation of the average beta value series in each pair of the seed regions was calculated. The functional connectivity matrix was constructed using the absolute values of the correlation coefficients, and was converted into a binary graph, by using the network cost as threshold. The network cost was defined as:

$$C_G = \frac{K}{N(N-1)/2} \tag{3.1}$$

Where N and K were the total number of nodes and edges respectively; $N(N-1)/2$ was the number of all possible subnetworks in the graph G (Latora and Marchiori, 2001).

Resting-state fMRI studies often use only positive or only negative values in the correlation coefficient matrices to construct brain networks, with concerns that positive and negative functional connectivities in brain regions may have different physiological substrates during resting-state (Shehzad et al., 2009; Di et al., 2014). The current study

chose to use the absolute values of the functional correlation coefficients, which is consistent with the majority of existing task-based functional brain network studies, with an understanding that strong functional connections of brain regions in both positive and negative ways represent strong regional interactions for sensory and cognitive information transferring during the task (Meunier et al., 2014). Again, consistent with existing task-based functional network studies, we chose to use the binarized, instead of weighted connectivity matrix that were often implemented in resting-state studies (Tomasi et al., 2014; Iraji et al., 2016).

We investigated the network topological properties over a wide range of the cost values from 0.1 to 0.5 (with increments of 0.01). This selected cost value interval was widely suggested to allow the small-world properties to be properly estimated and the subnetworks to be connected with enough discriminatory power in functional connectivity (Watts and Strogatz, 1998; Bullmore and Sporns, 2009; Xia et al., 2014). Global-efficiency, $E_{glob}(G)$ and local-efficiency, $E_{loc}(G)$, were defined using the following:

$$E_{glob}(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{l_{ij}} \tag{3.2}$$

$$E_{loc}(G) = \frac{1}{N} \sum_{i \in G} E_{glob}(G_i) \tag{3.3}$$

where l_{ij} was the shortest path length between node i and j; $E_{glob}(G_i)$ the global efficiency of the sub-network G_i that was constructed by the set of nodes that were immediate neighbors of node i (Latora and Marchiori, 2001). The graph is considered a small-world network if it met the criteria: $E_{glob}(G_{regular})<E_{glob}(G)<E_{glob}(G_{random})$ and $E_{loc}(G_{random}) < E_{loc}(G) < E_{loc}(G_{regular})$, where $E_{glob}(G_{regular})$, $E_{glob}(G_{random})$, $E_{loc}(G_{regular})$ and $E_{loc}(G_{random})$ represent the global- and local-efficiency of the node- and edge-matched regular and random networks, respectively (Achard and Bullmore, 2007) (Figure 3.8). In this study, the network cost range from 0.15 to 0.3 was determined based on these criteria (Figure 3.9).

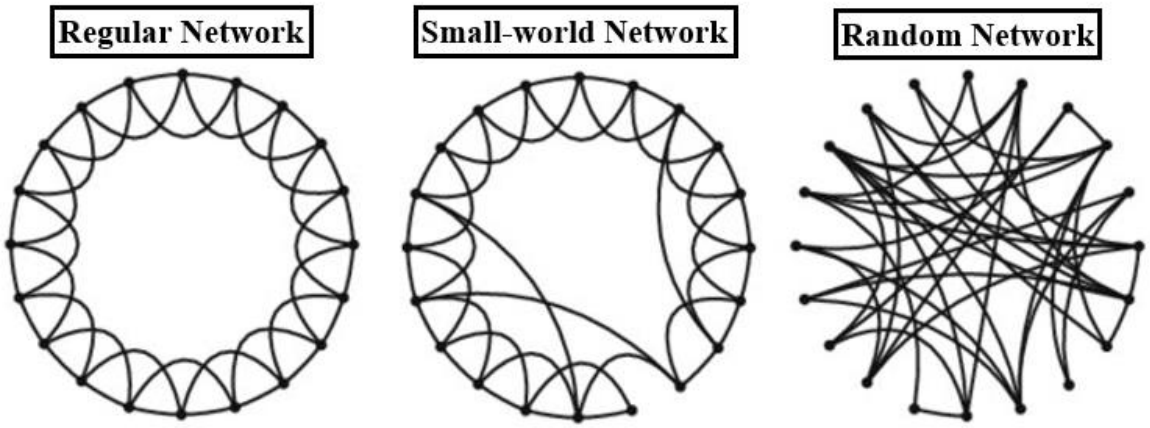

Figure 3.8 Diagrams of regular, small-world and random networks.

Nodal-efficiency, $E_{nodal}(G,i)$, was defined as following:

$$E_{nodal}(G,i) = \frac{1}{N-1}\sum_{j\in G}\frac{1}{l_{ij}} \tag{3.4}$$

where l_{ij} was the shortest path length between node i and j. It was a local measurement to evaluate the communication efficiency between a node i and all other nodes in the network G (Latora and Marchiori, 2001).

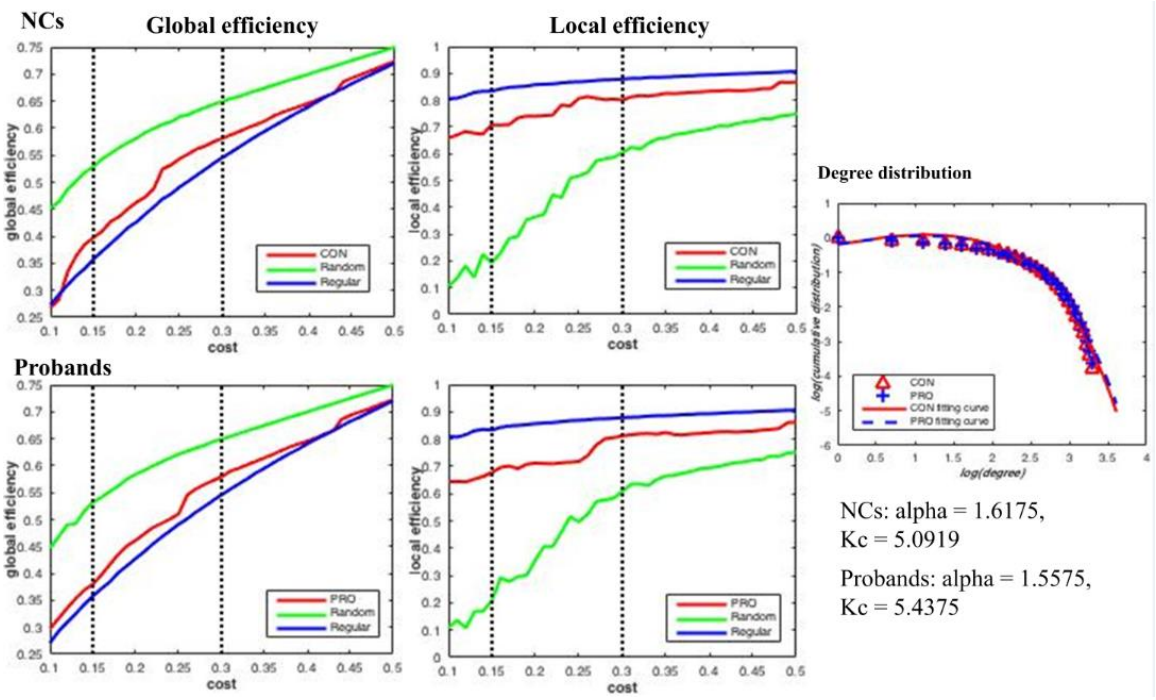

Figure 3.9 The small-world properties of the functional brain networks. The global and local efficiency curves of both controls and probands showed small world pattern over a range of 0.15≤cost≤0.3. The Degree distribution curves fitted by an exponentially truncated power of the form, $P(k) \sim k^{\alpha-1}e^{-k/k_c}$, showed an estimated exponent α=1.618 and a cutoff degree k_c=5.092 for the control group and α=1.558, k_c=5.438 for the probands group.

Network hubs in each diagnostic group were also investigated. Degree (D) and betweenness-centrality (BC) were utilized to determine whether a node acts as a network hub. The definition of degree (D_i) of node i was the number of edges connected to that

node, while the betweenness-centrality (BC_i) of node i was defined by the number of all the shortest paths between two nodes (other than node i) that pass through node i (Sporns et al., 2007). In each individual, the average values of the D and BC measures of each node over the network cost range of 0.15-0.3 were calculated and then normalized by converting into z scores using a normal distribution. The network hubs in each group were identified by estimating the grand sample means of D and BC in each diagnostic group. These standardized values were then tested with a normal distribution. A node i was defined as an acting network hub if 1- $\Phi(z_i)<\alpha$, where $\Phi(\cdot)$ was the standard normal cumulative distribution function, and α=0.05 was the level of significance (Li et al., 2012b).

3.2.5 Group Statistic Analyses

Group comparisons of clinical, behavioral and demographical characteristics and fMRI task performance measures were carried out using chi-square tests for discrete variables, and/or unpaired two-sample t-tests for continuous variables between the controls and ADHD probands, and further between the two ADHD subgroups (i.e., ADHD-R and ADHD-P).

Group comparisons of the functional brain network topological measures, including efficiency measures $E_{glob}(G)$, $E_{loc}(G)$ and $E_{nodal}(G,i)$ at each i, and network hub measures D_i and BC_i at each i, were carried out using the one-way analysis of covariance (ANCOVA) and post-hoc t-tests among the three groups; controls, ADHD-R

and ADHD-P, with gender as a fixed-effect covariate and age as a random-effect covariate. Bonferroni correction was used to control multiple comparisons at α=0.05.

Pearson's correlation analyses were applied in the ADHD-R and ADHD-P groups, respectively, to determine associations between the ADHD symptoms (measured using the T-scores of the DSM-IV Inattentive and Hyperactive-Impulsive Indices and the fMRI task performance measures) and the brain network measures that showed significant alterations in either ADHD-R or ADHD-P relative to controls. Again, bonferroni correction was applied to control multiple comparisons at α=0.05.

3.3 Results

3.3.1 Clinical, Behavioral and Demographic Measures

As shown in Table 3.2, demographic measures did not show significant between-group differences. All participants achieved a >85% rate for response accuracy when performing the fMRI task. Task performance measures, including response accuracy rate, omission error rate, commission error rate, did not show between-group differences. Compared to the controls, the ADHD probands showed significantly larger standard deviation of the reaction time for the cued targets (p=0.005) and all targets (p=0.006), and significantly longer mean reaction time for the cued targets (p=0.048).

Table 3.2 Demographic and Clinical Characteristics in Groups of Controls and ADHD Probands (and Further in the Sub-groups of Remitters and Persisters of the ADHD Probands)

	Controls (N=33)	**Probands (N=35)**		**ADHD-R (N=18)**	**ADHD-P (N=17)**	
	Mean (SD)	**Mean (SD)**	***p***	**Mean (SD)**	**Mean (SD)**	***p***
Age	24.27 (2.2)	24.69 (2.1)	*0.44*	24.79 (2.2)	24.55 (2.2)	*0.77*
Full-scale IQ	102.81(15.5)	99.11 (14.5)	*0.33*	99.57 (14.9)	98.46 (14.9)	*0.84*
Conners' Adult ADHD Rating Scale (T-score)						
Inattentive	45.19 (8.3)	55.29 (12.6)	*<0.001*	49.94 (10.9)	62.69 (11.1)	*<0.01*
Hyperactive/impulsive	42.73 (6.1)	52.36 (12.3)	*<0.001*	46.06 (9.0)	61.08 (11.0)	*<0.001*
ADHD Total	43.42 (7.7)	55.48 (14.2)	*<0.001*	48.44 (11.2)	65.23 (12.1)	*<0.001*
ADHD semi-structured interview (number of symptoms)	0.81(1.3)	6.26 (4.8)	*<0.001*	2.52 (1.8)	10.69 (3.2)	*<0.01*
	N (%)	**N (%)**	***p***	**N (%)**	**N (%)**	***p***
Male	28 (84.8)	29 (80.6)	*0.64*	16 (84.2)	13 (76.5)	*0.56*
Right-handed	31 (93.9)	32 (88.9)	*0.46*	17 (89.5)	15 (88.2)	*0.91*
Race			*0.21*			*0.42*
Caucasian	12 (36.4)	17 (47.2)		9 (47.4)	8 (47.1)	
African American	12 (36.4)	5 (13.9)		4 (21.1)	1 (5.9)	
More than one race	6 (18.2)	6 (16.7)		4 (21.1)	2 (11.8)	
Asian	2 (6.1)	0		0	0	
Ethnicity			*0.33*			*0.96*
Hispanic/Latino	10 (30.3)	15 (41.7)		8 (42.1)	7 (41.2)	
Current mood disorder	4 (12.1)	6 (16.7)	*0.59*	3 (15.8)	3 (17.6)	*0.88*
Current anxiety disorder	8 (24.2)	10 (27.8)	*0.74*	3 (15.8)	7 (41.2)	*0.09*
Current substance disorder	7 (21.2)	13(36.1)	*0.17*	8 (42.1)	5 (29.4)	*0.43*

3.3.2 Brain Network Topological Measures

The network global- and local-efficiency measures did not significantly differ among the controls, ADHD-R and ADHD-P. Relative to ADHD-P, ADHD-R showed significantly higher nodal-efficiency in both right (p=0.027) and left middle frontal gyrus (MFG) (p=0.029).

Group comparisons of the network hub measures showed that compared to the controls, ADHD-P had significantly lower D in right MFG (p=0.045), and significantly lower BC in left superior frontal gyrus (SFG) (p=0.016), left MFG (p=0.002), right inferior frontal gyrus (IFG) (p=0.004) and left precentral nodes (p=0.003).

Figure 3.10 and Table 3.3 showed the anatomical regions and locations of the acting network hubs (see definition in Section 2.5), which were significantly more active than the average of all the nodes within each of the three diagnostic groups. Distinct patterns of acting network hub distribution were observed among these groups. The group of controls showed acting network hubs in bilateral inferior parietal lobule (IPL) and functional connections between the left IPL and left SFG hubs, that were not found in the groups of ADHD-R and ADHD-P. The right MFG, and right globus pallidus and putamen were shown as acting network hubs only in the ADHD-R, while left MFG and left precentral gyrus were shown as acting network hubs only in the ADHD-P. In addition, a

unique pattern of significant bilateral MFG functional connectivity in the subgroup of ADHD-P (p=0.005) was observed when compared to controls and the ADHD-R.

3.3.3 Associations Between Brain and Behavioral Measures

As shown in Figure 3.11, right IFG, right IPL, and bilateral MFG played different roles in ADHD symptom manifestation in the ADHD-R and ADHD-P. Specifically, higher nodal efficiency of right IFG was significantly associated with lower inattentive (r=-0.592, p=0.01) and hyperactive/impulsive (r=-0.544, p=0.02) symptom severity scores in ADHD-R, but significantly associated with higher inattentive (r=0.614, p=0.026) and hyperactive/impulsive (r=0.62, p=0.024) symptom severity scores in the ADHD-P. Stronger bilateral MFG functional connectivity was significantly associated with increased inattentive symptoms (r=0.484, p=0.042) in the ADHD-R but not in the ADHD-P. Higher nodal efficiencies of right IPL was strongly associated with increased inattentive symptoms in the ADHD-P (r=0.56, p=0.046). In addition, reduced average reaction time for cued targets was significantly associated with higher right IPL nodal efficiency in the group of ADHD-R (r=-0.583, p=0.011).

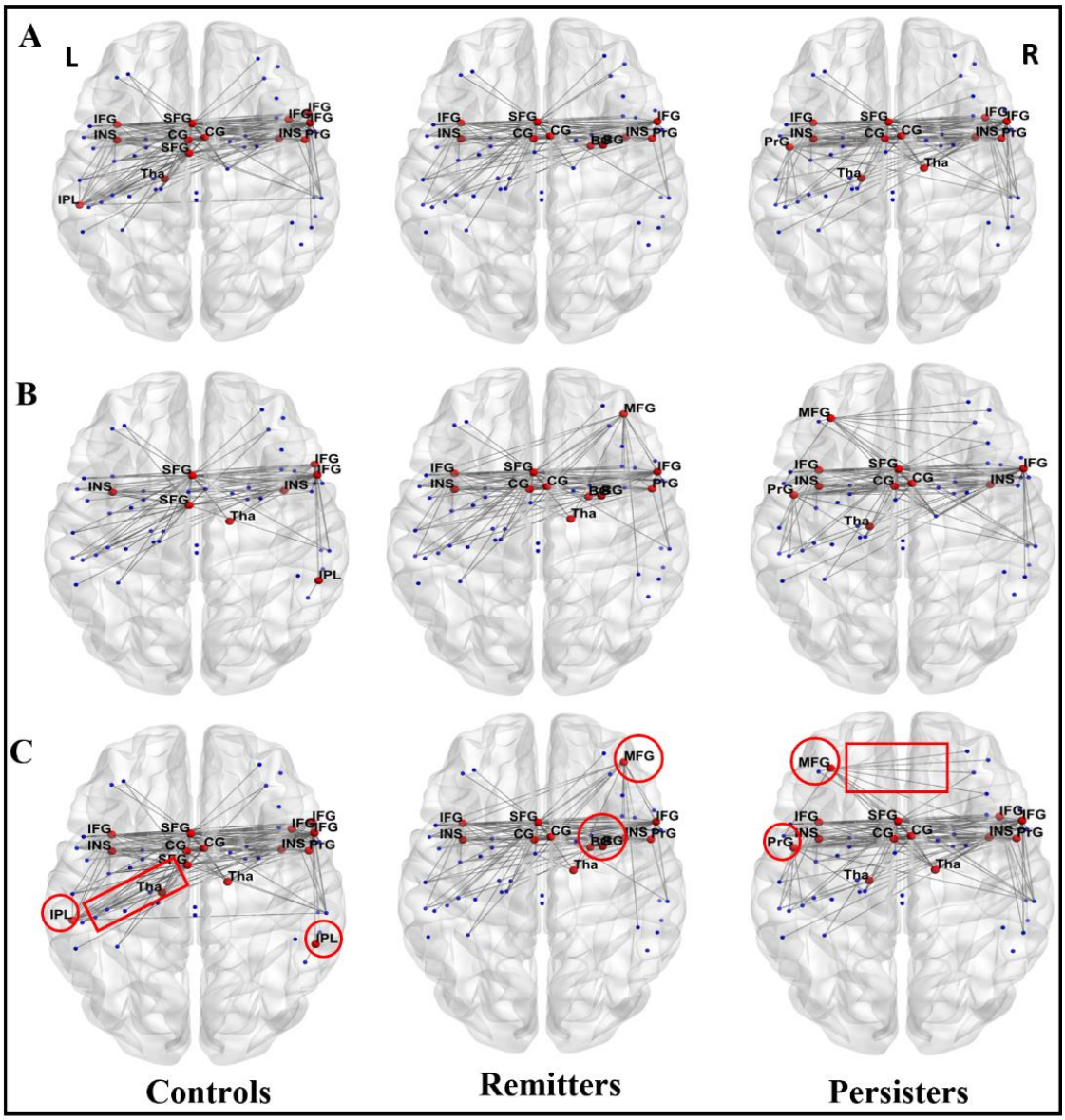

Figure 3.10 Network hubs in the groups of controls, ADHD remitters; and persiters. (A): Hubs identified with only degree measures; (B): Hubs identified with only between-centrality measure; (C): Combination of hubs identified using degree and between-centrality measures.

Table 3.3 Network Hubs in ADHD Persisters, Remitters, and Normal Controls

			Controls		ADHD-R		ADHD-P	
Node ID	**Anatomical Region**	**BA**	**Degree**	**BC**	**Degree**	**BC**	**Degree**	**BC**
1	L. Superior frontal gyrus	8	√	√	√	√	√	√
2	**L. Superior frontal gyrus**	**6**	**√**	**√**				
3	**L. Middle frontal gyrus**	**9/46**						**√**
4	R. Middle frontal gyrus	9/46		√				√
7	**R. Middle frontal gyrus**	**9/46**				**√**		
9	R. Inferior frontal gyrus	45	√	√				
10	L. Inferior frontal gyrus	44	√		√	√	√	√
11	R. Inferior frontal gyrus	44	√				√	
13	R. Inferior frontal gyrus	4	√	√	√	√	√	√
15	**L. Precentral gyrus (tongue and larynx region)**	**4**					**√**	**√**
16	R. Precentral gyrus (tongue and larynx region)	4	√		√	√	√	
26	**R. Inferior parietal lobule**	**40**		**√**				
28	L. Inferior parietal gyrus	40	√					
35	L. Insula		√	√	√	√	√	√
36	R. Insula		√	√	√		√	√
37	L. Cingulate gyrus	24	√		√	√	√	√
38	R. Cingulate gyrus	24	√		√	√	√	√
40	**R. Globus pallidus**				**√**	**√**		
45	**R. Dorsolateral putamen**				**√**	**√**		
49	**L. Thalamus**		**√**				**√**	**√**
50	R. Thalamus			√		√	√	

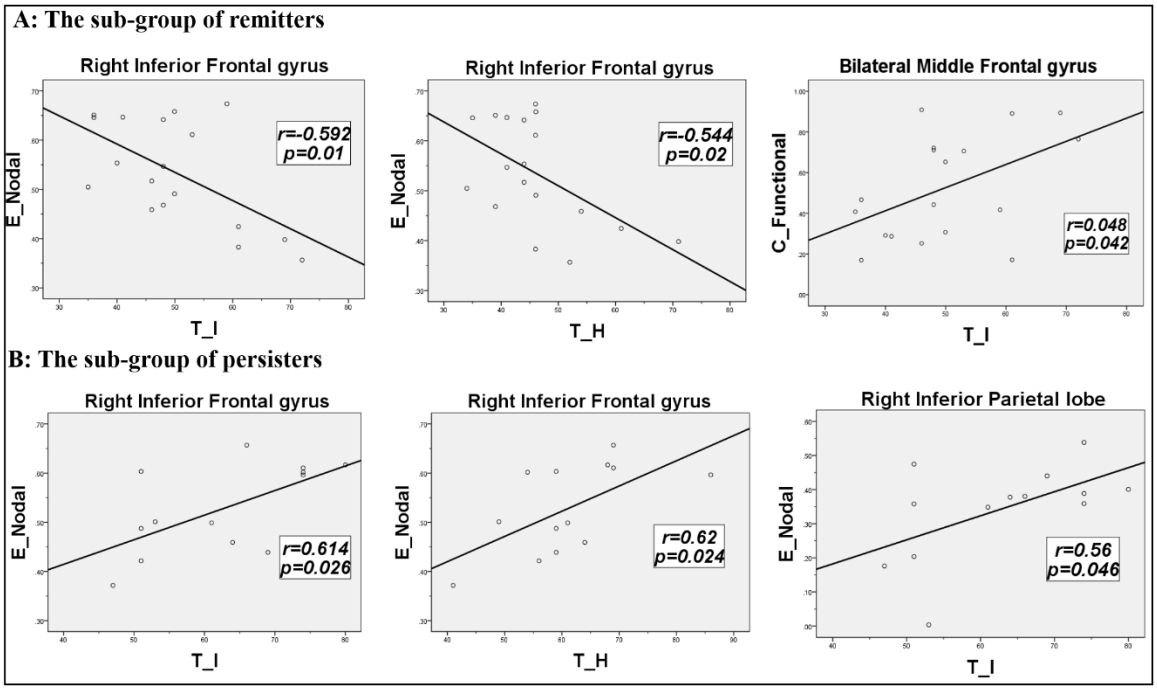

Figure 3.11 Regions that showed significant correlations between their nodal efficiency and the clinical symptom measures, including the T-scores of inattentive and hyperactive/impulsive symptoms, in the remitters and persisters study cohort.

3.4 Discussion

To our knowledge, this is the first study in the field to report distinct topological properties of the cue-evoked attention processing brain network in persisters and remitters who were diagnosed with ADHD during childhood. Specifically, missing network hubs in IPL, and lacking fronto-parietal functional communications were observed in the ADHD probands (both ADHD-R and ADHD-P) relative to the controls. Furthermore, the ADHD-P showed even lower nodal efficiency in bilateral MFG relative to the ADHD-R, and a unique pattern of hyper-interactions between bilateral MFG during the cue-evoked attention processing.

A variety of existing clinical and neuroimaging studies have suggested that functional and structural abnormalities in the frontal and parietal areas (especially MFG and IPL) are associated with the onset of childhood ADHD. Both task-based and resting-state fMRI studies have reported functional alterations in MFG in children with ADHD (Yang et al., 2011; Li et al., 2012a; Li et al., 2013; Posner et al., 2014; Schulz et al., 2017). Decreased MFG cortical thickness has also been frequently reported in children with ADHD (Shaw et al., 2006; Li et al., 2007; Batty et al., 2010), and was found to be associated with the persistence of more severe childhood ADHD into adulthood (Shaw et al., 2006; Rajendran et al., 2013; Shaw et al., 2013). On the other hand, improved MFG activation and connectivity with other brain regions were found to be associated with the reduction of ADHD symptoms over development (Shaw et al., 2006; Shaw et al., 2010). The parietal lobe is another critical component of the attention network, which has been frequently reported in studies of ADHD (Bush, 2011). One of our previous fMRI studies showed significantly decreased right IPL activation for cognitive and motor control in adults with childhood ADHD that has persisted relative to the ADHD-R and healthy controls (Schulz et al., 2017). DTI studies suggested that disrupted white matter integrity in tracts connecting IPL and prefrontal cortices significantly contribute to the persistence of childhood ADHD into adulthood (Makris et al., 2008; Cortese et al., 2013). Together with these existing results, our findings suggest that severe functional impairments of right hemisphere frontal areas, especially right MFG, may impact normal fronto-parietal

functional interactions for sensory and cognitive information processing, and play an important role for the persistence of ADHD symptoms in young adults.

We further found that the nodal efficiency of right IFG was significantly negatively correlated with the ADHD symptom severity scores and reduced mean reaction time for cued targets in ADHD-R, while significantly positively correlated with the ADHD symptom severity scores in the ADHD-P. And nodal efficiency of right IPL was significantly positively correlated with increased inattentive symptoms in the ADHD-P. These findings depict distinct roles of the right frontal lobes for the remission or persistence of ADHD symptoms, respectively; and further suggest the significant involvement of right frontal and parietal lobes for symptom persistence of childhood ADHD.

In the results of this study, the putamen nucleus of the right striatum acted as a network hub for cue-evoked attention processing in ADHD-R. Based on the nature of the degree and betweenness-centrality measures, a network hub has significantly more connections with other nodes in the network, but not necessarily has higher nodal efficiency for transferring functional information in the network (Li et al., 2012b). Structural and functional alterations in the striatum have been widely reported in children (Xia et al., 2012) and adults (Wang et al., 2013) with ADHD. The super active putamen hub for cue-evoked attention processing in the ADHD-R may suggest the importance of optimal putamen function for symptom remission of childhood ADHD.

Some limitations of the study need to be considered. First, the study included both male and female subjects. Although it is unclear yet whether the neuropathological underpinnings of ADHD have gender differences, clinical studies have observed different symptoms and comorbidity profiles in male and female patients (Quinn and Madhoo, 2014). To partially remove gender-related effect, we added sex as a fixed effect covariate in the group-level analyses. We further compared the network property measures between the 16 male ADHD-R and 13 male ADHD-P and found a pattern of between-group differences similar with the primary results reported in the results section. Female-specific tests were not conducted due to the very limited number of female participants in the study. Second, the sample size of this study is relatively small. Therefore, the findings may have the preliminary nature. Our future research will focus on investigating the neuroanatomical bases of the functional and behavioral aberrances associated with symptom persistence of childhood ADHD, in a much larger sample that improves the statistic power.

3.5 Conclusion

This study found that right frontal lobe functional impairments that may relate to inefficient fronto-parietal functional interactions for sensory and cognitive information processing and symptom persistence in young adults with childhood ADHD.

CHAPTER 4

GRAY MATTER AND WHITE MATTER STRUCTURAL CORRELATES OF REMISSION AND PERSISTENCE IN ADULTS WITH CHILDHOOD ATTENTION-DEFICIT/HYPERACTIVITY DISORDER

4.1 Introduction

4.1.1 Background

A large number of existing studies suggest that ADHD symptoms in children are associated with widespread neuroanatomical alterations of brain. Substantial structural neuroimaging studies have found ADHD symptoms in childhood to be associated with decreased regional GM volume in frontal cortex, striatum, thalamus and cerebellum (Ellison-Wright et al., 2008; Bledsoe et al., 2011; Mahone et al., 2011). Reduced regional cortical GM thickness in frontal and parietal cortices have also been linked with ADHD symptoms (Batty et al., 2010; Almeida Montes et al., 2013). WM structural deficits, especially reduced WM volume and/or FA in the fronto-parietal, fronto-limbic, corona radiata, cerebellar- and temporo-occipital, and internal capsule fiber tracts have been consistently demonstrated in children with ADHD (Durston et al., 2004; Nagel et al., 2011; Peterson et al., 2011; Qiu et al., 2011; Xia et al., 2012).

The majority of existing clinical and neuroimaging studies in ADHD have focused on understanding the neural correlates of symptoms in cross-sectional samples of children or young adults. Far fewer studies have examined neural substrates associated with the

diverse adult outcomes of childhood ADHD. Neuroanatomical studies showed diverse results in adults with childhood ADHD. With structural MRI, Proal et al. found that compared to matched controls, ADHD probands had significantly decreased GM volume in prefrontal lobe, cerebellum, thalamus, and caudate, regardless of ADHD symptom remission or persistence (Proal et al., 2011); while Shaw et al. showed that significantly reduced cortical thickness was linked with symptom persistence (Shaw et al., 2013). A DTI study suggested that ADHD probands had WM disruptions in the superior longitudinal fasciculus (SLF) and cortico-limbic areas regardless of symptom remission or persistence (Gehricke et al., 2017); another study found that greater adult inattentiveness, but not hyperactivity/impulsivity, was associated with lower FA in inferior occipito-frontal fasciculus and uncinated fasciculus (Shaw et al., 2015); where Cortese et al. indicated no significant WM differences between the ADHD-R and ADHD-P (Cortese et al., 2013). The inconsistent findings from these neuroimaging studies in adults with childhood ADHD may be partially explained by differences in imaging modalities, analytic methods, and study cohorts. These existing studies have demonstrated neuroanatomical alterations in adults with childhood ADHD. However, most of them applied only single imaging modality (either structural MRI or DTI) to investigate GM morphometrical or WM integrity properties, without reporting both the GM and WM patterns in the same study cohort, and their impact on the adult outcome of childhood ADHD. This study aimed to fill this gap by applying both structural MRI and DTI in the same study sample to identify the

structural markers in GM and WM, that are associated with symptom persistence and remission in young adults with childhood ADHD (Luo et al., 2020b). Based on findings of previous studies from our group and others, we hypothesized that optimal structural development associated with the frontal and parietal lobes, such as greater regional GM thickness, higher FA of the WM tracts that connect subcortical structures (i.e., thalamus, caudate) and frontal/parietal cortices, may play an important role in symptom remission in young adults with childhood ADHD.

4.1.2 Bayes Inference and Monte Carlo Markov Chain Utilized in Project 2

For both structural MRI and DTI data analyses, Bayesian Inference was widely performed in Aim 2 with a Bayesian probabilistic model to calculating an expected probability, estimating the density, or other properties of the probability distribution. The direct calculation of the desired quantity from a model of interest is intractable for all but the most trivial probabilistic models. Instead, the expected probability or density must be approximated by other means. The desired calculation is typically a sum of a discrete distribution of many random variables or integral of a continuous distribution of many variables and is intractable to calculate. This problem exists in both schools of probability, although is perhaps more prevalent or common with Bayesian probability and integrating over a posterior distribution for a model. The typical solution is to draw independent samples from the probability distribution, then repeat this process many times to

approximate the desired quantity. This is referred to as Monte Carlo sampling or Monte Carlo integration, named for the city in Monaco that has many casinos. The problem with Monte Carlo sampling is that it does not work well in high-dimensions. This is firstly because of the curse of dimensionality, where the volume of the sample space increases exponentially with the number of parameters (dimensions). Secondly, and perhaps most critically, this is because Monte Carlo sampling assumes that each random sample drawn from the target distribution is independent and can be independently drawn. This is typically not the case or intractable for inference with Bayesian structured or graphical probabilistic models.

The solution to sampling probability distributions in high-dimensions is to use Markov Chain Monte Carlo (MCMC). Like Monte Carlo methods, MCMC was first developed around the same time as the development of the first computers and was used in calculations for particle physics required as part of the Manhattan project for developing the atomic bomb. It is the combination of Monte Carlo and Markov Chain.

Monte Carlo is a technique for randomly sampling a probability distribution and approximating a desired quantity. Monte Carlo methods typically assume that we can efficiently draw samples from the target distribution. From the samples that are drawn, we can then estimate the sum or integral quantity as the mean or variance of the drawn samples. A useful way to think about a Monte Carlo sampling process is to consider a complex two-dimensional shape, such as a spiral. We cannot easily define a function to describe the

spiral, but we may be able to draw samples from the domain and determine if they are part of the spiral or not. Together, a large number of samples drawn from the domain will allow us to summarize the shape (probability density) of the spiral.

Markov chain is a systematic method for generating a sequence of random variables where the current value is probabilistically dependent on the value of the prior variable. Specifically, selecting the next variable is only dependent upon the last variable in the chain. Consider a board game that involves rolling dice, such as snakes and ladders (or chutes and ladders). The roll of a die has a uniform probability distribution across 6 stages (integers 1 to 6). You have a position on the board, but your next position on the board is only based on the current position and the random roll of the dice. Your specific positions on the board form a Markov chain. Another example of a Markov chain is a random walk in one dimension, where the possible moves are 1, -1, chosen with equal probability, and the next point on the number line in the walk is only dependent upon the current position and the randomly chosen move.

Combining these two methods, Markov Chain and Monte Carlo, allows random sampling of high-dimensional probability distributions that honors the probabilistic dependence between samples by constructing a Markov Chain that comprise the Monte Carlo sample. Specifically, MCMC is for performing inference (e.g. estimating a quantity or a density) for probability distributions where independent samples from the distribution cannot be drawn, or cannot be drawn easily. As such, Monte Carlo sampling cannot be

used. Instead, samples are drawn from the probability distribution by constructing a Markov Chain, where the next sample that is drawn from the probability distribution is dependent upon the last sample that was drawn. The idea is that the chain will settle on (find equilibrium) on the desired quantity we are inferring. Yet, we are still sampling from the target probability distribution with the goal of approximating a desired quantity, so it is appropriate to refer to the resulting collection of samples as a Monte Carlo sample, e.g. extent of samples drawn often forms one long Markov chain. The idea of imposing a dependency between samples may seem odd at first, but may make more sense if we consider domains like the random walk or snakes and ladders games, where such dependency between samples is required.

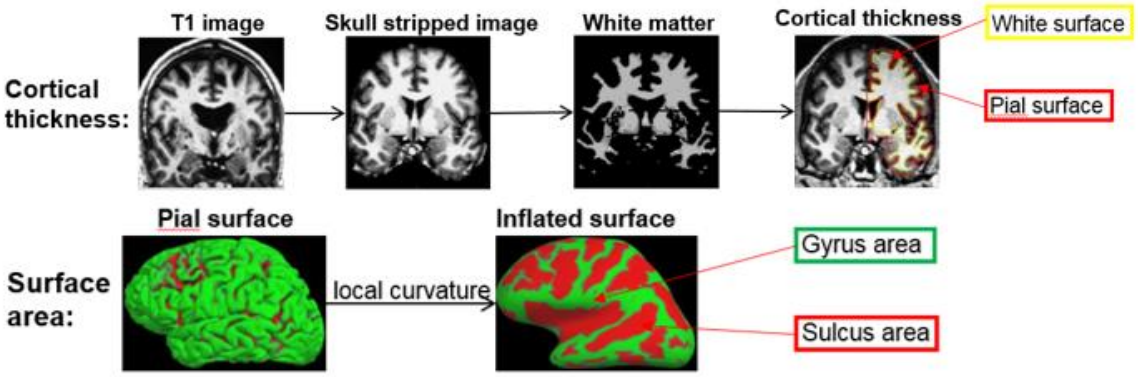

Figure 4.1 Workflow of structural magnetic resonance imaging data analysis.

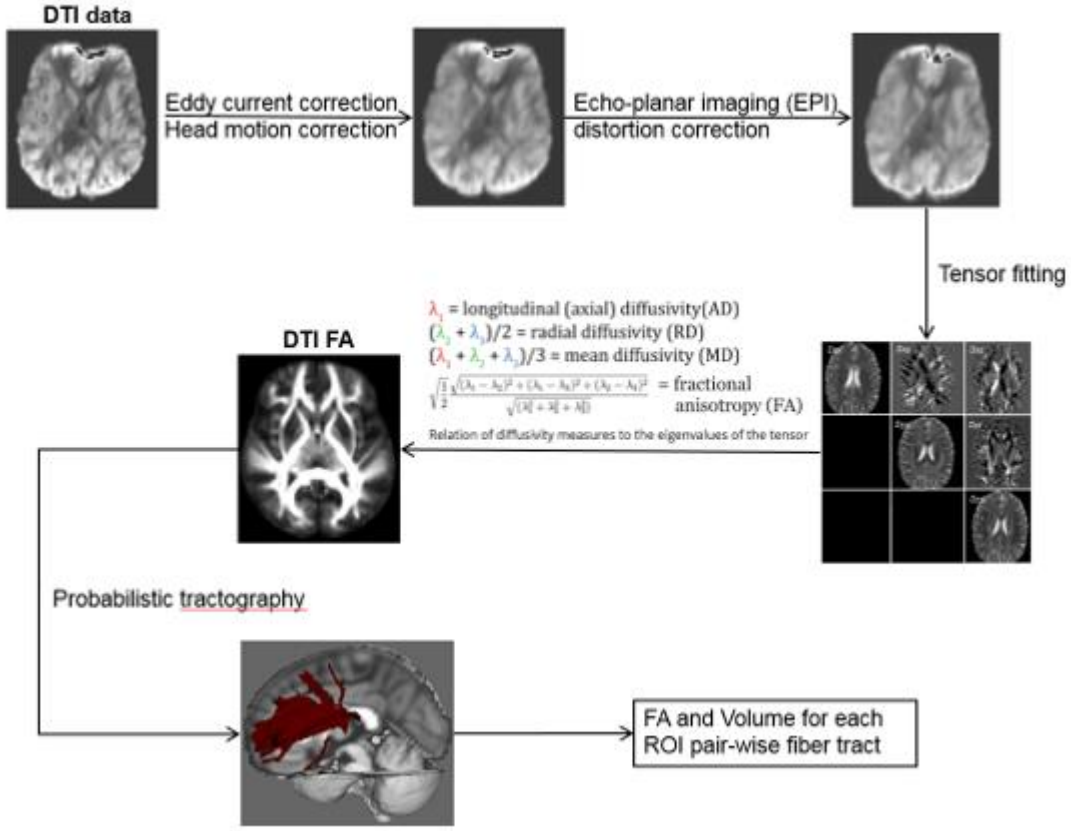

Figure 4.2 Workflow of diffusion tensor imaging data analysis.

4.2 Experimental Strategy

4.2.1 Participants

A total of 32 ADHD probands and 35 controls were involved in this project. Among the 32 ADHD probands, 16 were classified as ADHD-P and 16 as ADHD-R, and were able to provide usable T1-weighted and DTI data. All the ADHD probands had a history of treatment with short-acting psychostimulants. Mean duration of treatment was 2.03 years (SD=3.21) for the subgroup of ADHD-R and 4.18 years (SD=4.12) for the subgroup of ADHD-P (t=-1.604, *p*=0.12). Clinical and demographic information are listed in Table 4.1.

4.2.2 Individual-level Structural Magnetic Resonance Imaging Data Analyses

T1-weighted data were reconstructed into a 3-dimensional cortical model for thickness and area estimations using FreeSurfer v.5.3.0 (https://surfer.nmr.mgh.harvard.edu). Each data point was first registered with the Talairach atlas to compute the transformation matrix using an affine registration method, which was developed and distributed by the MNI. Then intensity variations caused by magnetic field inhomogeneities were corrected using Voronoi partitioning algorithm. The skull was stripped using a deformable template model. Cutting planes were defined to separate the left and right hemispheres and to remove the cerebellum and brainstem. Two mess surfaces (mess of grids created using surface tessellation technique) were then generated between WM and GM (white matter surface), as well as between GM and cerebrospinal fluid (pial surface). The distance between the two closet vertices of the white matter and pial surfaces presented the cortical thickness at that specific location, validated using training data (Rosas et al., 2002). Regional cortical thickness and area in 68 bilateral cortical regions were estimated based on the Desikan atlas (Desikan et al., 2006).

Each of 37 subcortical structures/nuclei was first labelled after the initial registration with the Talairach atlas, and then refined based on a manually labelled model constructed according to prior knowledge of spatial relationships acquired with a training data set (Fischl et al., 2002). Volume of each subcortical structure was then calculated.

To adjust head-size variation related influence on these cortical and subcortical GM measures, the head-size scaling factor of each subject was calculated by normalizing the T1-weighted data with the template provided in FSL/SIENA (Smith et al., 2002). The normalized thickness and area of each cortical region and volume of each subcortical structure were finally estimated by multiplying the original value with the scaling factor of that subject.

4.2.3 Individual-level Diffusion Tensor Imaging Data Analyses

DTI data from each subject was first processed using the Diffusion Toolbox (FDT Version 3.0) from FSL (Behrens et al., 2007). After eddy current and head motion corrections, the diffusion-weight images were registered to the additionally acquired non-diffusion-weighted reference image (b0 image) using an affine, 12 degrees of freedom registration. The FA value and principle diffusion direction at each brain voxel were calculated. WM probabilistic tractography between each pair of 18 ROIs was constructed using the FSL/BEDPOSTX toolbox (Behrens et al., 2007). These 18 ROIs (including thalamus, putamen and caudate nuclei from striatum, hippocampus, and frontal, parietal, occipital, temporal, and insular cortices in both hemispheres) were created based on the Harvard-Oxford Cortical Atlases and the Julich Histological Atlas from the MNI standard space, and mapped to the DTI data. We used the multi-fiber probabilistic connectivity-based method to determine the number of pathways between each seed and

target ROIs. The default setting of parameters for Markov Chain Monte Carlo estimation of the probabilistic tractography was utilized: 5000 individual pathways were drawn on the principle fiber direction of each voxel within the seed ROI; curvature threshold of 80° to exclude implausible pathways; a maximum number of 2000 travel steps of each sample pathway and a 0.2 mm step length. The number of pathways that existed through each voxel from the remainder of the brain was labeled. The non-zero labeling voxels were taken as the initial elements of the tracts between the seed and target ROIs. The brain voxels with low probability of connection were removed from the tract, if one had a number of pathways that was less than the average of the pathway numbers from all the non-zero labeling voxels. At the end, a total of 20 cortico-cortical (including bilateral fronto-parietal, fronto-occipital, fronto-temporal, fronto-insular, parieto-occipital, parieto-temporal, parieto-insular, occipito-temporal, occipito-insular, temporo-insular) and 40 subcortico-cortical (including bilateral thalamo-frontal, thalamo-parietal, thalamo-occipital, thalamo-temporal, thalamo-insular, putamen-frontal, putamen-parietal, putamen-occipital, putamen-temporal, putamen-insular, caudate-frontal, caudate-parietal, caudate-occipital, caudate-temporal, caudate-insular, hippocampo-frontal, hippocampo-parietal, hippocampo-occipital, hippocampo-temporal, hippocampo-insular) WM fiber tracts were generated. Average FA and volume (number of voxels times voxel size) of each identified WM tract were estimated.

4.2.4 Group Statistical Analyses

The clinical, neurocognitive and demographic measures were compared using chi-square tests for discrete variables and unpaired two-sample t-tests for continuous variables, between groups of controls and ADHD probands, and further between the two ADHD subgroups (ADHD-R and ADHD-P) using SPSS18 (SPSS Inc, Somers, NY).

The structural MRI- and DTI-based neuroimaging measures (including regional cortical thickness, surface area, volume of each subcortical structure, FA and volume of each WM fiber tract) were compared between the groups of controls and ADHD probands, as well as between the subgroups of ADHD-R and ADHD-P, using ANCOVA with gender, age, IQ and socioeconomic status (SES) as covariates. Bonferroni correction for multiple comparisons (at a corrected α=0.05) was applied to control potential false positive results of these group comparisons. For the structural MRI-based measures, the number of ROIs (a total of 105 ROIs, including 68 bilateral cortical regions and 37 subcortical structures) were controlled; while for DTI-based measures, the number of WM tracts (a total of 60 tracts) were controlled.

Partial correlation analysis was utilized to assess associations between the GM and WM brain measures that showed between-group differences and the clinical symptom measures (the T-scores for inattentive and hyperactive-impulsive symptoms derived from the CAARS collected during the visit of MRI scan) in the group of ADHD probands Age,

gender, IQ and SES were added as covariates. Bonferroni correction was used to correct the number of partial correlation procedures (a total of 16) at a corrected α=0.05.

4.3 Results

4.3.1 Clinical, Behavioral and Demographic Measures

As shown in Table 4.1, there were no significant demographic differences between the groups although relative to controls, ADHD probands tended to have lower IQ and SES.

4.3.2 Brain Anatomical Measures

Significantly decreased volume in right putamen was observed in ADHD probands when compared to controls (p=0.045). Compared to the ADHD-P group, those with ADHD-R showed significantly increased cortical surface area in bilateral parahippocampal gyri (Left: p=0.05; Right: p=0.008), left paracentral gyrus (p=0.012), and right transverse temporal gyrus (p=0.037) (see Table 4.2).

Group comparisons of WM measures showed significantly decreased volume of the left parieto-insular fiber tracts (p=0.041) in ADHD probands relative to controls. Compared to ADHD-R, the subgroup of ADHD-P showed significantly decreased volume in two cortico-cortical fiber tracts (right hippocampo-frontal (p=0.037) and right parieto-insular (p=0.038)), and in the WM tracts connecting bilateral caudate nuclei of the striatum with all the five cortical ROIs of the same hemispheres (p<0.001) (see Table 4.3).

Table 4.1 Demographic and Clinical Characteristics in Groups of Controls and ADHD Probands (including Remitted and Persistent)

	Controls (N=35)	**Probands (N=32)**		**ADHD-R (N=16)**	**ADHD-P (N=16)**	
	Mean (SD)	**Mean (SD)**	***p***	**Mean (SD)**	**Mean (SD)**	***p***
Age	24.24 (2.3)	24.60 (2.1)	*0.51*	24.81 (2.3)	24.39 (1.9)	*0.57*
Full-scale IQ*	104.21 (15.7)	96.81 (14.3)	*0.07*	99.58 (14.2)	94.11 (11.5)	*0.24*
Conners' Adult ADHD Rating Scale (T-score)						
Inattentive	45.74 (8.9)	55.72 (12.8)	*0.001*	49.99 (11.5)	61.44 (11.6)	*<0.01*
Hyperacitive/impulsive	42.89 (6.3)	52.35 (12.0)	*<0.001*	45.57 (9.3)	59.13 (10.7)	*0.001*
ADHD Total	43.43 (8.3)	55.22 (13.9)	*<0.001*	48.12 (11.9)	62.31 (12.2)	*0.002*
ADHD semistructured interview (number of symptoms)	0.83 (1.4)	6.30 (4.7)	*<0.001*	2.04 (1.6)	10.73 (3.0)	*<0.01*
	N (%)	**N (%)**	***p***	**N (%)**	**N (%)**	***p***
Male	30 (85.7)	27 (84.4)	*0.88*	14 (87.5)	13 (81.3)	*0.63*
Right-handed	31 (88.6)	28 (87.5)	*0.89*	14 (87.5)	14 (87./5)	*1*
Race			*0.41*			*0.70*
Caucasian	14 (40.0)	17 (53.1)		8 (50.0)	9 (56.3)	
African American	13 (37.1)	7 (21.9)		4 (25.0)	3 (18.8)	
More than one race	6 (17.1)	8 (25)		4 (25.0)	4 (25.0)	
Asian	2 (5.7)	0 (0)		0 (0)	0 (0)	
Ethnicity			*0.21*			*0.72*
Hispanic/Latino	12 (34.3)	15 (46.9)		7 (43.8)	8 (50.0)	
Current mood disorder	4 (11.4)	8 (25.0)	*0.15*	2 (12.5)	6 (37.5)	*0.10*
Current anxiety disorder	10 (28.6)	10 (31.3)	*0.81*	3 (18.8)	7 (43.8)	*0.13*
Current substance disorder	7 (20)	15 (46.9)	*0.02*	7 (43.8)	8 (50.0)	*0.72*

*Assessed in adolescence

Table 4.2 Gray Matter Neuroimaging Measures that Show Significant Between-group Differences

Group	Anatomical location	Measure	F-value	*p*-value after Bonferroni correction
CON>PRO	R. Putamen	Volume	8.245	*0.048*
ADHD-R>ADHD-P	L./R. Parahippocampal gyrus	Regional Cortical Surface Area	8.311/13.547	*0.048/0.006*
	L. Paracentral gyrus		10.294	*0.018*
	R. Transverse temporal gyrus		8.906	*0.036*

Table 4.3 White Matter Neuroimaging Measures that Show Significant Between-group Differences

Group	White matter fiber tract	Measure	F-value	*p*-value after Bonferroni correction
CON>PRO	L. parieto-insular tract	Volume	9.622	*0.039*
ADHD-R>ADHD-P	L./R. caudate-frontal tracts	Volume	42.983/33.712	*<0.001/<0.001*
	L./R. caudate-parietal tracts		53.104/32.716	*<0.001/<0.001*
	L./R. caudate-occipital tracts		56.444/32.722	*<0.001/<0.001*
	L./R. caudate-temporal tracts		56.349/32.687	*<0.001/<0.001*
	L./R. caudate-insular tracts		56.448/32.66	*<0.001/<0.001*
	R. hippocampo-frontal tract		14.197	*0.036*
	R. parieto-insular tract		13.262	*0.036*

4.3.3 Associations Between Brain and Behavioral Measures

Dimensional analyses between the GM and WM measures (listed in Tables 4.2 and 4.3) and the clinical symptom measures indicated that among the ADHD probands, greater FA of the left caudate-parietal WM fiber tract was significantly associated with reduced hyperactive/impulsive symptoms (Figure 4.3, r=-0.389, *p*=0.037).

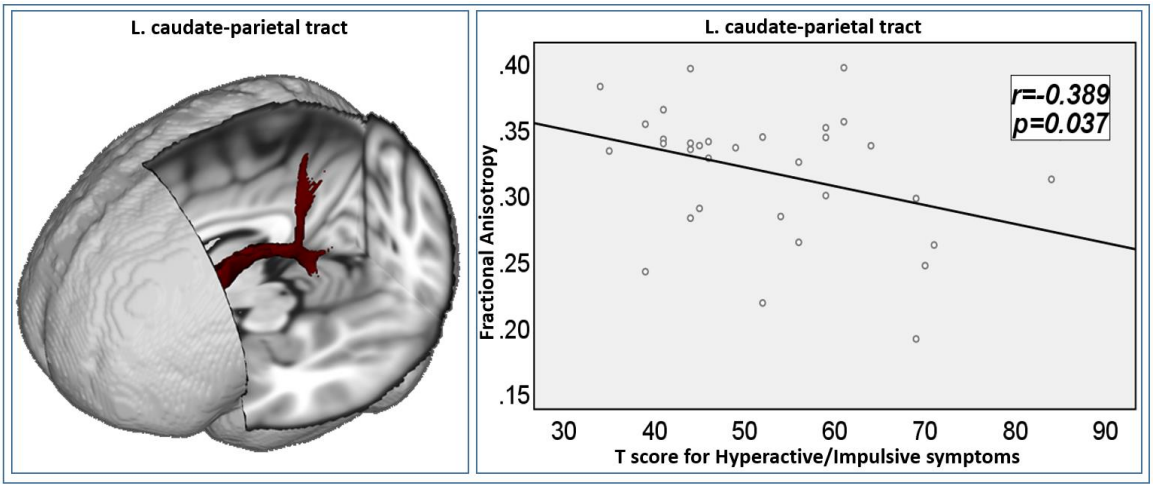

Figure 4.3 In the group of ADHD probands, greater fractional anisotropy of the left caudate-parietal white matter fiber tract was significantly associated with reduced hyperactive-impulsive symptoms measured by the DSM standard T-score.

4.4 Discussion

The present study investigated GM and WM structural differences between young adults with childhood ADHD and group-matched controls, and between the subgroups of remitters and persisters within the ADHD probands. Compared to controls, significantly

reduced GM volume of the putamen in right hemisphere was observed in the ADHD probands. The putamen and caudate nucleus together form the dorsal striatum, and play a key role in the CSTC loops for attention and higher order cognitive processes (Alexander et al., 1986; Ring and Serra-Mestres, 2002). A large number of structural MRI and fMRI studies have reported the linkage of putamen-related anatomical and functional abnormalities and onset of ADHD in children (Max et al., 2002; Ellison-Wright et al., 2008; Nakao et al., 2011; Frodl and Skokauskas, 2012). Putamen-related structural alterations have also been tested in neuroimaging studies focusing on adults with ADHD which yielded inconsistent results, with some reported reduced putamen volume in adults with ADHD (Seidman et al., 2011; Onnink et al., 2014), and others reported increased putamen volume (Greven et al., 2015) or no significant differences (Seidman et al., 2006) when compared to group-matched controls. The inconsistence of these existing studies may have been caused by technical differences for putamen extractions, and sample-related biases such as the very wide age ranges involved in these studies (Greven et al., 2015). Adding into the debating literature, our result of significantly reduced putamen GM volume in young adults with childhood ADHD (regardless of their clinical outcomes) suggests its significant linkage with the emergence of ADHD during their childhood.

Comparing to controls, we also found that the ADHD probands had significantly reduced volume of the left hemisphere parieto-insular WM tract; while relative to the ADHD-R, the ADHD-P had significantly smaller volume of the right hemisphere

parieto-insular WM tract. The parieto-insular WM fiber tract is an important structural component of the vestibular system, and has been suggested to link with static and dynamic balance control (Perennou et al., 2000; Ustinova et al., 2001; Shum and Pang, 2009; Frank and Greenlee, 2018). Vestibular system deficiency, which can cause inappropriate postural condition or impaired balance function, has been found to associate with cognitive deficits and behavioral symptoms in ADHD patients (Clark et al., 2008; Shum and Pang, 2009; Haghshenas et al., 2014). Merging with the results of existing studies, our findings of the underdeveloped parieto-insular WM fiber tracts in adults with childhood ADHD, especially in those with persistent ADHD symptoms, suggest that parieto-insular WM structural alterations may interact with the vestibular system functional alterations, and together contribute to the onset and symptom persistence of ADHD.

Within the probands, we further found that the ADHD-R had significantly larger surface area in bilateral parahippocampal, left paracentral, and right transverse temporal gyri, as well as significantly greater volume of WM fiber tracts connecting caudate with the frontal, parietal, occipital, temporal, and insular cortices when compared to the ADHD-P. Existing studies have reported that ADHD-R had increased parahippocampal cortical thickness compared to ADHD-P (Proal et al., 2011). Further studies have implicated that parahippocampal gyrus interacts with the ventralateral prefrontal cortex (VLPFC), both significantly contribute to appropriate inhibitory control (Deacon et al., 1983; Schulz et al., 2005). Parahippocampal cortical volume reduction has been observed in both children and

adolescents with ADHD, compared to group-matched controls (Carmona et al., 2005; Noordermeer et al., 2017).

Caudate plays a critically important role in cognitive control (Grahn et al., 2008; Chiu et al., 2017). Structural and functional deficits associated with caudate have been widely observed in children and adults with ADHD (Frodl and Skokauskas, 2012; Onnink et al., 2014; Szekely et al., 2017). Substantial structural MRI studies have revealed that children with ADHD had smaller caudate volume relative to controls (Castellanos et al., 2002). Task-based fMRI studies showed significantly decreased caudate activation in children with ADHD (Vaidya et al., 2005) and adults with childhood ADHD (Szekely et al., 2017), during attention and inhibitory control processes. Our findings of significantly smaller volume of the WM fiber tracts connecting caudate with all five cortices bilaterally in the ADHD-P suggest that caudate-associated widespread WM underdevelopment may play important roles in symptom persistence of ADHD. This hypothesis can also be supported by multiple existing DTI studies that showed immature WM organizations involving caudate and cortical structures in children and adults with ADHD (Casey et al., 1997; Castellanos et al., 2002; Ashtari et al., 2005; Shang et al., 2013).

In addition, we found that the FA of left caudate-parietal tracts was significantly negatively correlated with the CAARS T-score for hyperactive/impulsive symptoms in ADHD probands. The caudate-parietal WM tract is one of the most important structural component of the CSTC loops, which subserves maintaining the modifications of spatial

attention via reinforcement learning, and supports the integration of reward, attention, and executive control (Jarbo and Verstynen, 2015). Reduced parietal activation during cognitive control has been linked to the persistence of ADHD symptoms in adults with childhood ADHD (Schulz et al., 2017; Szekely et al., 2017). Reduced caudate and parietal lobe activation during inhibitory control processing were found to be associated with increased inattentive and impulsive symptoms in adults with ADHD diagnosed in childhood (Schneider et al., 2010). Together with these existing findings, we suggest that optimal structural development in the caudate-parietal WM tract may partially modulate the functional integrity of caudate and parietal cortex, and together contribute to symptom remission in adults with childhood ADHD.

4.5 Conclusion

In summary, together with existing findings, results of this study suggest that WM structural development in tracts that connect caudate with cortical areas, especially in the caudate-parietal path, is a critical determining factor of outcomes in adults with childhood ADHD. The current study has some limitations. First, our cohort consisted of both male and female subjects, but many more males. It is still unclear whether the neuropathological underpinnings of ADHD differ between males and females. To partially remove gender-related effects, sex was added as a fixed effect covariate in the group-level analyses. Second, the sample size of this study is relatively small. Therefore, the findings

must be considered preliminary. Future work will need a much larger cohort from a longitudinal study consisting of multi-scan neuroimaging data, to determine the neural underpinnings of longitudinal trajectories of childhood ADHD.

CHAPTER 5

MULTIMODAL NEUROIMAGING-BASED PREDICTION OF ADULT OUTCOMES IN CHILDHOOD-ONSET ATTENTION-DEFICIT/HYPERACTIVITY DISORDER USING ENSEMBLE LEARNING TECHNIQUES

5.1 Introduction

5.1.1 Background

ADHD is a highly prevalent heterogeneous neurodevelopmental disorder. Diagnostic standards for ADHD are clinical symptom-based, and rely primarily on subjective reports collected from multiple sources, which often cause biases and inconsistencies of the diagnoses. ELTs, which integrate results from multiple basic classifiers by using voting (Lam and Suen, 1997; Ruta and Gabrys, 2005), bagging (Breiman, 1996), stacking (Wolpert, 1992), or boosting (Schapire, 1990; Yoav Freund and Schapire, 1997; Johnston et al., 2014) strategies, have been recently developed in the big data science field, to deal with complicated feature variations, biases, and optimized prediction performances (Wang et al., 2011; Deng and Platt, 2014). ELTs have been applied in three recent studies to discriminate patients with ADHD from controls (Eloyan et al., 2012; Tenev et al., 2014; Zhang-James et al., 2019). Eloyan and colleagues applied a voting-based ELT, along with hold-out technique for CV, in structural MRI and resting-state fMRI data from children with ADHD and controls, and reported an important group discrimination predictor of dorsomedial-dorsolateral functional connectivity in the motor network (Eloyan et al.,

2012). Voting-based ELT has also been applied in EEG data collected from adults with ADHD and controls, without reporting the most important discrimination predicators (Tenev et al., 2014). Very recently, Zhang-James et al. applied ELTs in structural MRI data from patients with ADHD (both adults and children) and controls, and suggested that GM volume of bilateral caudate and thalamus and orbitofrontal surface area significantly contribute to successful group discrimination (Zhang-James et al., 2019). However, clarifications about optimization strategies was lacking and low accuracy of discriminations (<0.65) was reported.

The current study applied ELTs to structural MRI, DTI, and task-based fMRI data collected from a sample of adults with childhood ADHD who were clinically followed from ages 7-11 years and never-ADHD controls who have been followed since adolescence (Luo et al., 2020a). All currently available optimization strategies (i.e., voting, bagging, boosting and stacking techniques) were tested in a pool of semi-final classification results generated by seven basic classifiers (including K-Nearest Neighbors (KNN), SVM, logistic regression (LR), Naïve Bayes (NB), linear discriminant analysis (LDA), random forest (RF), and multilayer perceptron (MLP)). A nested CV including an inner LOOCV and an outer 5-fold CV were applied with grid search to tune the hyperparameters and minimize the overfitting. The high-dimensional neuroimaging features for classification included regional cortical GM thickness and surface area, GM volume of subcortical structures estimated from structural MRI data, volume and FA of

major WM fiber tracts derived from DTI data, the pair-wise regional connectivity and global/nodal topological properties (i.e., global-, local-, and nodal-efficiency, etc.) of the cue-evoked attention processing network computed from task-based fMRI data. Based on findings from existing studies (Proal et al., 2011; Clerkin et al., 2013; Shaw et al., 2013; Francx et al., 2015; Shaw et al., 2015; Luo et al., 2018), we hypothesized that structural and functional alterations in frontal, parietal, and subcortical areas and their interactions would significantly contribute to accurate discrimination of ADHD probands (adults diagnosed with ADHD in childhood) from controls; while abnormal fronto-parietal hyper-communications in right hemisphere would play an important role in inattentive and hyperactive/impulsive symptom persistence in adults with childhood ADHD. Finally, we hypothesized that classification performance parameters (accuracy, AUC) of the ROC, etc.) derived from ELT-based procedures would be superior to those of basic model-based procedures.

5.1.2 Introduction to Machine Learning Classification Models

5.1.2.1 K-Nearest Neighbors. KNN algorithm is a type of supervised machine learning algorithm which can be used for both classification as well as regression predictive problems. However, it is mainly used for classification predictive problems in industry. KNN algorithm uses ‘feature similarity’ to predict the values of new data points

which further means that the new data point will be assigned a value based on how closely it matches the points in the training set (Figure 5.1).

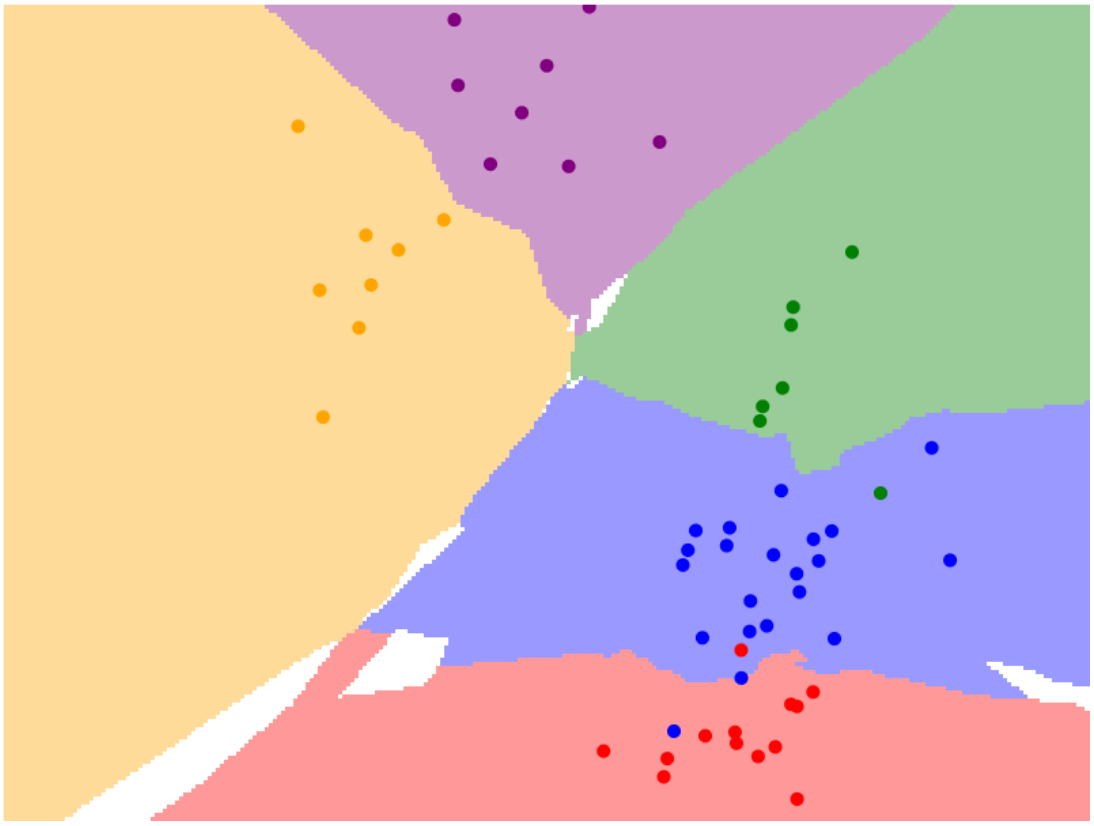

Figure 5.1 Diagram showing the classification performance of KNN algorithm. Each point in the plane is colored with the class that would be assigned to it using the KNN algorithm; points for which the KNN algorithm results in a tie are colored white.

5.1.2.2 Support Vector Machine. A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. To separate the two classes

of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e., the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence. Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine when the number of features exceeds 3 (Figure 5.2).

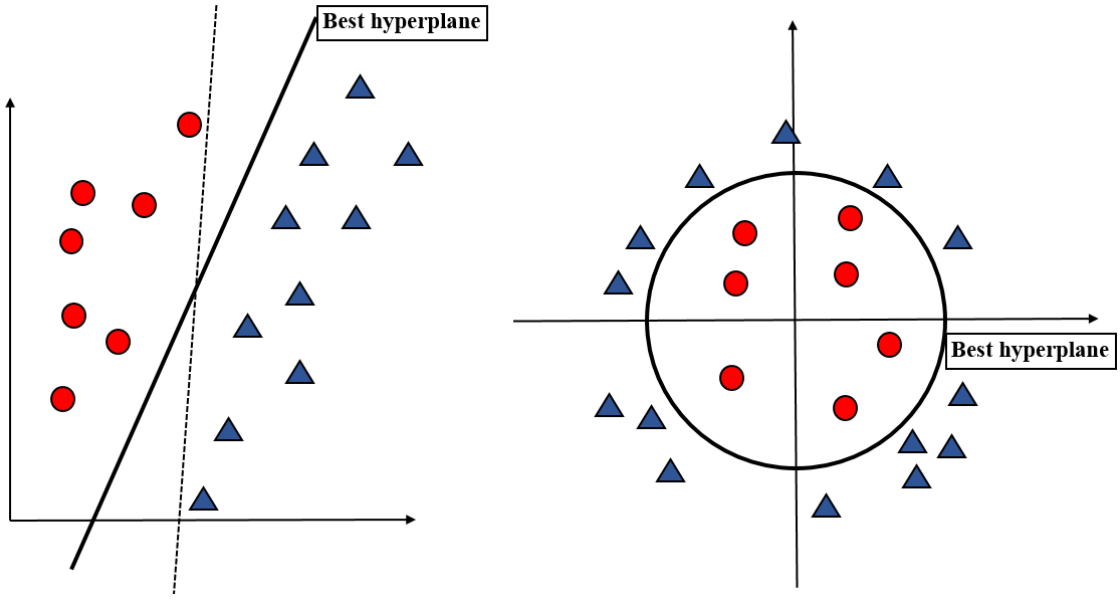

Figure 5.2 Best hyperplane for linear and non-linear data in SVM algorithm.

5.1.2.3 Logistic Regression. Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. Mathematically, a binary logistic model has a dependent variable with two possible values, such as pass/fail which is represented by an indicator variable, where the two values are labeled 0 and 1 (Figure 5.3). In the logistic model, the log-odds for the value labeled 1 is a linear combination of one or more independent variables; the independent variables can each be a binary variable or a continuous variable. The corresponding probability of the value labeled 1 can vary between 0 and 1, hence the labeling; the function that converts log-odds to probability is the logistic function.

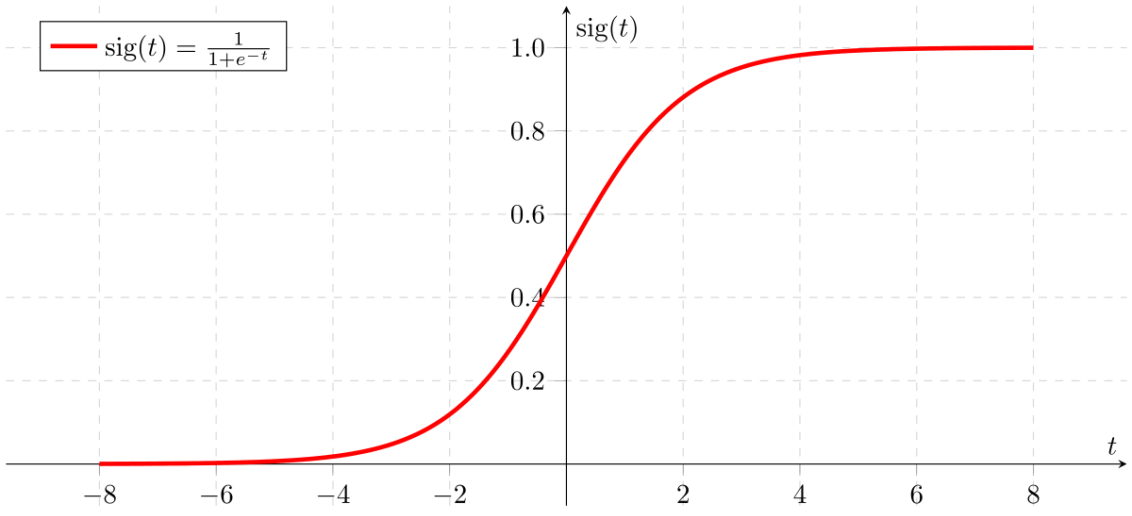

Figure 5.3 A simple logistic regression using sigmoid function.
Source: https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/. Accessed on April 10, 2020.

5.1.2.4 Naïve Bayes. A Naive Bayes classifier is a probabilistic machine learning model that's used for classification task. The crux of the classifier is based on the Bayes theorem, which is defined as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{5.1}$$

Using Bayes theorem, we can find the probability of A happening, given that B has occurred. Here, B is the evidence and A is the hypothesis. The assumption made here is that the predictors/features are independent. That is presence of one particular feature does not affect the other. Hence it is called naive.

5.1.2.5 Linear Discriminant Analysis. LDA is a dimensionality reduction technique used as a preprocessing step in machine learning and pattern classification applications. The main goal of dimensionality reduction techniques is to reduce the dimensions by removing the redundant and dependent features by transforming the features from higher dimensional space to a space with lower dimensions, which maximizes the between class variance and minimize the within class variance. LDA is a supervised classification technique which takes labels into consideration. This category of dimensionality reduction is used in biometrics, bioinformatics and chemistry.

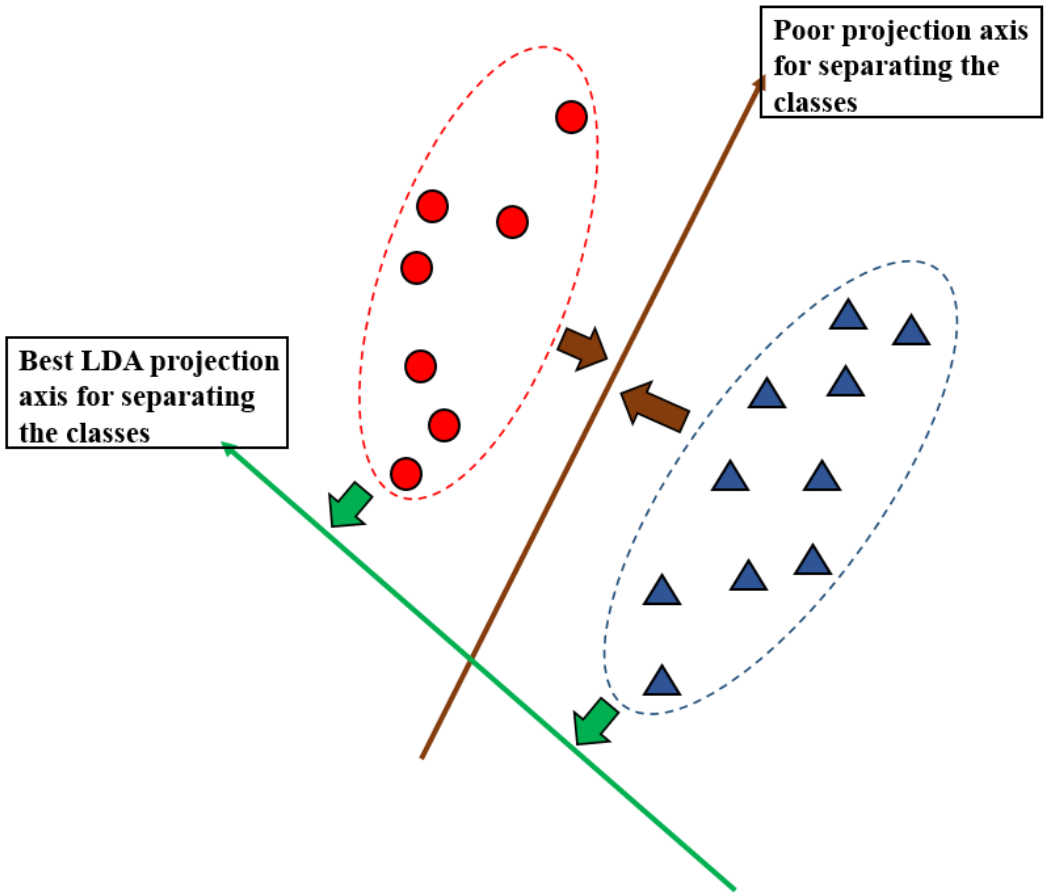

Figure 5.4 Different projections in linear discriminant analysis.

5.1.2.6 Random Forest. The random forest is a model made up of many decision trees. Rather than just simply averaging the prediction of trees (which we could call a "forest"), this model uses two key concepts that gives it the name random, which includes (a) random sampling of training data points when building trees; (b) random subsets of features considered when splitting nodes.

When training, each tree in a random forest learns from a random sample of the data points. The samples are drawn with replacement, known as bootstrapping, which means that some samples will be used multiple times in a single tree. The idea is that by training each tree on different samples, although each tree might have high variance with respect to a particular set of the training data, overall, the entire forest will have lower variance but not at the cost of increasing the bias. At test time, predictions are made by averaging the predictions of each decision tree. This procedure of training each individual learner on different bootstrapped subsets of the data and then averaging the predictions is known as bagging, short for bootstrap aggregating.

The other main concept in the random forest is that only a subset of all the features are considered for splitting each node in each decision tree. Generally, this is set to square root of the number of features for classification.

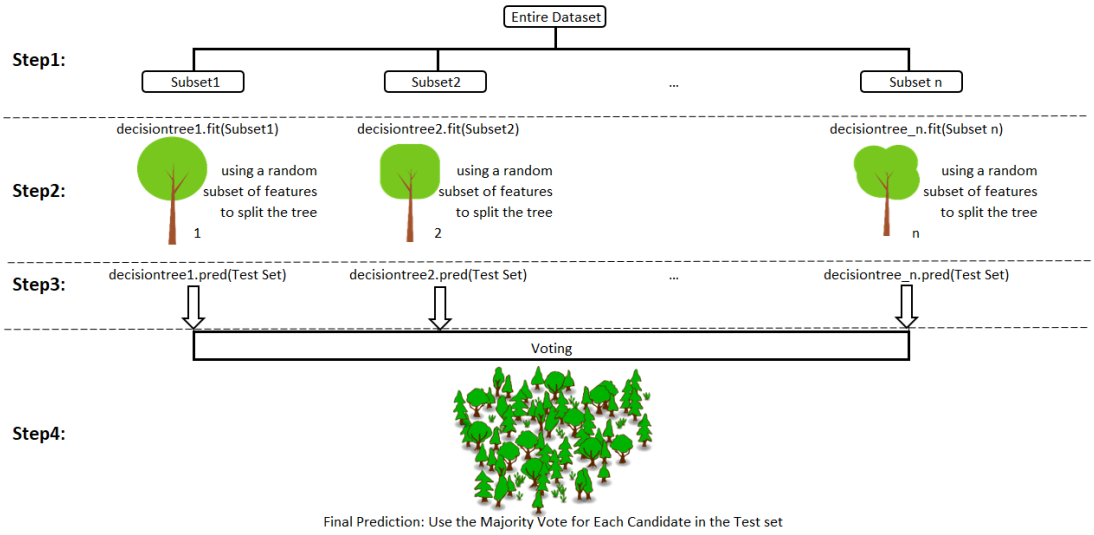

Figure 5.5 Workflow of random forest.
Source:
https://towardsdatascience.com/basic-ensemble-learning-random-forest-adaboost-gradient-boosting-step-by-step-explained-95d49d1e2725. Accessed on April 10, 2020.

5.1.2.7 Multilayer Perceptron. A MLP is a class of feedforward artificial neural network. The term MLP is used ambiguously, sometimes loosely to refer to any feedforward ANN, sometimes strictly to refer to networks composed of multiple layers of perceptrons (with threshold activation). Multilayer perceptrons are sometimes colloquially referred to as "vanilla" neural networks, especially when they have a single hidden layer. An MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training.

Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

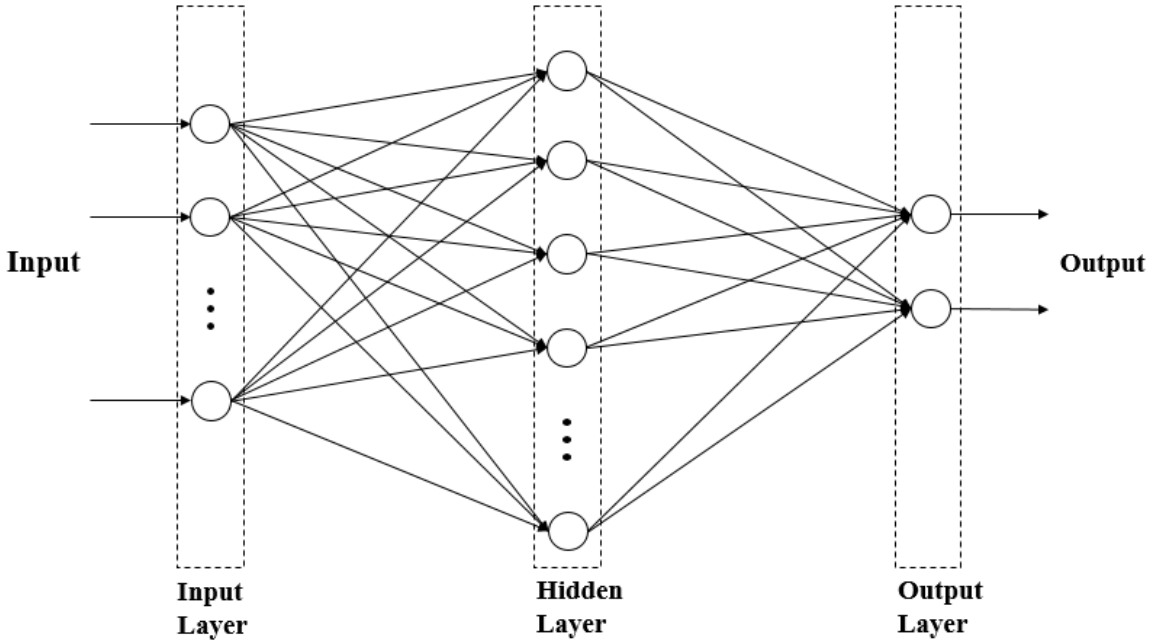

Figure 5.6 Structure of multilayer perceptron.

5.1.2.8 Hierarchical Clustering. In data mining and statistics, hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types: agglomerative and divisive. Agglomerative is a bottom-up approach with each observation starts in its own cluster, and pairs of clusters are merges as one moves up the hierarchy; while divisive is a top-down approach with all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

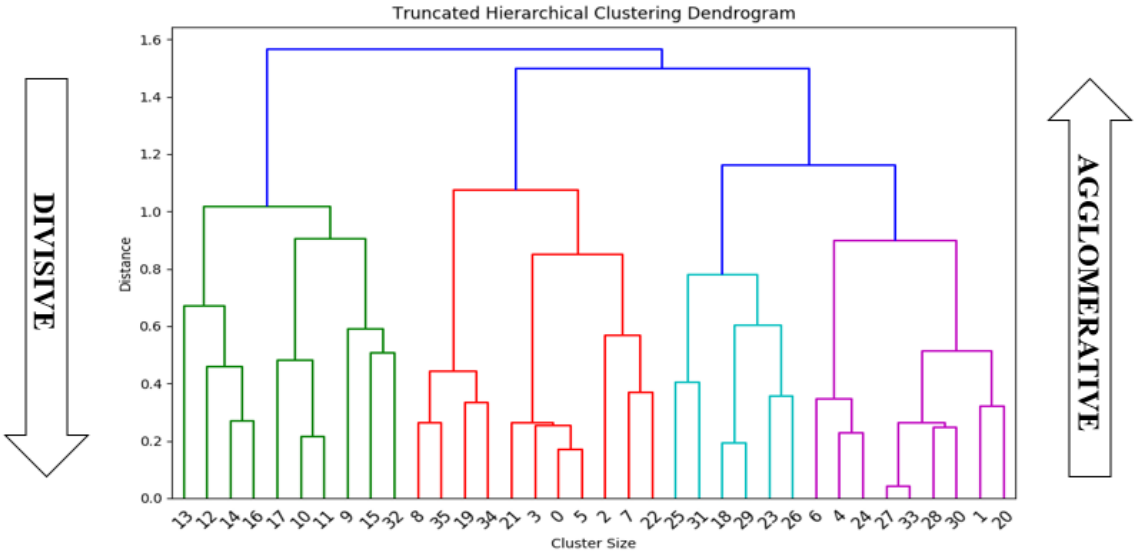

Figure 5.7 Structure of hierarchical clustering.

5.1.3 Introduction to Ensemble Learning Technique

5.1.3.1 Voting. Voting is one of the simplest ways of combining the predictions from multiple machine learning algorithms. It works by first creating two or more

standalone models from your training dataset. A Voting Classifier can then be used to wrap your models and average the predictions of the sub-models when asked to make predictions for new data. The predictions of the sub-models can be weighted, but specifying the weights for classifiers manually or even heuristically is difficult. Thus, the majority voting is widely utilized in many existing studies.

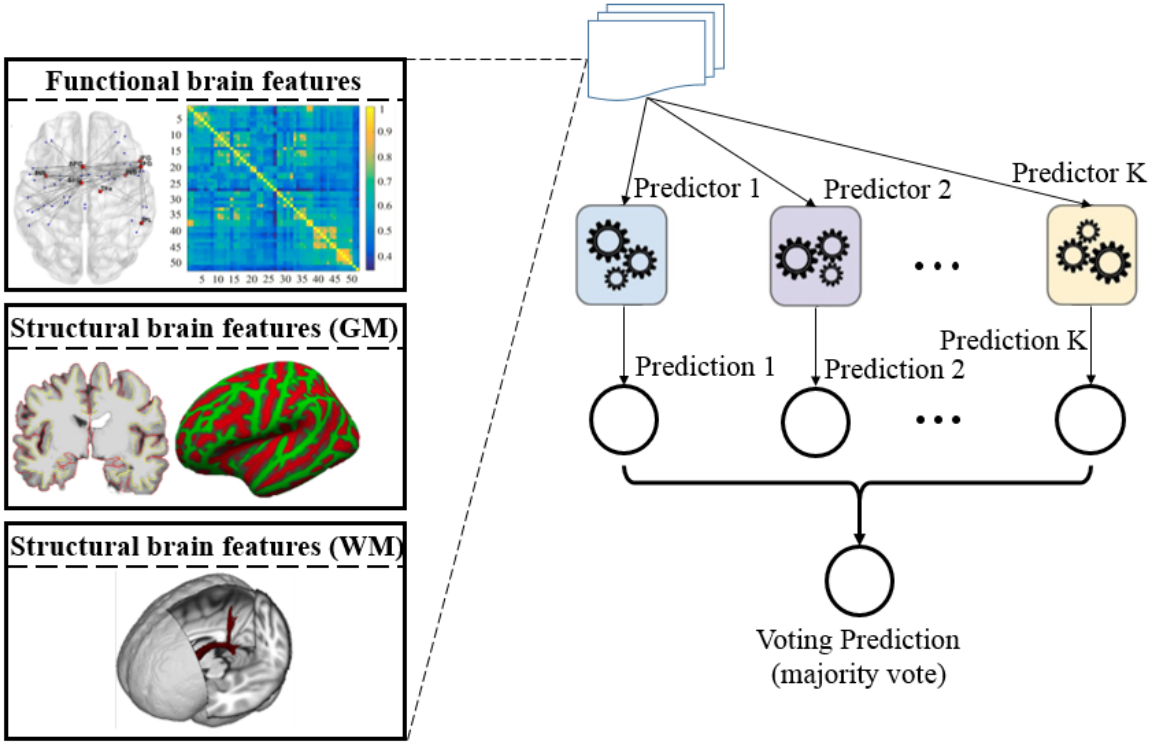

Figure 5.8 Workflow of voting.

5.1.3.2 Bagging. Bootstrap Aggregation (or Bagging for short), is a simple and very powerful ensemble method. An ensemble method is a technique that combines the predictions from multiple machine learning algorithms together to make more accurate

predictions than any individual model. Bootstrap Aggregation is a general procedure that can be used to reduce the variance for those algorithm that have high variance. An algorithm that has high variance are decision trees, like classification and regression trees (CART). Decision trees are sensitive to the specific data on which they are trained. If the training data is changed (e.g. a tree is trained on a subset of the training data) the resulting decision tree can be quite different and in turn the predictions can be quite different. Bagging is the application of the Bootstrap procedure to a high-variance machine learning algorithm, typically decision trees.

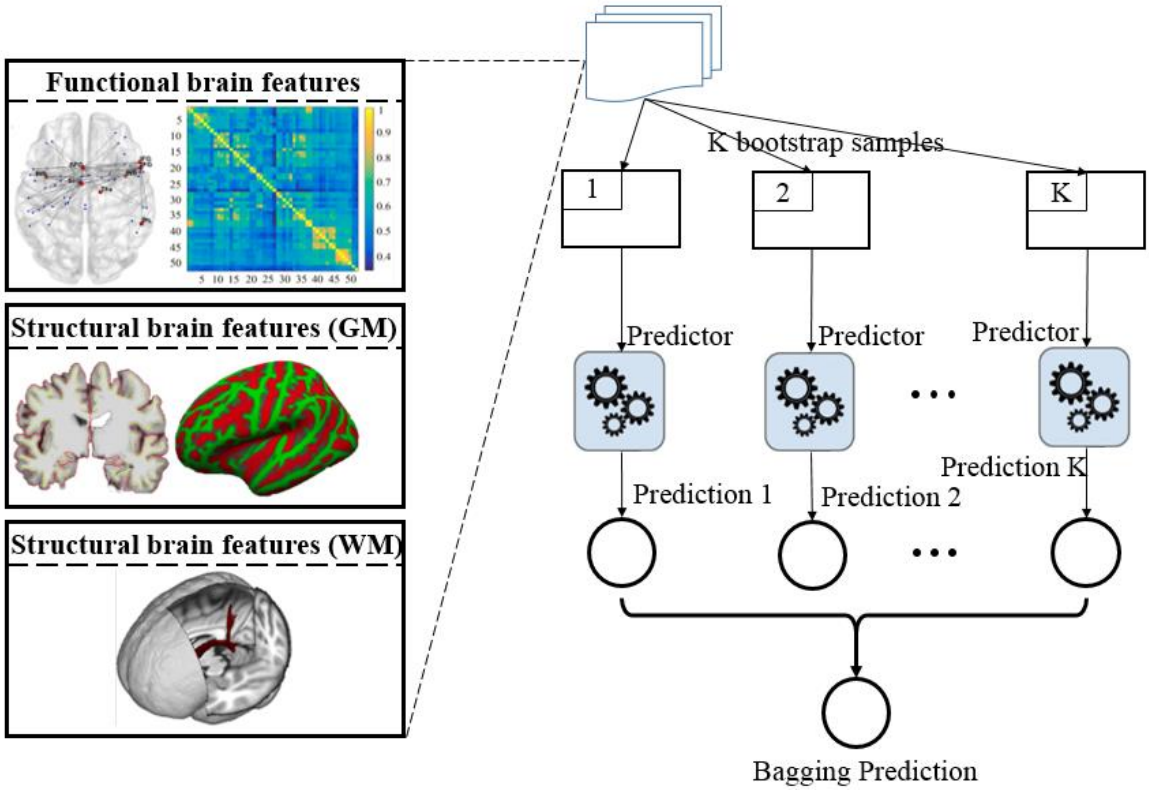

Figure 5.9 Workflow of bagging.

5.1.3.3 Boosting. Boosting is a kind of algorithm that trains a bunch of individual models in a sequential way. Each individual model learns from mistakes made by the previous model. AdaBoost is a boosting ensemble model and works especially well with the decision tree. Boosting model's key is learning from the previous mistakes, e.g. misclassification data points increasing the weight of misclassified data points. The core principle of AdaBoost is to fit a sequence of weak learners (i.e., models that are only slightly better than random guessing, such as small decision trees) on repeatedly modified versions of the data. The predictions from all of them are then combined through a weighted majority vote (or sum) to produce the final prediction. The data modifications at each so-called boosting iteration consist of applying weights $\boldsymbol{\omega_1}$, $\boldsymbol{\omega_2}$,…, $\boldsymbol{\omega_N}$ to each of the training samples. Initially, those weights are all set to $\boldsymbol{\omega_i = 1/N}$, so that the first step simply trains a weak learner on the original data. For each successive iteration, the sample weights are individually modified and the learning algorithm is reapplied to the reweighted data. At a given step, those training examples that were incorrectly predicted by the boosted model induced at the previous step have their weights increased, whereas the weights are decreased for those that were predicted correctly. As iterations proceed, examples that are difficult to predict receive ever-increasing influence. Each subsequent weak learner is thereby forced to concentrate on the examples that are missed by the previous ones in the sequence.

The AdaBoost makes a new prediction by adding up the weight (of each tree) multiply the prediction (of each tree). Obviously, the tree with higher weight will have more power of influence the final decision.

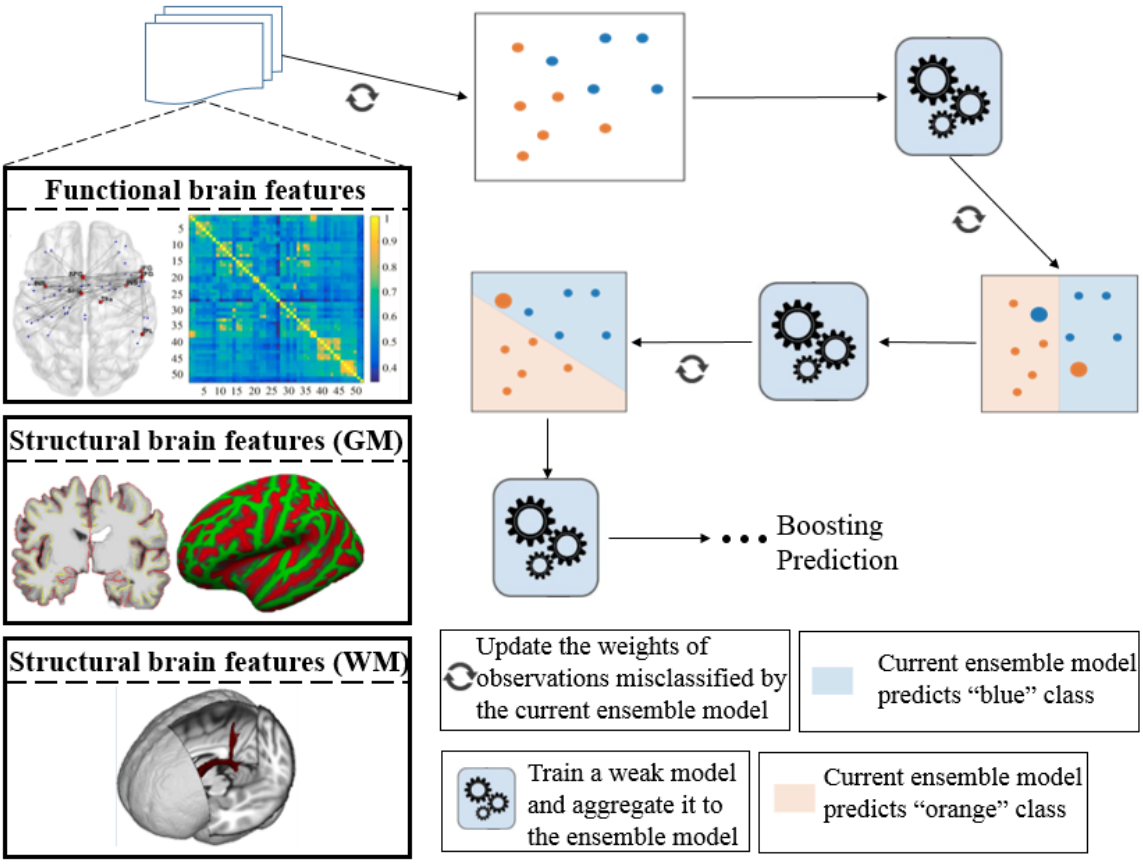

Figure 5.10 Workflow of boosting.

5.1.3.4 Stacking. Stacking mainly differ from bagging and boosting on two points. First stacking often considers heterogeneous weak learners (different learning algorithms are combined) whereas bagging and boosting consider mainly homogeneous weak learners. Second, stacking learns to combine the base models using a meta-model whereas bagging

and boosting combine weak learners following deterministic algorithms. The idea of stacking is to learn several different weak learners and combine them by training a meta-model to output predictions based on the multiple predictions returned by these weak models. Thus, to build stacking model, two things need to be defined: The L learners utilized to fit and the meta-model that combines them. The algorithm of stacking is as followed:

1) Split the training data in two folds.

2) Choose L weak learners and fit them to data of the first fold.

3) For each of L weak learners, make predictions for observations in the second fold.

4) Using predictions from weak learners as inputs to fit the meta-model on the second fold.

In the previous steps, the dataset was split in two folds because predictions on data that have been used for the training of the weak learners are not relevant for the training of the meta-model**.** Thus, an obvious drawback of this split in two parts is that half of the data is used to train the base models and half of the data is utilized to train the meta-model. In order to overcome this limitation, some kind of "k-fold cross-training" approach such that all the observations can be used to train the meta-model: for any observation, the prediction of the weak learners are done with instances of these weak learners trained on the k-1 folds that do not contain the considered observation can be followed. In other words, it consists in training on k-1 fold in order to make predictions on the remaining fold and that iteratively so that to obtain predictions for observations in any folds. Doing so, relevant predictions for

each observation of the dataset can be produced and then train meta-model on all these predictions.

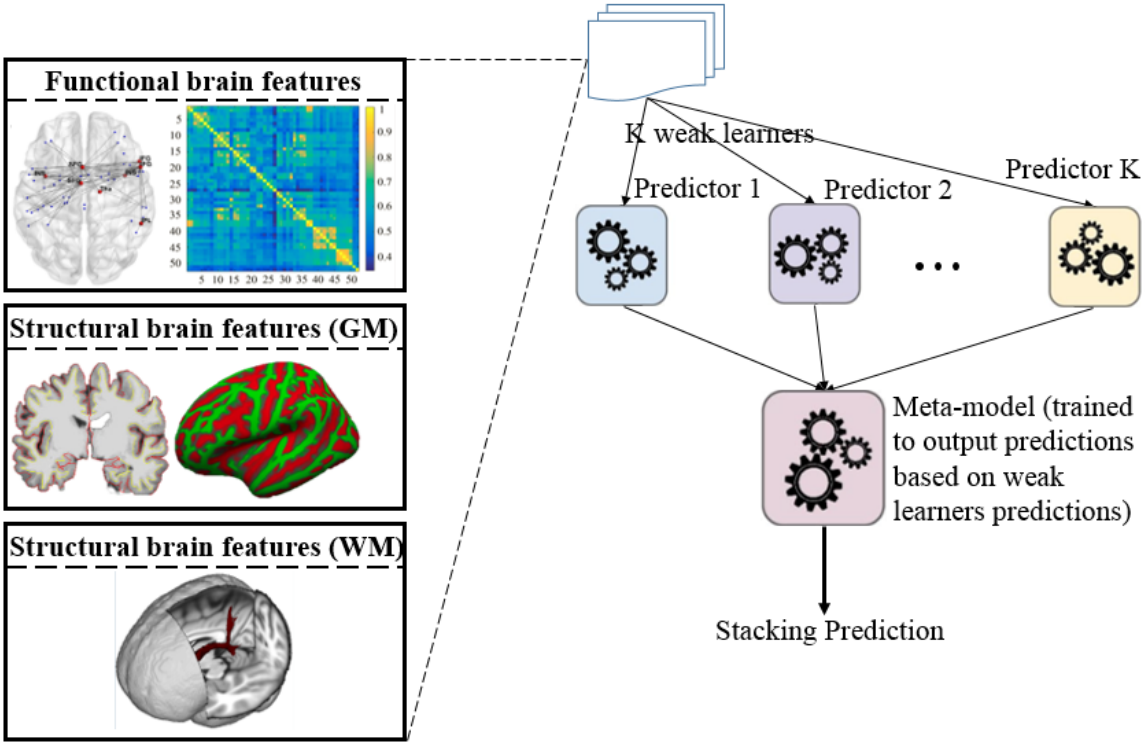

Figure 5.11 Workflow of stacking.

5.1.4 Introduction to Regression Models

Regression analysis is a statistical technique that models and approximates the relationship between a dependent and one or more independent variables. Four commonly used regression models include ordinary least squares (OLS), ridge, least absolute shrinkage and selection operator (LASSO), and elastic net regression.

5.1.4.1 Ordinary Least Squares In statistics, OLS is a type of linear least squares method for estimating the unknown parameters in a linear regression model. OLS chooses the parameters of a linear function of a set of explanatory variables by the principle of least squares: minimizing the sum of the squares of the differences between the observed dependent variable (values of the variable being observed) in the given dataset and those predicted by the linear function.

The equation for this model is referred to as the cost function and is a way to find the optimal error by minimizing and measuring it. The gradient descent algorithm is used to find the optimal cost function by going over a number of iterations. Let's kick off with the basics: the simple linear regression model, in which you aim at predicting n observations of the response variable, Y, with a linear combination of m predictor variables, X, and a normally distributed error term with variance σ^2:

$$Y = X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2) \tag{5.2}$$

As we don't know the true parameters, β, we have to estimate them from the sample. In the OLS approach, we estimate them as $\hat{\beta}$ in such a way, that the sum of squares of residuals is as small as possible. In other words, we minimize the following loss function:

$$L_{OLS}(\hat{\beta}) = \sum_{i=1}^{n} (y_i - x_i{}'\hat{\beta})^2 = \left|\left|y - X\hat{\beta}\right|\right|^2 \tag{5.3}$$

in order to obtain the infamous OLS parameter estimates, $\hat{\beta}_{OLS} = (X'X)^{-1}(X'Y)$. One situation is the data showing multi-collinearity, this is when predictor variables are correlated to each other and to the response variable. The high correlation of two variables can inflate the standard error of their coefficients which may make them seem statistically insignificant.

To produce a more accurate model of complex data we can add a penalty term to the OLS equation. A penalty adds a bias towards certain values. These are known as L1 regularization (LASSO regression) and L2 regularization (ridge regression). The best model we can hope to come up with minimizes both the bias and the variance:

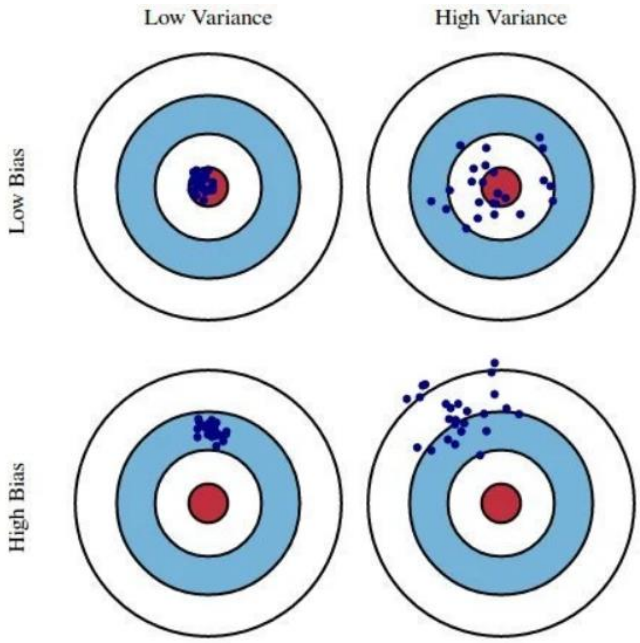

Figure 5.12 Variance/bias trade off.
Source: https://hackernoon.com/an-introduction-to-ridge-lasso-and-elastic-net-regression-cca60b4b934f. Accessed on April. 10, 2020.

5.1.4.2 Ridge Regression. Ridge regression uses L2 regularization which adds the following penalty term to the OLS equation.

$$L_{ridge}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{m} \hat{\beta}_j{}^2 = \left|\left|y - X\hat{\beta}\right|\right|^2 + \lambda\left|\left|\hat{\beta}\right|\right|^2 \quad (5.4)$$

Solving this for $\hat{\beta}$ gives the ridge regression estimates $\hat{\beta}_{ridge} = (X'X + \lambda I)^{-1}(X'Y)$, where I denotes the identity matrix. The L2 term is equal to the square of the magnitude of the coefficients. In this case if λ is zero then the equation is the basic OLS but if it is greater than zero then we add a constraint to the coefficients. This constraint results in minimized coefficients (aka shrinkage) that trend towards zero the larger the value of lambda. Shrinking the coefficients leads to a lower variance and in turn a lower error value. Therefore, ridge regression decreases the complexity of a model but does not reduce the number of variables, it rather just shrinks their effect.

5.1.4.3 LASSO Regression. LASSO regression uses the L1 penalty term and stands for Least Absolute Shrinkage and Selection Operator. The penalty applied for L2 is equal to the absolute value of the magnitude of the coefficients:

$$L_{lasso}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{m} |\hat{\beta}_j| \quad (5.5)$$

Similar to ridge regression, a lambda value of zero spits out the basic OLS equation, however given a suitable lambda value LASSO regression can drive some coefficients to zero. The larger the value of lambda the more features are shrunk to zero. This can eliminate some features entirely and give us a subset of predictors that helps mitigate multi-collinearity and model complexity. Predictors not shrunk towards zero signify that they are important and thus L1 regularization allows for feature selection (sparse selection).

5.1.4.4 Elastic Net Regression. A third commonly used model of regression is the Elastic Net which incorporates penalties from both L1 and L2 regularization:

$$L_{elasticnet}(\hat{\beta}) = \frac{\sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2}{2n} + \lambda(\frac{1-\alpha}{2}\sum_{j=1}^{m}\hat{\beta}_j{}^2 + \alpha\sum_{j=1}^{m}|\hat{\beta}_j|) \quad (5.6)$$

In addition to setting and choosing a lambda value elastic net also allows us to tune the alpha parameter where $\alpha = 0$ corresponds to ridge and $\alpha = 1$ to LASSO. Therefore, an alpha value between 0 and 1 could be chosen to optimize the elastic net. Effectively this will shrink some coefficients and set some to 0 for sparse selection.

5.1.5 Introduction to Cross Validation

Cross validation, sometimes called out-of-sample testing, is any of various similar model validation techniques for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. In a prediction problem, a model is usually given a dataset of known data on which training is run (training dataset), and a dataset of unknown data against which the model is tested (called the validation dataset or testing set). The goal of cross-validation is to test the model's ability to predict new data that was not used in estimating it, in order to flag problems like overfitting or selection bias and to give an insight on how the model will generalize to an independent dataset (i.e., an unknown dataset, for instance from a real problem). To reduce variability, in most methods multiple rounds of cross-validation are performed using different partitions, and the validation results are combined (e.g. averaged) over the rounds to give an estimate of the model's predictive performance.

In summary, cross-validation combines (averages) measures of fitness in prediction to derive a more accurate estimate of model prediction performance. K-fold cross validation and leave-one-out cross validation are most commonly utilized cross validation techniques.

In K-fold cross validation, the original sample is randomly partitioned into K equal sized subsamples. Of the K subsamples, a single subsample is retained as the validation

data for testing the model, and the remaining K-1 subsamples are used as training data (Figure 5.13). The cross validation process is then repeated K times, with each of the K subsamples used exactly once as the validation data. The K results can then be averaged to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once. Five-fold cross validation is commonly used, but in general K remains an unfixed parameter. When K equals to the number of data in the set, leave-one-out cross validation is treated as a particular case of K-fold cross validation (Figure 5.14).

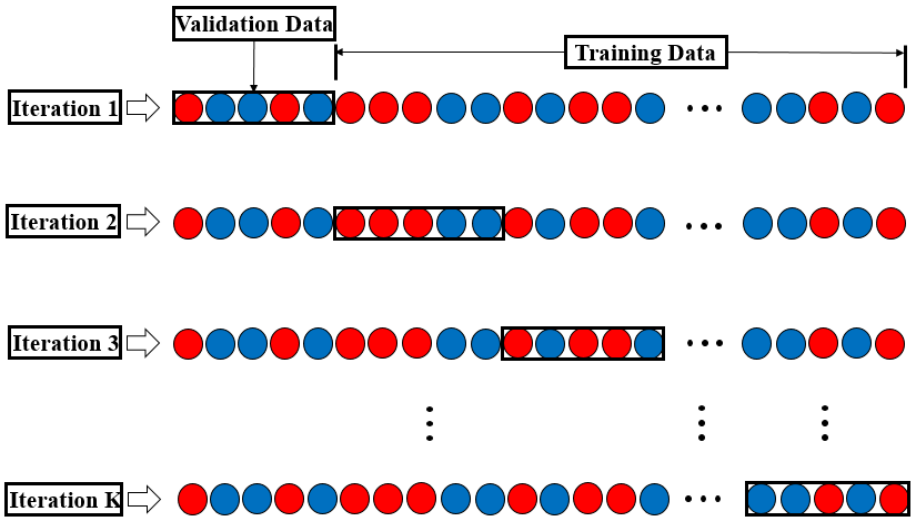

Figure 5.13 Workflow of k-fold cross validation.

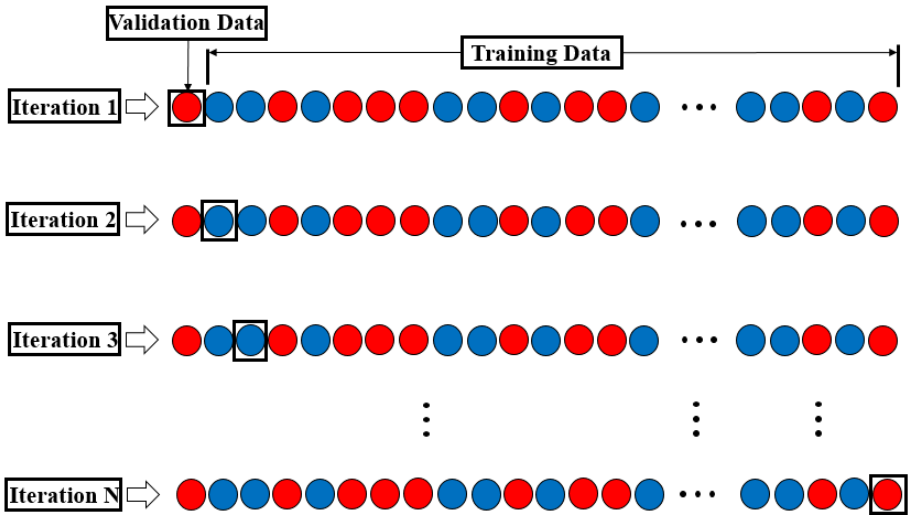

Figure 5.14 Workflow of leave-one-out cross validation.

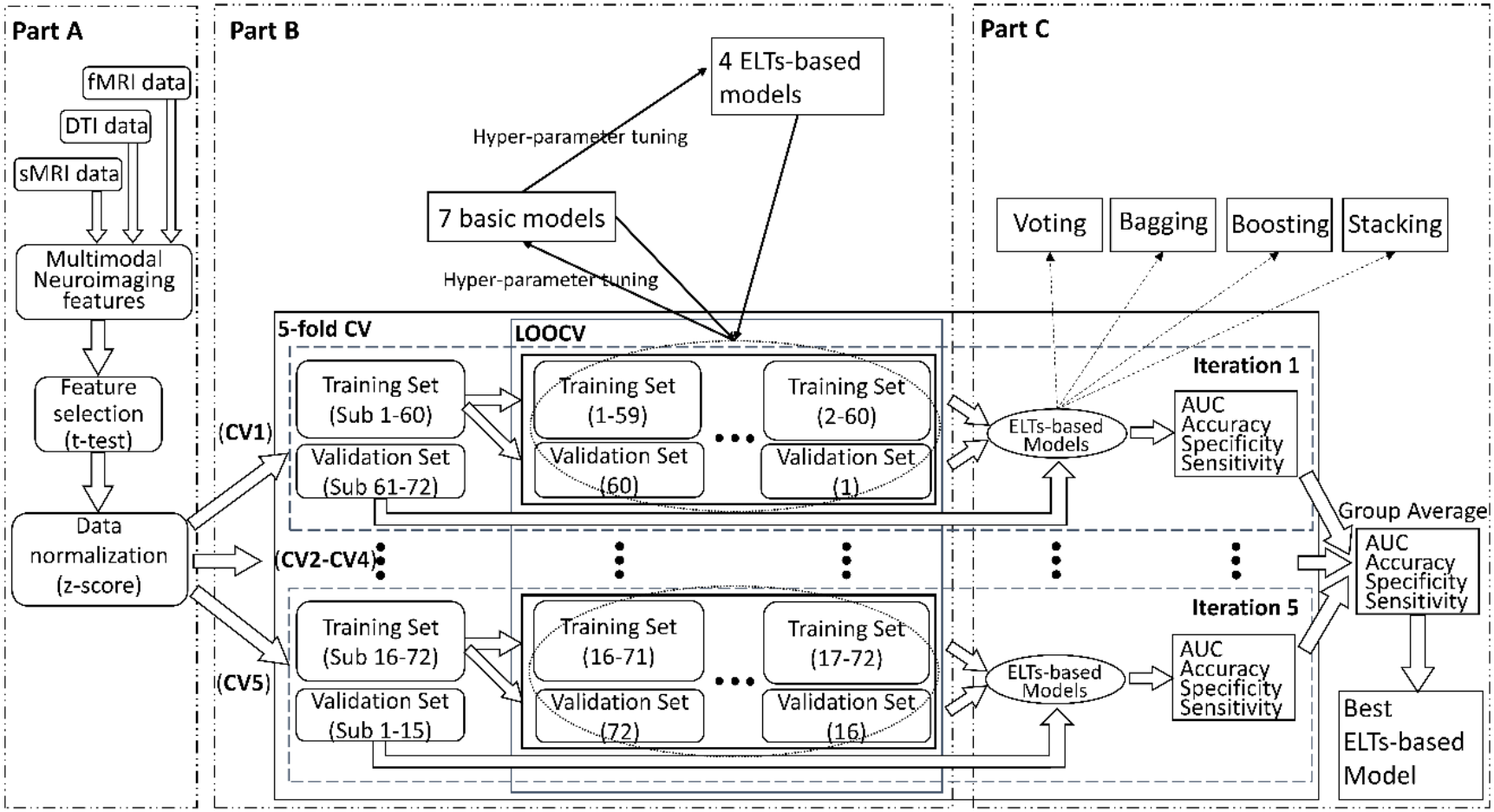

Figure 5.15 Ensemble learning flow chart for discrimination between ADHD and normal controls.

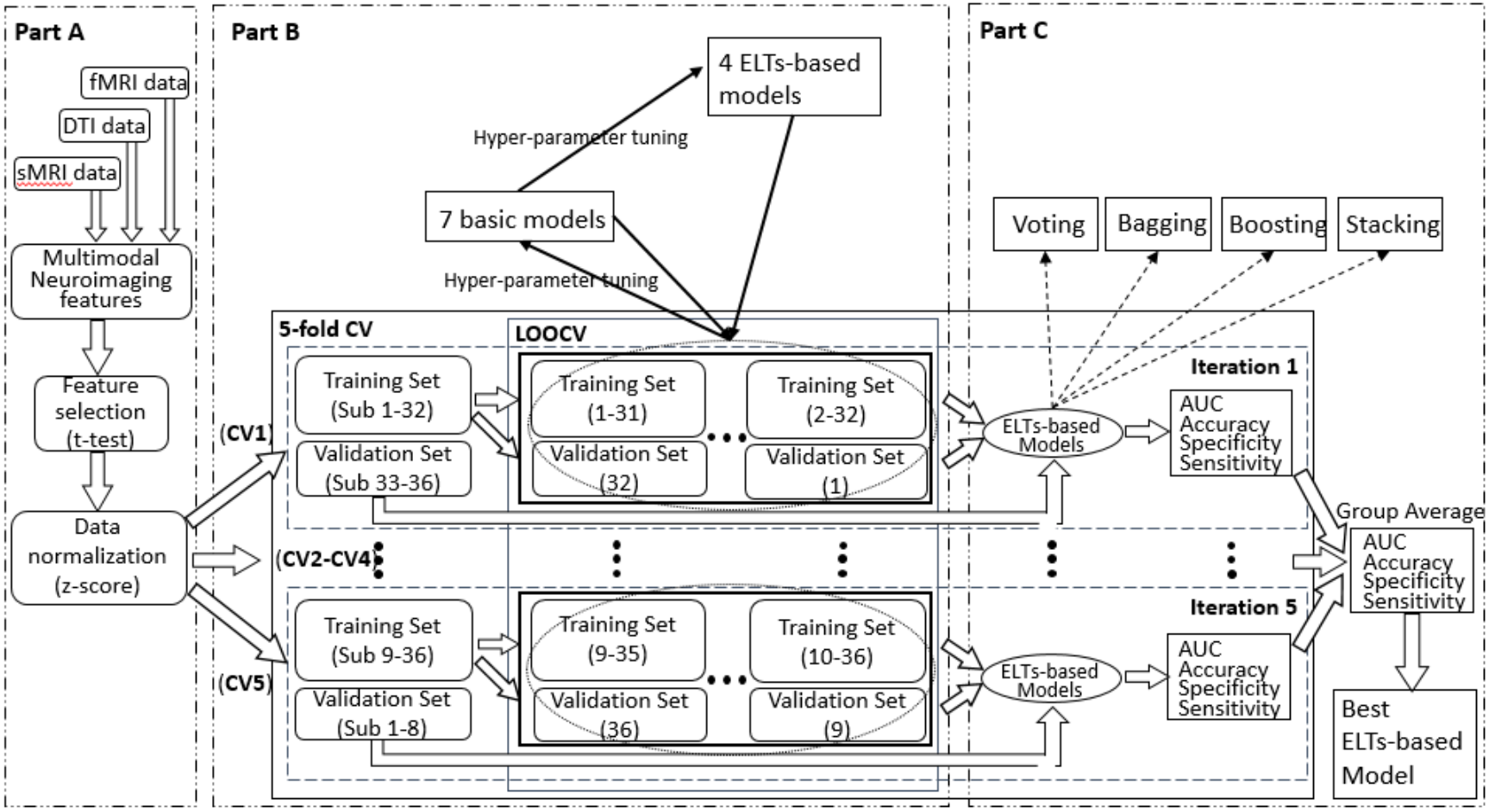

Figure 5.16 Ensemble learning flow chart for discrimination between ADHD persisters and ADHD remitters.

5.2 Experimental Strategy

5.2.1 Participants

Seventy-two young adults [mean (SD) age 24.4 (2.1) years] who provided good quality data from multimodal neuroimaging and clinical assessments, participated in this study. There were 36 ADHD probands diagnosed with ADHD-C in childhood and 36 group-matched comparison subjects with no history of ADHD. Among the 36 ADHD probands, 18 were classified as ADHD-R and the other 18 probands were classified as ADHD-P.

5.2.2 Multimodal Imaging Data Processing for Feature Extractions

For each subject, the T1-weighted data was reconstructed into a 3-dimensional cortical model for thickness and area estimations using FreeSurfer v.5.3.0 (https://surfer.nmr.mgh.harvard.edu). Each volume was first registered to the Talairach atlas. Intensity variations caused by magnetic field in homogeneities were corrected and non-brain tissue was removed. A cutting plane was used to separate the left and right hemispheres and to remove the cerebellum and brainstem. Two mess surfaces (mess of grids created using surface tessellation technique) were generated between GM and WM (WM surface), as well as between GM and cerebrospinal fluid (pial surface). The distance between the two closest vertices of the WM and pial surfaces represented the cortical thickness at that specific location. Cortical subregions were parcellated based on the

Desikan atlas. A total of 202 structural MRI features, including regional cortical GM thickness, surface area, and GM volume of subcortical structures were extracted from each subject.

The DTI data was corrected for eddy current-induced distortions due to the changing gradient field directions. Head motion was corrected with non-diffusion-weighted reference image (b0 image) using an affine, 12 degrees of freedom registration. Then the FA value and principle diffusion direction at each brain voxel were calculated. WM probabilistic tractography between each pair of 18 ROIs (bilateral thalami, putamen and caudate nuclei from striatum, hippocampus, and frontal, parietal, occipital, temporal, and insular cortices based on the Harvard-Oxford Cortical Atlases and Julich Histological Atlas) were constructed using the FSL/BEDPOSTX tool. The multi-fibre probabilistic connectivity-based method was applied to determine the number of pathways between the seed and each of the target clusters, with the default setting of parameters for the Markov Chain Monte Carlo estimation of the probabilistic tractography. At the end, a total of 120 DTI-based features, including the volume and FA of cortico-cortical and subcortico-cortical WM fiber tracts were extracted for each subject.

The fMRI data from each participant was preprocessed using Statistical Parametric Mapping version 8 (SPM8, Wellcome Trust Centre for Neuroimaging, London, United Kingdom; http://www.fil.ion.ucl.ac.uk/spm/) implemented on a MATLAB platform. The preprocessing procedures included slice timing correction, realignment, co-registration,

segmentation, normalization, and spatial smoothing. The first-level analyses were conducted using general linear model to generate the activation map responding to the cues. The group average activation maps for ADHD probands and controls were generated, respectively. A total of 52 cortical and subcortical seed regions, which was parceled according to the structural and functional connectivity-based Brainnetome atlas, were determined based on the results of the combination of the functional activation maps of the groups of ADHD probands and controls (Fan et al., 2016). To construct the cue-evoked attention processing network, the single-trial beta value series from the 48 cue-related events in the four runs were extracted. Among all the voxels in each of the 52 node ROIs, the average beta value series was calculated and used to create a 52×52 pair-wise Pearson correlation matrix. Then the GTTs were carried out. More details of the fMRI data processing can be found in (Luo et al., 2018). A total of 200 fMRI features, including the global- and local-efficiency of the entire network, the nodal efficiency, degree, and betweenness centrality measures of the 52 nodes, as well as their pair-wise functional connectivities, were generated for each subject.

5.2.3 Modeling of Ensemble Learning Architecture

Modeling of the ELTs for classifications between ADHD probands and controls, as well as between ADHD-P and ADHD-R, is described in Figures 5.15 and 5.16, respectively. Specifically, Part A of Figures 5.15 and 5.16 presents feature selection and preparation

flow. In order to decrease the risk of overfitting, two-sample t-tests were applied and a total of 20 neuroimaging features that showed the largest between group differences were first selected from the 522 multimodal neuroimaging features derived from structural MRI, DTI, and fMRI data. Then each value of the 20 selected features was normalized by using a z-score transformation in the feature-specific space. The normalized 20 top-ranked neuroimaging features were then entered to the training and validation procedures (Part B of Figures 5.15 and 5.16), which consisted of a nested CV (there were two CV loops, including an outer 5-fold CV loop to split the data into training set and validation set, and an inner loop to tune the hyperparameters for 7 basic models and 4 ELTs-based models using grid search in combination with LOOCV). More specifically, the 20 neuroimaging features were split into a total of 5 stratified folds such that each fold consisted of balanced 20% of the entire data. The five-fold CV was performed by using these 5 stratified folds, where each trial dedicated four folds for training data and the remaining one for validation. Then for each iteration in 5-fold CV, the corresponding training set was sent into the LOOCV processing. In each iteration, one subject was extracted from the training set to act as a validation data, and the remaining subjects were trained to construct the models. According to the classification performance of the validation data, the hyperparameters for each model were tuned and the optimal hyperparameters setup were selected using grid search. More details of the hyperparameters are described in Table 5.2. We utilized the LOOCV to tune the hyperparameters of 7 basic models, including KNN, SVM, LR, NB,

LDA, RF, and MLP. Based on the hyperparameters of basic models, we applied LOOCV to tune the hyperparameters of 4 ELTs-based models, including max Voting, Bagging, AdaBoost, Stacking. As shown in Part C of Figures 5.15 and 5.16, during iterations of 5-fold CV outer loop, the performance of each basic and ELTs-based models with the optimal hyperparameters derived from LOOCV inner loop iterations was evaluated. The group average of classification performance of each classifier derived from each iteration of 5-fold CV was generated. The 7 basic and 4 ELTs-based models were assessed according to the group average value of AUC of the ROC from iterations of 5-fold CV outer loop. The basic and ELTs-based models with the highest average AUC were selected as optimal classifiers. Based on the types of ELT-based models we evaluated and selected, the importance score corresponding each feature was then calculated using the ELT-based model and the corresponding basic models.

We also applied unsupervised learning (i.e., the hierarchical clustering) in our dataset. The hyperparameters, including the metric used to compute the linkage (affinity), the linkage criterion used to determine which distance between sets of observation (linkage) were also tuned by using grid search. Then the model with best classification performance (i.e., accuracy) was selected.

Table 5.1 Hyperparameters of 7 Basic Models and 4 ELTs-based Models

Classifiers	Hyperparameters
K-Nearest Neighbors	n_neighbors: [1, 3, 5, 7, 9]; algorithm: [‘auto’, ‘ball_tree’, ‘kd_tree’, ‘brute’]; p: [1, 2, 3]
Support Vector Machine	C: [0.001, 0.01, 0.1, 1, 10, 100, 1000]; gamma: [‘auto’, ‘scale’]; kernel: [‘linear’, ‘rbf’, ‘poly’, ‘sigmoid’]
Logistical Regression	solver: [‘newton-cg’, ‘lbfgs’, ‘sag’, ‘saga’]; multi_class: [‘ovr’, ‘multinomial’, ‘auto’]
Naïve Bayes	N/A
Random Forest	n_estimators: list(range(3, 60, 5)); criterion: [‘gini’, ‘entropy’]; min_samples_leaf: [3, 5, 10]; max_depth: [3, 4, 5, 6]; min_samples_split: [3, 5, 10]; bootstrap: [True, False]
Linear Discriminant Analysis	solver: [‘svd’, ‘lsqr’, ‘eigen’]
Multilayer Perceptron	activation: [‘identity’, ‘logistic’, ‘tanh’, ‘relu’]; solver: [‘lbfgs’, ‘sgd’, ‘adam’]; hidden_layer_sizes: np.arange(1, 72, 10); max_iter: [4000]
Ensemble Learning Technique-Voting	estimators; voting: [‘hard’, ‘soft’]
Ensemble Learning Technique-Bagging	base_estimator; n_estimators: list(range(10, 150, 10)); max_samples=[0.2, 0.3, 0.4, 0.5]; max_features=[0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
Ensemble Learning Technique-Boosting	base_estimator; n_estimators: list(range(10, 150, 10)); learning_rate: list(range(0.01, 1, 0.01))
Ensemble Learning Technique-Stacking	classifiers; meta_classifiers

5.2.4 Regression Models

Following the classification procedures, we constructed the regression models to identify the relations between the neuroimaging features and the clinical inattentive and hyperactive/impulsive symptom T-scores. Based on the ELT-based classification results, the top three neuroimaging features were selected based on the weight of each feature in the optimal discriminators between ADHD and NCs, as well as between ADHD-P and ADHD-R. Then, we applied OLS, ridge regression (Hoerl and Kennard, 1970), LASSO regression (Santosa and Symes, 1986; Tibshirani, 1996), elastic net regression (Zou and Hastie, 2005) to construct the prediction models for inattentive and hyperactive/impulsive T-scores, respectively. The same nested CV utilized in previous steps were also conducted in regression model construction. The hyperparameters included the regularization strength (alpha), solver to use in the computational routines (solver) for ridge regression, the constant that multiplies the L1 term (alpha) for LASSO regression, the constant that multiplies the penalty terms (alpha), the elastic net mixing parameter (l1_ratio) for elastic net regression. During the iteration of 5-fold CV outer loop, the performance of each regression model with the optimal hyperparameters derived from LOOCV inner loop iterations was evaluated. The group average of regression performance, including the Pearson correlation coefficient and mean squared error (MSE) between predicted and

observed values, of each regression model derived from each iteration of 5-fold CV were calculated.

5.2.5 Evaluation Measures

The performance of each classification procedure classifier was measured in terms of classification accuracy, sensitivity, and specificity. The accuracy of a machine learning classification algorithm is to measure how often the algorithm classifies a data point correctly. It is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.7}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Sensitivity describes the proportion of actual positive cases that are correctly identified as positive. It implies that there will be another proportion of actual positive cases, which would get predicted incorrectly as negative. The sensitivity is defined as:

$$Sensitivity\ (Recall) = \frac{TP}{TP + FN} \tag{5.8}$$

Specificity is a measure of the proportion of actual negatives, which got predicted as the negative. It implies that there will be another proportion of actual negative, which got predicted as positive and could be termed as false positives. It is defined as:

$$Specificity = \frac{TN}{TN + FP} \tag{5.9}$$

In addition, a ROC curve was plotted to illustrate the diagnostic ability of a binary classifier system as its discrimination threshold is varied. In the classification case, we calculated the confusion matrix for each iteration cycle of the classifier and calculated the AUC of ROC. AUC provides an aggregate measure of performance across all possible classification thresholds. One way of interpreting AUC is as the probability that the model ranks a random positive example more highly than a random negative example. The AUC of ROC is defined as:

$$AUC = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5.10}$$

Among the equation of AUC, Precision and Recall are defined, respectively, as:

$$Precision = \frac{TP}{TP + FP} \tag{5.11}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.12}$$

For the regression model, the Pearson correlation coefficient and MSE between predicted values and actual values were calculated. The Pearson correlation coefficient is a measure of the linear correlation between two variables. It is defined as:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{5.13}$$

where cov is the covariance, σ_X is the standard deviation of X, σ_Y is the standard deviation of Y.

The MSE of an estimator measures the average squared difference between the estimated values and the actual value, which is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{5.14}$$

Where Y_i and $\hat{Y}_i$ represent the actual and predicted value.

5.3 Results

5.3.1 Demographic, Clinical and Behavioral Measures

The demographic, clinical information and behavioral performance of all groups are summarized in Table 5.1. There were no significant demographic between-group differences. Moreover, all participants achieved a >85% rate for response accuracy when performing the fMRI task. Task performance measures, including reaction time, response accuracy rate, omission error rate, commission error rate did not show between-group differences (p>0.05).

Table 5.2 Demographic and Clinical Characteristics in Groups of Controls and ADHD Probands (and Further in the Sub-groups of Remitters and Persisters of the ADHD Probands)

	Controls (N=36)	**Probands (N=36)**		**ADHD-R (N=18)**	**ADHD-P (N=18)**	
	Mean (SD)	**Mean (SD)**	***p***	**Mean (SD)**	**Mean (SD)**	***p***
Age	24.3 (2.3)	24.66 (2.0)	*0.48*	24.79 (2.2)	24.52 (2.0)	*0.7*
Full-scale IQ	103.83(15.4)	97.96 (14.1)	*0.1*	99.22 (14.9)	96.71 (13.6)	*0.6*
CAARS (T-score)						
Inattentive	45.75 (8.8)	56.5 (13.2)	*<0.001*	49.83 (10.9)	63.17 (12.0)	*0.001*
Hyperactive/Impulsive	42.97 (6.2)	53.64 (12.9)	*<0.001*	46.17 (9.0)	61.11 (12.0)	*<0.001*
ADHD Total	43.89 (8.2)	56.5 (14.7)	*<0.001*	42.61 (7.5)	54.33 (8.8)	*<0.001*
ADHD semistructured interview (number of symptoms)	0.79 (1.6)	6.17 (5.2)	*<0.001*	2.64 (2.0)	10.24 (3.6)	*<0.01*
	N (%)	**N (%)**	***p***	**N (%)**	**N (%)**	***p***
Male	31 (86.1)	30 (83.3)	*0.74*	16 (88.9)	14 (77.8)	*0.37*
Right-handed	32 (88.9)	32 (88.9)	*1*	15 (83.3)	16 (88.9)	*0.63*
Race			*0.17*			*0.59*
Caucasian	15 (41.7)	21 (58.3)		9 (50.0)	12 (66.7)	
African American	13 (36.1)	7 (19.4)		4 (22.2)	3 (16.7)	
More than one race	6 (16.7)	8 (22.2)		5 (27.8)	3 (16.7)	
Asian	2 (5.6)	0 (0.0)		0 (0.0)	0 (0.0)	
Ethnicity			*0.09*			*0.74*
Hispanic/Latino	10 (27.8)	17 (47.2)		8 (44.4)	9 (50.0)	
Task performance measures	**Mean (SD)**	**Mean (SD)**	***p***	**Mean (SD)**	**Mean (SD)**	***p***
Reaction time average	395.8 (53.1)	422.8 (74.3)	*0.08*	431.1 (67.0)	439.1 (107.8)	*0.79*
Reaction time std	129.6 (24.8)	137.2 (29.9)	*0.25*	136.2 (27.6)	138.2 (32.8)	*0.84*
Anticipation error	1.86 (2.1)	1.74 (1.6)	*0.78*	1.69 (1.6)	1.78 (1.7)	*0.88*
Commission error	0.33 (0.8)	0.85 (1.4)	*0.07*	0.75 (1.6)	0.94 (1.3)	*0.7*
Omission error	4.97 (5.8)	8 (10.8)	*0.15*	4.38 (4.0)	11.22 (13.8)	*0.06*

5.3.2 Classification Model Performance

Tables 5.3 and 5.4 summarize the seven basic models classification performance between ADHD probands vs. controls, and between ADHD-P vs. ADHD-R, respectively. The classifier of SVM performed the best among the seven basic models regarding the AUC, accuracy, and specificity (AUC=0.87, accuracy=0.816, specificity=0.942) for the discrimination between ADHD probands and NCs; and again SVM performed the best among all the basic models regarding the AUC and accuracy (AUC=0.85, accuracy=0.7) for the classification between ADHD-P and ADHD-R.

Table 5.3 Results of Seven Basic Classifications Between the Groups of ADHD and Normal Controls

Classifiers	Specificity	Sensitivity	Accuracy	AUC
KNN	0.72	0.66	0.689	0.69
SVM	0.942	0.69	0.816	0.87
LR	0.756	0.742	0.75	0.85
NB	0.778	0.718	0.748	0.86
RF	0.866	0.75	0.705	0.82
LDA	0.734	0.774	0.754	0.78
MLP	0.782	0.746	0.764	0.84

Table 5.4 Results of Seven Basic Classifications Between the Groups of ADHD Persisters and ADHD Remitters

Classifiers	Specificity	Sensitivity	Accuracy	AUC
KNN	0.4	0.934	0.667	0.72
SVM	0.65	0.75	0.7	0.85
LR	0.6	0.682	0.642	0.85
NB	0.734	0.65	0.692	0.77
RF	0.734	0.6	0.667	0.76
LDA	0.568	0.518	0.542	0.63
MLP	0.634	0.75	0.692	0.84

Table 5.5 summarizes the ADHD probands vs. controls classification performances of the ELTs. Additional details of ADHD probands vs. controls classification performance of ELTs are shown in Table 5.6. The bagging-based ELT with SVM as the basic model performed the best among all ensemble models (AUC=0.89). Tables 5.7 and 5.8 show ADHD-P vs. ADHD-R classification performances of ELTs with diverse basic models, and again demonstrated that the bagging-based ELT with SVM as the basic model performed the best among all ensemble models (AUC=0.9).

Table 5.5 Results of Four ELTs-based Classifications Between the Groups of ADHD and Normal Controls

Classifiers	Specificity	Sensitivity	Accuracy	AUC
ELT-Voting	0.808	0.718	0.763	0.87
ELT-Bagging	0.734	0.798	0.766	0.89
ELT-Boosting	0.67	0.77	0.721	0.88
ELT-Stacking	0.756	0.742	0.75	0.82

Table 5.6 Details of Four ELTs-based Classifications Between the Groups of ADHD and Normal Controls

Classifiers	Specificity	Sensitivity	Accuracy	AUC
ELT-Voting				
SVM, NB, LR	0.808	0.718	0.763	0.87
ELT-Bagging				
KNN	0.894	0.66	0.777	0.85
SVM	0.734	0.798	0.766	0.89
LR	0.806	0.718	0.763	0.87
NB	0.832	0.746	0.789	0.84
RF	0.862	0.746	0.804	0.88
LDA	0.836	0.748	0.791	0.82
MLP	0.834	0.718	0.777	0.86
ELT-Boosting				
KNN	N/A	N/A	N/A	N/A
SVM	0.67	0.77	0.721	0.88
LR	0.862	0.718	0.789	0.87
NB	0.81	0.72	0.763	0.82
RF	0.862	0.634	0.746	0.83
LDA	N/A	N/A	N/A	N/A
MLP	N/A	N/A	N/A	N/A
ELT-Stacking				
First-level: SVM, LR; Second-level: LR	0.756	0.742	0.75	0.82

Table 5.7 Results of Four ELTs-based Classifications Between the Groups of ADHD Persisters and ADHD Remitters

Classifiers	Specificity	Sensitivity	Accuracy	AUC
ELT-Voting	0.8	0.65	0.725	0.82
ELT-Bagging	0.75	0.582	0.67	0.90
ELT-Boosting	0.75	0.682	0.717	0.86
ELT-Stacking	0.884	0.684	0.783	0.82

Table 5.8 Details of Four ELTs-based Classifications Between the Groups of ADHD Persisters and ADHD Remitters

Classifiers	Specificity	Sensitivity	Accuracy	AUC
ELT-Voting				
SVM, NB, LR	0.8	0.65	0.725	0.82
ELT-Bagging				
KNN	0.368	0.934	0.65	0.78
SVM	0.75	0.582	0.67	0.9
LR	0.55	0.616	0.65	0.86
NB	0.95	0.568	0.758	0.85
RF	0.784	0.684	0.73	0.82
LDA	0.7	0.95	0.825	0.85
MLP	0.634	0.782	0.67	0.86
ELT-Boosting				
KNN	N/A	N/A	N/A	N/A
SVM	0.75	0.682	0.717	0.86
LR	0.684	0.616	0.65	0.79
NB	0.734	0.668	0.7	0.78
RF	0.784	0.7	0.742	0.77
LDA	N/A	N/A	N/A	N/A
MLP	N/A	N/A	N/A	N/A
ELT-Stacking				
First-level: SVM, RF; Second-level: LR	0.884	0.684	0.783	0.82

5.3.3 Receiver Operating Characteristic Curves of Classification Models

The ROC curve for each basic classification procedure, including the unsupervised hierarchical clustering, for ADHD vs. NCs and ADHD-P vs. ADHD-R is plotted in Figures

5.17 and 5.18, respectively. In addition, The ROC curve for each ensemble learning classification procedure for ADHD vs. NCs and ADHD-P vs. ADHD-R is plotted in Figures 5.19 and 5.20, respectively. Results showed that classification performance parameters of the ELTs-based procedures were greatly improved compared to those of the basic model-based procedures. In addition, relative to the performance improvement between ensemble learning and basic models of the classification between ADHD and NCs, the performance improvement of classification between ADHD-P and ADHD-R is greater.

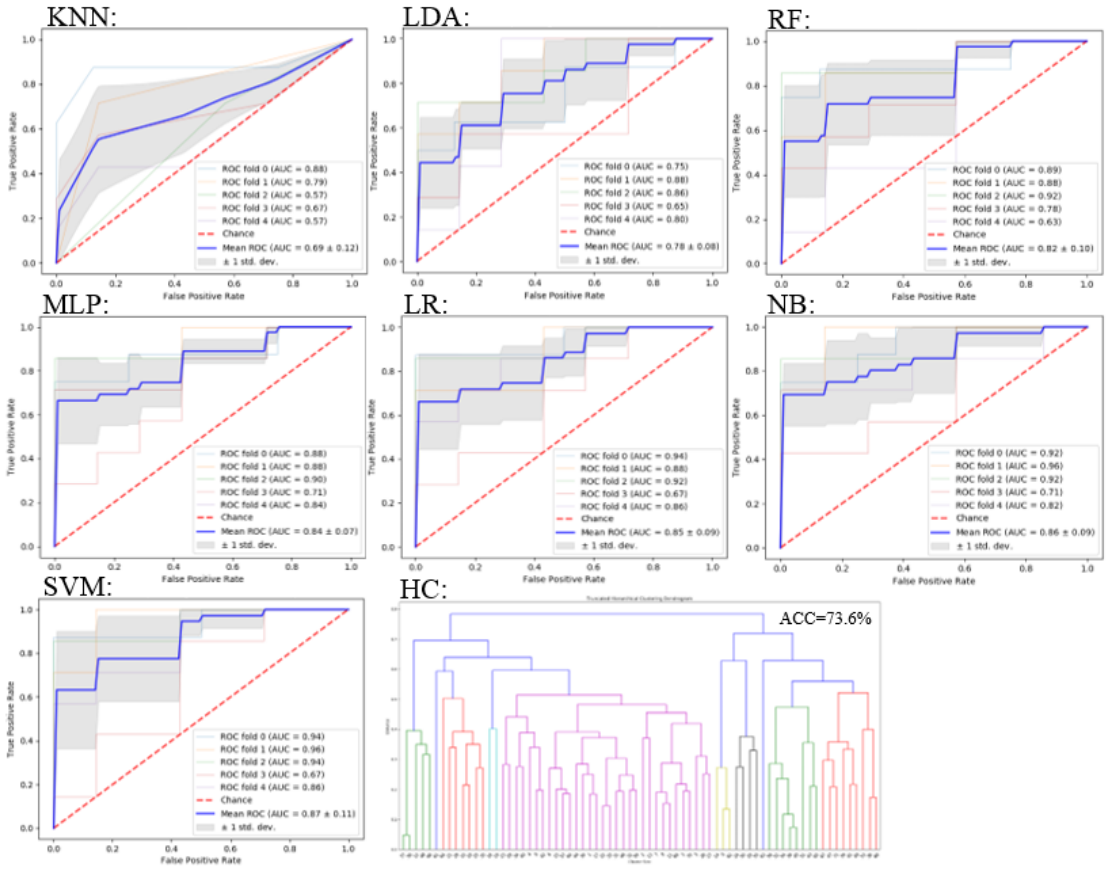

Figure 5.17 AUC of each basic classification procedure for discrimination between ADHD probands and normal controls.

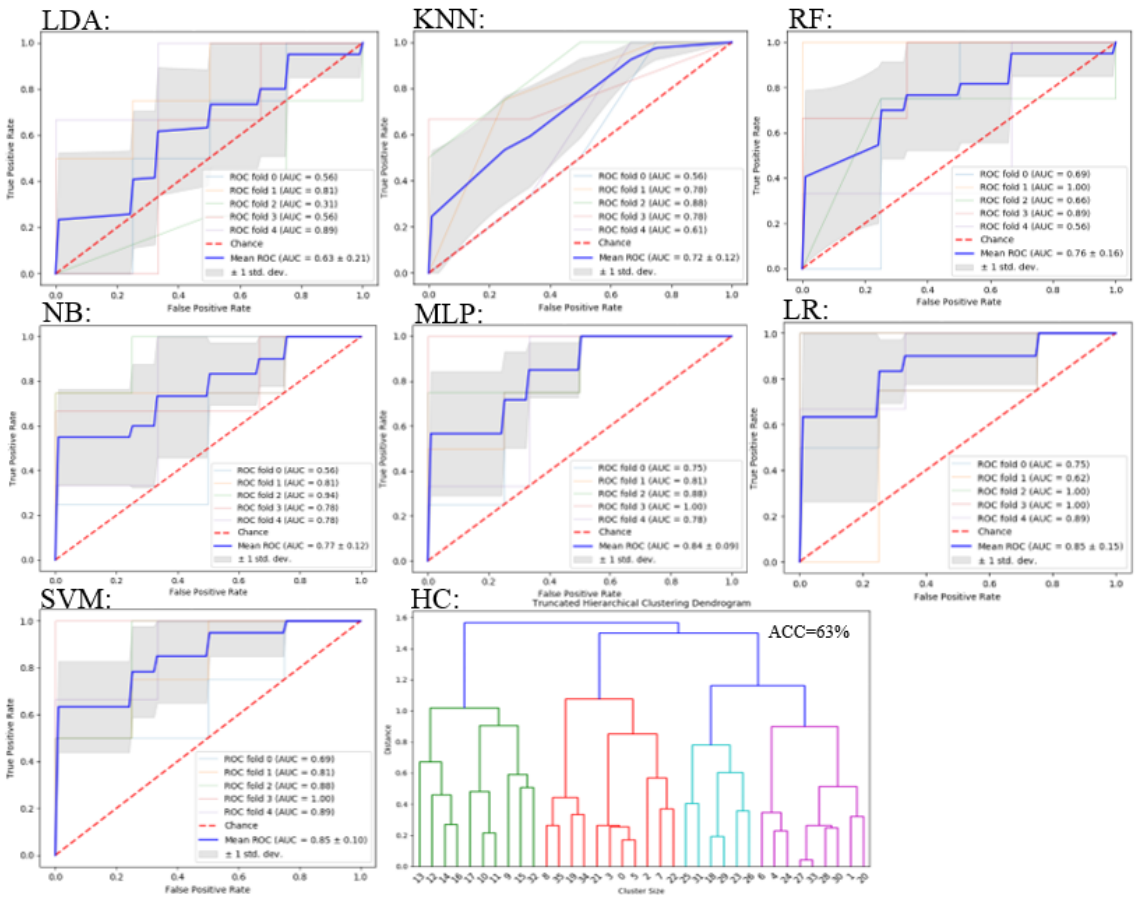

Figure 5.18 AUC of each basic classification procedure for discrimination between ADHD persisters and ADHD remitters.

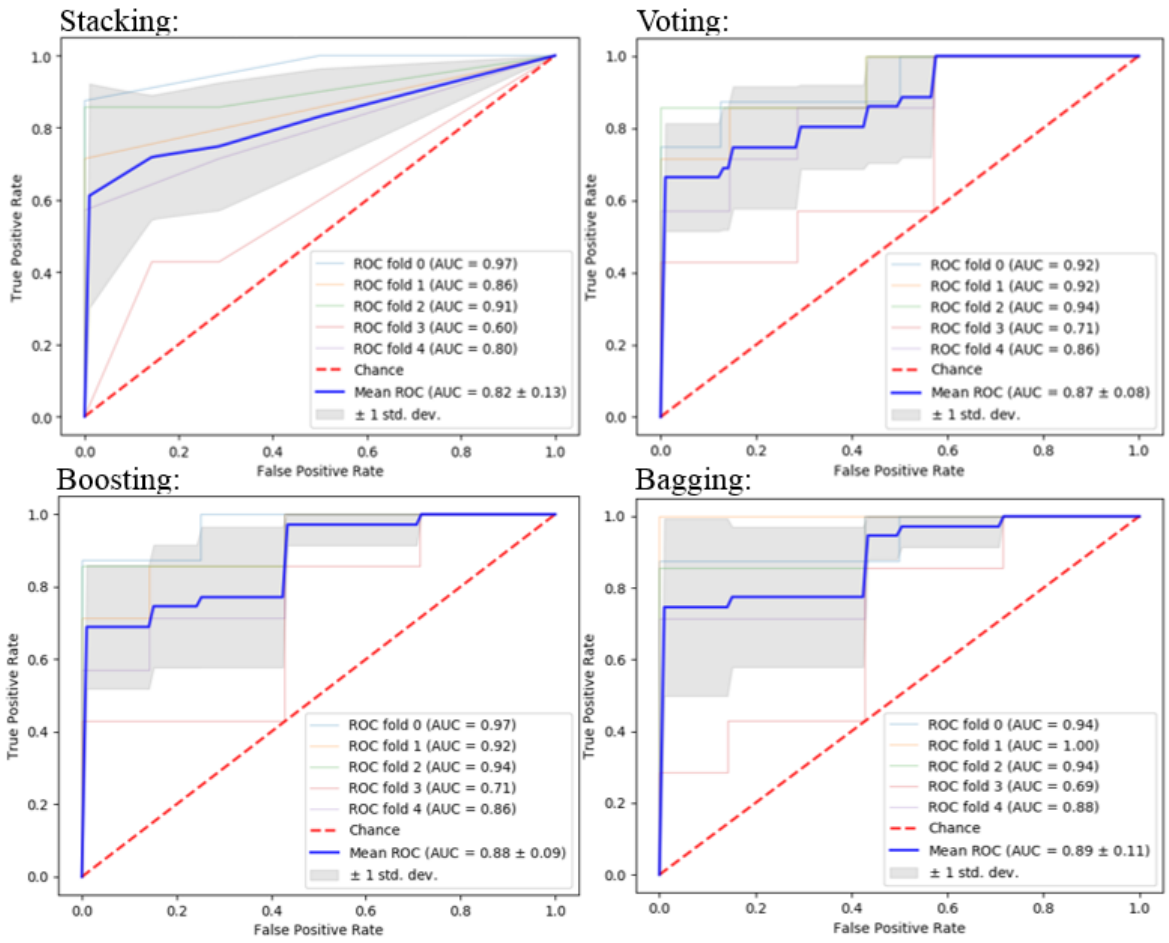

Figure 5.19 AUC of each ELT-based classification procedure for discrimination between ADHD probands and normal controls.

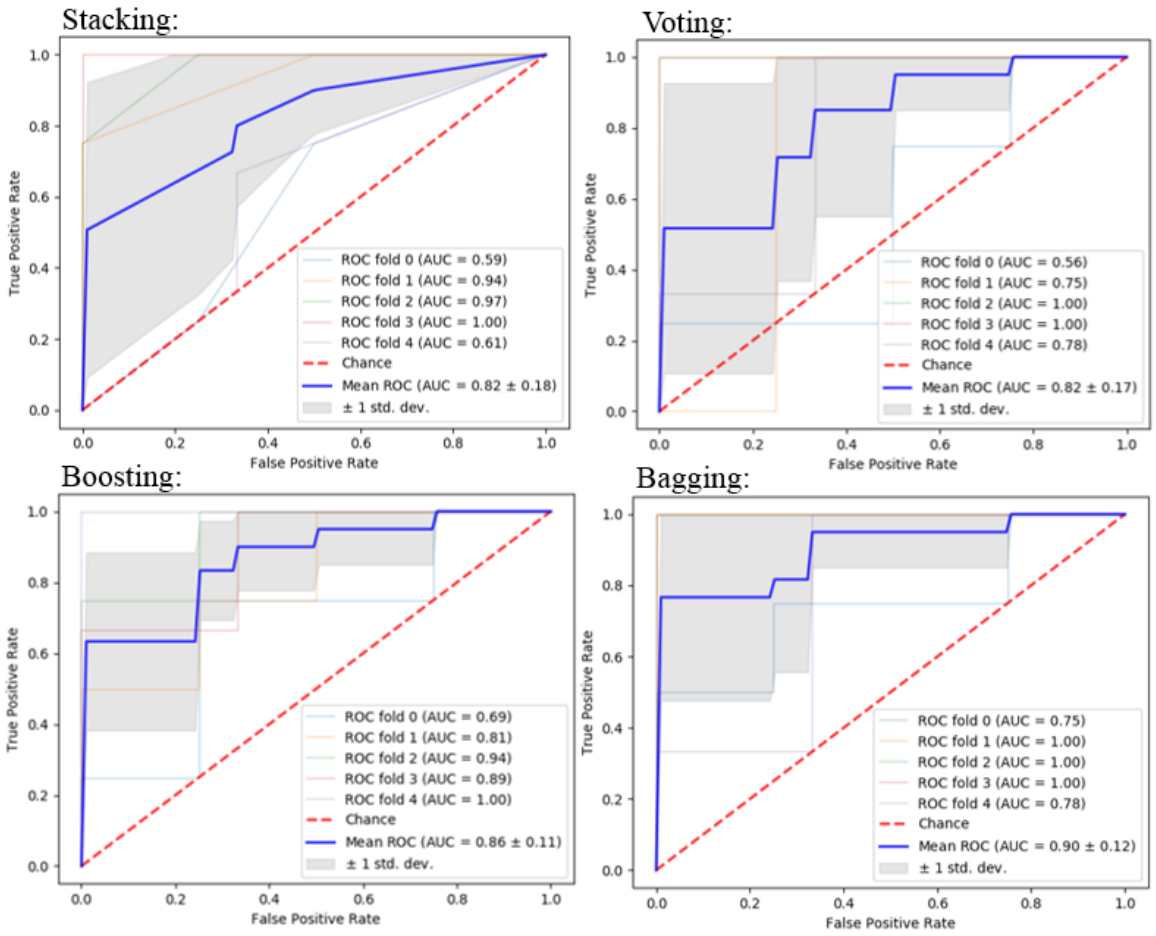

Figure 5.20 AUC of each ELT-based classification procedure for discrimination between ADHD persisters and ADHD remitters.

5.3.4 Importance Score of Classification Models

The importance score of top three features for the classifications between ADHD probands and NCs, and between ADHD-P and ADHD-R are shown in Table 5.9. More specifically, the nodal efficiency of right IFG, the functional connectivity between right MFG and right IPL, the volume of right amygdala served as the top three important features in the classification model between ADHD and NCs. The nodal efficiency of right MFG,

functional connectivity between right MFG and right IPL, and betweenness centrality of left putamen played the three most important characteristics in the classification between ADHD-P and ADHD-R.

Table 5.9 Importance Scores of Top Three Features in Classifications Between ADHD Probands and Normal Controls, as well as Between ADHD Persisters and ADHD Remitters

Feature	Importance Score
ADHD vs. NC	
Nodal efficiency of right Inferior Frontal gyrus	0.134
FC between right Middle Frontal gyrus and right Inferior Parietal lobule	0.111
Volume of right amygdala	0.1
ADHD-P vs. ADHD-R	
Nodal efficiency of right Middle Frontal gyrus	1.028
FC between right Middle Frontal gyrus and right Inferior Parietal lobule	0.852
Betweenness-centrality of left putamen	0.677

5.3.5 Regression Model and Importance Score

The regression results (Table 5.10) indicate that elastic net regression performed the best for the prediction of both inattentive and hyperactive/impulsive T-scores. Table 5.11 shows the importance scores of the top three features of elastic net regression for inattentive and hyperactive/impulsive symptom T-scores. Specifically, the top three features for the prediction of inattentive T-score were the nodal efficiency of right IFG, the functional connectivity between MFG and IPL in right hemisphere, the volume of right amygdala. The top three features for the prediction of hyperactive/impulsive T-score

included the nodal efficiency in right IFG, the functional connectivity between right MFG and right IPL, the nodal efficiency of right MFG.

Table 5.10 Pearson Correlation Coefficient and Mean Squared Error Performance of Regression Models

Regression	Pearson Correlation Coefficient	MSE
T-Inattentive		
OLS	r=0.4603; p<0.001	126.3
LASSO	r=0.4592; p<0.001	124.6
Ridge	r=0.4605; p<0.001	126.1
Elastic Net	r=0.4689; p<0.001	121.1
T-Hyperactive/Impulsive		
OLS	r=0.3329; p=0.0043	126.5
LASSO	r=0.3395; p=0.0035	123.3
Ridge	r=0.3334; p=0.0042	126.3
Elastic Net	r=0.3488; p=0.0027	119.8

Table 5.11 Importance Scores of Top Three Features in Elastic Net Regression for Inattentive and Hyperactive/Impulsive Symptom T-scores

Feature	r	*p*	Importance Score
Inattentive			
Nodal efficiency of right Inferior Frontal gyrus	-0.399	0.001	3.471
FC between right Middle Frontal gyrus and right Inferior Parietal lobule	0.405	<0.001	2.126
Volume of right Amygdala	-0.011	0.928	1.819
Hyperactive/Impulsive			
Nodal efficiency of right Inferior Frontal gyrus	-0.345	0.003	2.289
FC between right Middle Frontal gyrus and right Inferior Parietal lobule	0.361	0.002	2.134
Nodal efficiency of right Middle Frontal gyrus	-0.333	0.004	1.997

5.4 Discussion

To the best of our knowledge, this is the first study to apply ELT to multimodal neuroimaging features generated from structural MRI, DTI, and task-based fMRI data

collected from a sample of adults with childhood ADHD and controls, who have been clinically followed up since childhood. We found that the nodal efficiency in right IFG, functional connectivity between MFG and IPL in right hemisphere, and right amygdala volume were the most important features for discrimination between the ADHD probands and controls, while the nodal efficiency of right MFG, functional connectivity between right MFG and right IPL, and betweenness-centrality of left putamen played the most important roles for discrimination between the ADHD-P and ADHD-R. Moreover, the classification performance parameters of ELT-based procedures were superior to those of the basic classifiers.

5.4.1 Neurobiological Markers for Discriminations

Our current study observed the important roles of nodal efficiency in right IFG and functional connectivity between right MFG and right IPL for discrimination between ADHD probands and NCs. The abnormalities of these regions have been supported by a variety of existing neuroimaging and machine learning studies. Specifically, both task-based and resting-state fMRI studies have consistently reported the decreased functional activation in right IFG (Rubia et al., 1999; Silk et al., 2005; Cao et al., 2006; Konrad et al., 2006; Smith et al., 2006; Rubia et al., 2019) and reduced functional connectivity between right MFG and right IPL (Vance et al., 2007; Lin et al., 2015) in children with ADHD as compared with NCs. In addition, multivariate machine learning

and ELT-based studies have commonly reported that functional activation and connectivity in frontal and parietal areas are associated with improved classification between children with ADHD and NCs (Brown et al., 2012; Colby et al., 2012; Fair et al., 2012; dos Santos Siqueira et al., 2014; Deshpande et al., 2015; Iannaccone et al., 2015; Qureshi et al., 2017). They have supported the hypothesis that functional abnormalities in frontal and parietal areas, which are critical components of the attention network in human brain, especially stimulus-driven top-down control, are associated with the symptom emergence of childhood ADHD (Posner and Rothbart, 2009). Additionally, we found that the volume of right amygdala played a vital role in discrimination of ADHD probands and controls. The findings of amygdala anatomical abnormities in children with ADHD were supported by many previous studies. Amygdala plays as a critically important role in emotion regulation (Davidson et al., 2000; Banks et al., 2007; Domes et al., 2010) and thus structural anomalies associated amygdala have been widely observed in children (Van Dessel et al., 2019) and adults with ADHD (Tajima-Pozo et al., 2018), which suggests that the aberrant structure of amygdala may be associated with less control of impulsivity and delay aversion (Van Dessel et al., 2018).

Additionally, our findings point to the important roles of nodal efficiency in right MFG, functional connectivity between right MFG and right IPL for discrimination between ADHD-P and ADHD-R, and findings were supported by a variety of existing neuroimaging studies. More specifically, reduced activation and functional connectivity in

IFG, MFG, and fronto-parietal regions were observed in ADHD-P when compared to ADHD-R (Clerkin et al., 2013; Mattfeld et al., 2014; Schulz et al., 2017). The functional activation and connectivity in frontal and parietal regions during cognitive control were associated with the diverse adult outcomes of ADHD diagnosed in childhood, with symptom persistence linked to reduced activation or symptom recovery associated with higher connectivity within frontal areas (Clerkin et al., 2013; Mattfeld et al., 2014; Francx et al., 2015; Schulz et al., 2017). Although several existing multivariate machine learning and ELT-based studies have commonly reported that the anatomical features in frontal and parietal areas are associated with the classification performance between adults with ADHD and group-matched NCs (Chaim-Avancini et al., 2017; Zhang-James et al., 2019), no machine learning study has been conducted to identify the classification pattern for discrimination between ADHD-P and ADHD-R. We further observed that the features of nodal efficiency in right IFG, functional connectivity between right MFG and right IPL, and right amygdala volume were associated with inattentive symptom severity T-score, while the nodal efficiencies of right IFG and MFG and functional connectivity between MFG and IPL in right hemisphere were associated with hyperactive/impulsive symptom severity T-score. These findings suggest the significant involvement of frontal and parietal lobes in right hemisphere for both inattentive and hyperactive symptom persistence of childhood ADHD (Francx et al., 2015).

5.4.2 Performance of Classification and Regression Models

Moreover, we found that the classification performance parameters of ELT-based procedures were improved compared to those of basic models. The ELTs have been developed in the big data science field to adaptively combine multiple basic classifiers in order to strategically deal with feature variance and bias, and optimize prediction performances (Hansen and Salamon, 1990; Schapire, 1990; Dror et al., 2011; Balakrishnan et al., 2012). According to our classification results, bagging, sampling with replacement, would help to reduce the chance overfitting complex models. In our study, bagging with the basic model of SVC was applied to train our model and proved to be the best classifier for both discriminations. In addition, we used AUC statistic for model evaluation, instead on commonly used accuracy, which can be influenced by case-control imbalance in data sets (Hanley and McNeil, 1982; Fawcett, 2006). Our study showed a satisfactory performance of AUC with 0.89 and 0.9 for the discrimination between groups of ADHD and NCs, and between the groups of ADHD-P and ADHD-R, respectively. Although we had a relatively small sample size, our findings suggest that ELT-based models performed better than basic models.

In addition, the elastic net-based regression model demonstrated the best performance parameters when investigating the relations between the neuroimaging features and clinical symptom measures in the ADHD probands. The reason elastic net regression had the best performance was that it compromised the LASSO penalty (L1) and

the ridge penalty (L2) (Zou and Hastie, 2005). The LASSO (L1) penalty function performs variable selection and dimension reduction by shrinking coefficients (Tibshirani, 1996), while the ridge (L2) penalty function shrinks the coefficients of correlated variables toward their average (Hoerl and Kennard, 1970). Therefore, as for the study with relatively small sample size, the combined method obviously performed better than isolated ones, e.g., LASSO regression and ridge regression.

5.4.3 Limitations

Although the ELTs improved the model classification performance, especially for the cases when the base models had weak classification results, the current study has some limitations. First, our cohort consisted of both male and female subjects, but many more males. It is still unclear whether the discrimination models of ADHD differ between males and females. The future work may focus on constructing the classification models for both males and females. Second, the sample size of this study is relatively small. Although our study provided a considerable robust algorithm to reduce the overfitting, the relative small sample size may still influence the model's performance. Future work will need a much larger cohort to test the ELTs.

5.5 Conclusion

In summary, together with existing findings, results of this study suggest that structural and functional alterations in frontal, parietal, and amygdala areas and their functional interactions significantly contribute to accurate discrimination of ADHD probands from controls; while abnormal fronto-parietal functional communications in the right hemisphere plays an important role in symptom persistence in adults with childhood ADHD. Furthermore, the classification performance parameters (accuracy, AUC of the ROC, etc.) of the ELT-based procedures were improved than those of basic model-based procedures, which suggests that ELTs may have the potential to identify more reliable neurobiological markers for neurodevelopmental disorders.

CHAPTER 6

CONCLUSION AND FUTURE DIRECTIONS

This dissertation is the first study to identify the neurobiological markers for remission and persistence of childhood-onset ADHD using multimodal MRI and advanced machine learning techniques. It is one of the first studies to comprehensively examine the functional and structural alterations associated with onset and diverse adult outcomes of childhood ADHD, and utilized advanced machine learning technique to identify the importance of features. First, this research has been conducted to determine the neurobiological mechanisms associated with remission and persistence of childhood ADHD. Most existing ADHD neuroimaging studies have focused on the etiology of ADHD while few of them have been conducted to determine the pathology of ADHD. Second, as to fMRI data analysis project, the utilization of topological properties of cue-evoked attention processing network could help us investigate the complex organization of the functional brain network. Such a complexity arises from several, integrated, segregated, and distributed functional networks around critical areas involved in the specific cognitive function. The characteristic of small-world network enables information to travel quickly and efficiently even between far brain structures, as well as to prevent the uncontrolled spread of information across the whole network (Hilgetag and Kaiser, 2004). Graph theoretic technique provides powerful mathematical tools to study the behavior of complex

systems of interacting elements (Bullmore and Sporns, 2009). It has been widely used to characterize local and distributed interactions in the brain, and altered topological characteristics in functional brain networks have been observed in psychiatric disorders (Bassett and Bullmore, 2009). Third, this dissertation applied functional and structural MRI and DTI in the same study sample to identify the functional and structural markers, that are associated with symptom persistence and remission in young adults with childhood ADHD. Fourth, ensemble learning technique in combination with cross validation were applied in this research to deal with complicated feature variations, biases, and optimized prediction performances.

6.1 Implications of Current Findings

Our current findings of neurobiological mechanisms regarding emergence and diverse adult outcomes are tremendously valuable to the current diagnosis, longitudinal inference and treatment. Although the diagnostic criteria for ADHD have evolved over time, the assessment and tools for evaluation have remained essentially the same. The use of teacher- or parent-reported behavior-rating scales started in the late 1960s; now focus is on the behavioral criteria for ADHD as described in the DSM-5. Unfortunately, the objective assessments currently available for ADHD are of limited use in clarifying the diagnosis, including neuropsychological tests as well as neuroimaging. In addition, current treatment strategies for ADHD symptoms include medication-based, behavior-based, and combined

interventions. Stimulant medications that affect the dopaminergic system, including methylphenidate (Ritalin, Concerta, Metadata, Methylin) and certain amphetamines (Dexedrine, Dextrostat, Adderall), are most commonly prescribed for ADHD. Besides medications, behavior-based treatments, including education and/or behavior therapy, and social skills training have also been implemented in practice for ADHD interventions. Nevertheless, there is yet no curative treatment for ADHD without thoroughly understanding its heterogeneous and developmental pathophysiological mechanisms. Psychiatric and behavioral comorbidities, such as depression, anxiety disorder, bipolar disorder, substance use, and personality disorders, often co-occur with ADHD and result in increased difficulties for appropriate diagnoses and treatments. Although different pharmacological treatment strategies have been applied to ADHD patients with various comorbidities, evidence from a large body of studies showed that treatment responses from different patients are widely different in terms of the types of pharmacological treatments, dosage requirements, tolerability, response rates, and adverse-event profiles. Multiple factors may contribute to the treatment response heterogeneity in ADHD. Understanding the neural mechanisms that underpin this variability in the adult outcomes of childhood ADHD can inform novel treatment approaches and might provide biomarkers to help predict outcome.

In the light of such potential, the results from Aim1 of this dissertation highlight the importance of cue-evoked attention processing brain network. Among such brain

network, missing network hubs in IPL and lacking fronto-parietal functional communications were observed in ADHD probands relative to NCs. Furthermore, right frontal lobe functional impairments may relate to inefficient fronto-parietal functional interactions for sensory and cognitive information processing and symptom persistence in young adults with childhood ADHD. Thus, this finding could help better understand the relationship between inefficient fronto-parietal functional communications and inattentive symptom. It is believed that results from this dissertation would encourage future studies to identify the role of fronto-parietal functioning in the diverse adult outcomes of childhood ADHD. Results from Aim2 observed a reduced GM volume in right putamen and a decreased volume in left parieto-insular WM tract in ADHD probands relative to NCs; as well as an important role of striato-cortical, especially caudate-parietal WM fiber tracts determining the outcomes in adults with childhood ADHD. The findings indicated that optimal structural development in the caudate-parietal WM tract may partially modulate the functional integrity of caudate and parietal cortex, and together contribute to symptom remission in adults with childhood ADHD.

Although Aims 1 and 2 applied statistical analysis to identify the neural mechanisms of remission and persistence of childhood ADHD, two possibilities existed to be further identified. One possibility is that symptom remission in adulthood is due to the correction of childhood neural anomalies, whereas clinical persistence is tied to the persistence of neural anomalies. Alternatively, remission might arise from neural

reorganization as novel systems are recruited to help the individual compensate for core deficits of ADHD. These two models make different predictions about the remitted brain. In the first model, neural features in those with symptom remission will resemble those seen among individuals never affected by ADHD. If, however, neural reorganization and compensation drives remission, then the remitted brain will differ from the never affected, albeit in potentially beneficial ways. Finally, it is also possible that some anomalies that reflect the childhood presence of ADHD could persist, regardless of clinical recovery. By this reckoning, both those who have symptom remission and those with symptom persistence will show very similar atypical neural features, despite different clinical presentation.

Current recommendations for diagnostic evaluation of possible ADHD include a comprehensive history taking of prenatal, perinatal, and family history; school performance; environmental factors; and a detailed physical examination. During the physical examination, particular attention should be paid to vital signs, and a mental health assessment used to probe for comorbid conditions should be performed. Although neuroimaging provides a window into the developing brain, allowing us to examine safely and noninvasively brain anatomy, function, biochemistry, and connectivity. When applied to neurodevelopmental disorders, such as ADHD, imaging in vivo could provide objective tools to inform diagnosis, prognosis, and stimulate discovery of novel therapeutics. Thus, the efforts to translate neuroimaging into clinical tools should be highlighted.

The most significant contribution of this dissertation towards understanding the neurobiological mechanisms of onset and remission/persistence of childhood ADHD and the translation from neuroimaging to clinical tools is the utilization of machine learning, i.e., ensemble learning technique in Aim 3. Although this dissertation implemented novel analytic approaches applied to multimodal neuroimaging data to help bring imaging into the clinic of the future. The one objective of this dissertation is that we can translate the modest, but significant neural differences between groups—those with and without ADHD or those with remitted ADHD and persistent ADHD—to the individual level. This dissertation achieved a 0.89 of AUC for discrimination between ADHD proands and NCs with the important roles of nodal efficiency in right IFG and functional connectivity between right MFG and right IPL, as well as the volume of right amygdala as well as a 0.9 of AUC for discrimination between ADHD-P and ADHD-R with the important roles of nodal efficiency of right MFG, functional connectivity between right MFG and right IPL, and BC of left putamen. The results of Aim 3 indicated that for both diagnosis and recovery treatment, relative to neuroanatomy, the functional neuroimaging studies should be central in demonstrating how psychostimulants may work, and in the future might be used to provide neurofeedback to shift brain activation into more neurotypical ranges.

6.2 Limitations and Future Directions

Several limitations in the methodology and the research process employed in this dissertation must be noted. First, a major limitation in these studies is that imaging data were acquired in adulthood. Thus, the similarity between the remitted and never affected groups could arise if, in the group with remission, there were more typical neural features in childhood that had been carried forward into adulthood. A critical next stage is to collect prospectively both clinical and imaging data from childhood into adulthood and use these data to define the bonds between neural and clinical trajectories. Second, The scientific community has also been interested in understanding how neurotransmitter systems are involved in ADHD because animal models, neuroimaging studies, and pharmacologic studies provide support for the involvement of dopaminergic and adrenergic derangements in ADHD (Solanto, 2002; Del Campo et al., 2011). However, no evidence-based methods for assessing these neurotransmitter systems have been developed and shown to have utility in the ADHD diagnostic assessment. Future studies should be conducted to identify the relationship between neurotransmitter systems and neuroimaging findings. In addition, this dissertation has a limited sample size since the characteristic of longitudinal study. To overcome this barrier, collaborative studies that will provide more robust measures of the functional and structural brain abnormalities in ADHD should be conducted before neuroimaging becomes clinically useful. Such collaborations face challenges and need to

be further solved, such as integrating data acquired using different scanners and sequences, ideally reflecting a process central to ADHD or one of its underlying.

REFERENCES

Achard, S., & Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. *PLoS Comput Biol* 3**,** e17.

Adeyemo, B.O., Biederman, J., Zafonte, R., Kagan, E., Spencer, T.J., Uchida, M., Kenworthy, T., Spencer, A.E., & Faraone, S.V. (2014). Mild traumatic brain injury and adhd: A systematic review of the literature and meta-analysis. *J Atten Disord* 18**,** 576-584.

Agnew-Blais, J.C., Polanczyk, G.V., Danese, A., Wertz, J., Moffitt, T.E., & Arseneault, L. (2016). Evaluation of the persistence, remission, and emergence of attention-deficit/hyperactivity disorder in young adulthood. *JAMA Psychiatry* 73**,** 713-720.

Alexander, G.E., Delong, M.R., & Strick, P.L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 9**,** 357-381.

Almeida Montes, L.G., Prado Alcantara, H., Martinez Garcia, R.B., De La Torre, L.B., Avila Acosta, D., & Duarte, M.G. (2013). Brain cortical thickness in adhd: Age, sex, and clinical correlations. *J Atten Disord* 17**,** 641-654.

Anastopoulos, A.D., King, K.A., Besecker, L.H., O'rourke, S.R., Bray, A.C., & Supple, A.J. (2018). Cognitive-behavioral therapy for college students with adhd: Temporal stability of improvements in functioning following active treatment. *J Atten Disord***,** 1087054717749932.

Antshel, K.M., Hargrave, T.M., Simonescu, M., Kaul, P., Hendricks, K., & Faraone, S.V. (2011). Advances in understanding and treating adhd. *BMC Med* 9**,** 72.

Ashtari, M., Kumra, S., Bhaskar, S.L., Clarke, T., Thaden, E., Cervellione, K.L., Rhinewine, J., Kane, J.M., Adesman, A., Milanaik, R., Maytal, J., Diamond, A., Szeszko, P., & Ardekani, B.A. (2005). Attention-deficit/hyperactivity disorder: A preliminary diffusion tensor imaging study. *Biol Psychiatry* 57**,** 448-455.

Azevedo, F.A., Carvalho, L.R., Grinberg, L.T., Farfel, J.M., Ferretti, R.E., Leite, R.E., Jacob Filho, W., Lent, R., & Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol* 513**,** 532-541.

Balakrishnan, S., Wang, R., Scheidegger, C., Maclellan, A., Hu, Y., Archer, A., Krishnan, S., Applegate, D., Ma, G.Q., & Au, S.T. (2012). "Combining predictors for recommending music:The false positives' approach to kdd cup track 2", in: *Proceedings of KDD Cup 2011*. (eds.) D. Gideon, K. Yehuda & W. Markus. (Proceedings of Machine Learning Research: PMLR).

Banks, S.J., Eddy, K.T., Angstadt, M., Nathan, P.J., & Phan, K.L. (2007). Amygdala-frontal connectivity during emotion regulation. *Soc Cogn Affect Neurosci* 2**,** 303-312.

Barkley, R.A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of adhd. *Psychol Bull* 121**,** 65-94.

Barkley, R.A., & Brown, T.E. (2008). Unrecognized attention-deficit/hyperactivity disorder in adults presenting with other psychiatric disorders. *CNS Spectr* 13**,** 977-984.

Bartels, A., & Zeki, S. (2000). The architecture of the colour centre in the human visual brain: New results and a review. *Eur J Neurosci* 12**,** 172-193.

Bassett, D.S., & Bullmore, E. (2006). Small-world brain networks. *Neuroscientist* 12**,** 512-523.

Bassett, D.S., & Bullmore, E.T. (2009). Human brain networks in health and disease. *Curr Opin Neurol* 22**,** 340-347.

Batty, M.J., Liddle, E.B., Pitiot, A., Toro, R., Groom, M.J., Scerif, G., Liotti, M., Liddle, P.F., Paus, T., & Hollis, C. (2010). Cortical gray matter in attention-deficit/hyperactivity disorder: A structural magnetic resonance imaging study. *J Am Acad Child Adolesc Psychiatry* 49**,** 229-238.

Becker, K., El-Faddagh, M., Schmidt, M.H., Esser, G., & Laucht, M. (2008). Interaction of dopamine transporter genotype with prenatal smoke exposure on adhd symptoms. *J Pediatr* 152**,** 263-269.

Behrens, T.E., Berg, H.J., Jbabdi, S., Rushworth, M.F., & Woolrich, M.W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage* 34**,** 144-155.

Berridge, C.W., Devilbiss, D.M., Andrzejewski, M.E., Arnsten, A.F., Kelley, A.E., Schmeichel, B., Hamilton, C., & Spencer, R.C. (2006). Methylphenidate preferentially increases catecholamine neurotransmission within the prefrontal cortex at low doses that enhance cognitive function. *Biol Psychiatry* 60**,** 1111-1120.

Bidwell, L.C., Karoly, H.C., Hutchison, K.E., & Bryan, A.D. (2017). Adhd symptoms impact smoking outcomes and withdrawal in response to varenicline treatment for smoking cessation. *Drug Alcohol Depend* 179**,** 18-24.

Bidwell, L.C., Willcutt, E.G., Mcqueen, M.B., Defries, J.C., Olson, R.K., Smith, S.D., & Pennington, B.F. (2011). A family based association study of drd4, dat1, and 5htt and continuous traits of attention-deficit hyperactivity disorder. *Behav Genet* 41**,** 165-174.

Biederman, J., Kwon, A., Aleardi, M., Chouinard, V.A., Marino, T., Cole, H., Mick, E., & Faraone, S.V. (2005). Absence of gender effects on attention deficit hyperactivity disorder: Findings in nonreferred subjects. *Am J Psychiatry* 162**,** 1083-1089.

Biederman, J., Mick, E., & Faraone, S.V. (2000). Age-dependent decline of symptoms of attention deficit hyperactivity disorder: Impact of remission definition and symptom type. *Am J Psychiatry* 157**,** 816-818.

Bledsoe, J.C., Semrud-Clikeman, M., & Pliszka, S.R. (2011). Neuroanatomical and neuropsychological correlates of the cerebellum in children with attention-deficit/hyperactivity disorder--combined type. *J Am Acad Child Adolesc Psychiatry* 50**,** 593-601.

Bonvicini, C., Faraone, S.V., & Scassellati, C. (2016). Attention-deficit hyperactivity disorder in adults: A systematic review and meta-analysis of genetic, pharmacogenetic and biochemical studies. *Mol Psychiatry* 21**,** 1643.

Bonvicini, C., Faraone, S.V., & Scassellati, C. (2018). Common and specific genes and peripheral biomarkers in children and adults with attention-deficit/hyperactivity disorder. *World J Biol Psychiatry* 19**,** 80-100.

Booth, J.R., Burman, D.D., Meyer, J.R., Lei, Z., Trommer, B.L., Davenport, N.D., Li, W., Parrish, T.B., Gitelman, D.R., & Mesulam, M.M. (2005). Larger deficits in brain networks for response inhibition than for visual selective attention in attention deficit hyperactivity disorder (adhd). *J Child Psychol Psychiatry* 46**,** 94-111.

Breiman, L. (1996). Bagging predictors. *Machine Learning* 24**,** 18.

Brookes, K., Xu, X., Chen, W., Zhou, K., Neale, B., Lowe, N., Anney, R., Franke, B., Gill, M., Ebstein, R., Buitelaar, J., Sham, P., Campbell, D., Knight, J., Andreou, P., Altink, M., Arnold, R., Boer, F., Buschgens, C., Butler, L., Christiansen, H., Feldman, L., Fleischman, K., Fliers, E., Howe-Forbes, R., Goldfarb, A., Heise, A., Gabriels, I., Korn-Lubetzki, I., Johansson, L., Marco, R., Medad, S., Minderaa, R., Mulas, F., Muller, U., Mulligan, A., Rabin, K., Rommelse, N., Sethna, V., Sorohan, J., Uebel, H., Psychogiou, L., Weeks, A., Barrett, R., Craig, I., Banaschewski, T., Sonuga-Barke, E., Eisenberg, J., Kuntsi, J., Manor, I., Mcguffin, P., Miranda, A., Oades, R.D., Plomin, R., Roeyers, H., Rothenberger, A., Sergeant, J., Steinhausen, H.C., Taylor, E., Thompson, M., Faraone, S.V., & Asherson, P. (2006a). The analysis of 51 genes in dsm-iv combined type attention deficit hyperactivity disorder: Association signals in drd4, dat1 and 16 other genes. *Mol Psychiatry* 11**,** 934-953.

Brookes, K.J., Mill, J., Guindalini, C., Curran, S., Xu, X., Knight, J., Chen, C.K., Huang, Y.S., Sethna, V., Taylor, E., Chen, W., Breen, G., & Asherson, P. (2006b). A common haplotype of the dopamine transporter gene associated with attention-deficit/hyperactivity disorder and interacting with maternal use of alcohol during pregnancy. *Arch Gen Psychiatry* 63**,** 74-81.

Brown, M.R., Sidhu, G.S., Greiner, R., Asgarian, N., Bastani, M., Silverstone, P.H., Greenshaw, A.J., & Dursun, S.M. (2012). Adhd-200 global competition: Diagnosing adhd using personal characteristic data can outperform resting state fmri measurements. *Front Syst Neurosci* 6**,** 69.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10**,** 186-198.

Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nat Rev Neurosci* 13**,** 336-349.

Bush, G. (2011). Cingulate, frontal, and parietal cortical dysfunction in attention-deficit/hyperactivity disorder. *Biol Psychiatry* 69**,** 1160-1167.

Bush, G., Valera, E.M., & Seidman, L.J. (2005). Functional neuroimaging of attention-deficit/hyperactivity disorder: A review and suggested future directions. *Biol Psychiatry* 57**,** 1273-1284.

Cao, M., Shu, N., Cao, Q., Wang, Y., & He, Y. (2014a). Imaging functional and structural brain connectomics in attention-deficit/hyperactivity disorder. *Mol Neurobiol* 50**,** 1111-1123.

Cao, M., Wang, J.H., Dai, Z.J., Cao, X.Y., Jiang, L.L., Fan, F.M., Song, X.W., Xia, M.R., Shu, N., Dong, Q., Milham, M.P., Castellanos, F.X., Zuo, X.N., & He, Y. (2014b). Topological organization of the human brain functional connectome across the lifespan. *Dev Cogn Neurosci* 7**,** 76-93.

Cao, Q., Zang, Y., Sun, L., Sui, M., Long, X., Zou, Q., & Wang, Y. (2006). Abnormal neural activity in children with attention deficit hyperactivity disorder: A resting-state functional magnetic resonance imaging study. *Neuroreport* 17**,** 1033-1036.

Cao, Q., Zang, Y., Zhu, C., Cao, X., Sun, L., Zhou, X., & Wang, Y. (2008). Alerting deficits in children with attention deficit/hyperactivity disorder: Event-related fmri evidence. *Brain Res* 1219**,** 159-168.

Carmona, S., Vilarroya, O., Bielsa, A., Tremols, V., Soliva, J.C., Rovira, M., Tomas, J., Raheb, C., Gispert, J.D., Batlle, S., & Bulbena, A. (2005). Global and regional gray matter reductions in adhd: A voxel-based morphometric study. *Neurosci Lett* 389**,** 88-93.

Casey, B.J., Castellanos, F.X., Giedd, J.N., Marsh, W.L., Hamburger, S.D., Schubert, A.B., Vauss, Y.C., Vaituzis, A.C., Dickstein, D.P., Sarfatti, S.E., & Rapoport, J.L. (1997). Implication of right frontostriatal circuitry in response inhibition and attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 36**,** 374-383.

Casey, B.J., Nigg, J.T., & Durston, S. (2007). New potential leads in the biology and treatment of attention deficit-hyperactivity disorder. *Curr Opin Neurol* 20**,** 119-124.

Castellanos, F.X., Lee, P.P., Sharp, W., Jeffries, N.O., Greenstein, D.K., Clasen, L.S., Blumenthal, J.D., James, R.S., Ebens, C.L., Walter, J.M., Zijdenbos, A., Evans, A.C., Giedd, J.N., & Rapoport, J.L. (2002). Developmental trajectories of brain volume abnormalities in children and adolescents with attention-deficit/hyperactivity disorder. *JAMA* 288**,** 1740-1748.

Chacko, A., Bedard, A.V., Marks, D., Gopalan, G., Feirsen, N., Uderman, J., Chimiklis, A., Heber, E., Cornwell, M., Anderson, L., Zwilling, A., & Ramon, M. (2018). Sequenced neurocognitive and behavioral parent training for the treatment of adhd in school-age children. *Child Neuropsychol* 24**,** 427-450.

Chaim-Avancini, T.M., Doshi, J., Zanetti, M.V., Erus, G., Silva, M.A., Duran, F.L.S., Cavallet, M., Serpa, M.H., Caetano, S.C., Louza, M.R., Davatzikos, C., & Busatto, G.F. (2017). Neurobiological support to the diagnosis of adhd in stimulant-naive adults: Pattern recognition analyses of mri data. *Acta Psychiatr Scand* 136**,** 623-636.

Chang, C.H., Yu, C.J., Du, J.C., Chiou, H.C., Chen, H.C., Yang, W., Chung, M.Y., Chen, Y.S., Hwang, B., Mao, I.F., & Chen, M.L. (2018). The interactions among organophosphate pesticide exposure, oxidative stress, and genetic polymorphisms of dopamine receptor d4 increase the risk of attention deficit/hyperactivity disorder in children. *Environ Res* 160**,** 339-346.

Chang, C.W., Ho, C.C., & Chen, J.H. (2012). Adhd classification by a texture analysis of anatomical brain mri data. *Front Syst Neurosci* 6**,** 66.

Chen, W., Zhou, K., Sham, P., Franke, B., Kuntsi, J., Campbell, D., Fleischman, K., Knight, J., Andreou, P., Arnold, R., Altink, M., Boer, F., Boholst, M.J., Buschgens, C., Butler, L., Christiansen, H., Fliers, E., Howe-Forbes, R., Gabriels, I., Heise, A., Korn-Lubetzki, I., Marco, R., Medad, S., Minderaa, R., Muller, U.C., Mulligan, A., Psychogiou, L., Rommelse, N., Sethna, V., Uebel, H., Mcguffin, P., Plomin, R., Banaschewski, T., Buitelaar, J., Ebstein, R., Eisenberg, J., Gill, M., Manor, I., Miranda, A., Mulas, F., Oades, R.D., Roeyers, H., Rothenberger, A., Sergeant, J., Sonuga-Barke, E., Steinhausen, H.C., Taylor, E., Thompson, M., Faraone, S.V., & Asherson, P. (2008). Dsm-iv combined type adhd shows familial association with sibling trait scores: A sampling strategy for qtl linkage. *Am J Med Genet B Neuropsychiatr Genet* 147B**,** 1450-1460.

Chen, Y., Wang, S., Hilgetag, C.C., & Zhou, C. (2013). Trade-off between multiple constraints enables simultaneous formation of modules and hubs in neural systems. *PLoS Comput Biol* 9**,** e1002937.

Cheng, W., Ji, X., Zhang, J., & Feng, J. (2012). Individual classification of adhd patients by integrating multiscale neuroimaging markers and advanced pattern recognition techniques. *Front Syst Neurosci* 6**,** 58.

Chiu, Y.C., Jiang, J., & Egner, T. (2017). The caudate nucleus mediates learning of stimulus-control state associations. *J Neurosci* 37**,** 1028-1038.

Clark, D.L., Arnold, L.E., Crowl, L., Bozzolo, H., Peruggia, M., Ramadan, Y., Bornstein, R., Hollway, J.A., Thompson, S., Malone, K., Hall, K.L., Shelton, S.B., Bozzolo, D.R., & Cook, A. (2008). Vestibular stimulation for adhd: Randomized controlled trial of comprehensive motion apparatus. *J Atten Disord* 11**,** 599-611.

Clerkin, S.M., Schulz, K.P., Berwid, O.G., Fan, J., Newcorn, J.H., Tang, C.Y., & Halperin, J.M. (2013). Thalamo-cortical activation and connectivity during response preparation in adults with persistent and remitted adhd. *Am J Psychiatry* 170**,** 1011-1019.

Clerkin, S.M., Schulz, K.P., Halperin, J.M., Newcorn, J.H., Ivanov, I., Tang, C.Y., & Fan, J. (2009). Guanfacine potentiates the activation of prefrontal cortex evoked by warning signals. *Biol Psychiatry* 66**,** 307-312.

Cocchi, L., Bramati, I.E., Zalesky, A., Furukawa, E., Fontenelle, L.F., Moll, J., Tripp, G., & Mattos, P. (2012). Altered functional brain connectivity in a non-clinical sample of young adults with attention-deficit/hyperactivity disorder. *J Neurosci* 32**,** 17753-17761.

Colby, J.B., Rudie, J.D., Brown, J.A., Douglas, P.K., Cohen, M.S., & Shehzad, Z. (2012). Insights into multimodal imaging classification of adhd. *Front Syst Neurosci* 6**,** 59.

Cortese, S., Imperati, D., Zhou, J., Proal, E., Klein, R.G., Mannuzza, S., Ramos-Olazagasti, M.A., Milham, M.P., Kelly, C., & Castellanos, F.X. (2013). White matter alterations at 33-year follow-up in adults with childhood attention-deficit/hyperactivity disorder. *Biol Psychiatry* 74**,** 591-598.

Cortese, S., Kelly, C., Chabernaud, C., Proal, E., Di Martino, A., Milham, M.P., & Castellanos, F.X. (2012). Toward systems neuroscience of adhd: A meta-analysis of 55 fmri studies. *Am J Psychiatry* 169**,** 1038-1055.

Craddock, R.C., Jbabdi, S., Yan, C.G., Vogelstein, J.T., Castellanos, F.X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S., & Milham, M.P. (2013). Imaging human connectomes at the macroscale. *Nat Methods* 10**,** 524-539.

Cubillo, A., Halari, R., Smith, A., Taylor, E., & Rubia, K. (2012). A review of fronto-striatal and fronto-cortical brain abnormalities in children and adults with attention deficit hyperactivity disorder (adhd) and new evidence for dysfunction in adults with adhd during motivation and attention. *Cortex* 48**,** 194-215.

Da Silva, B.S., Cupertino, R.B., Rovaris, D.L., Schuch, J.B., Kappel, D.B., Muller, D., Bandeira, C.E., Victor, M.M., Karam, R.G., Mota, N.R., Rohde, L.A., Contini, V., Grevet, E.H., & Bau, C.H.D. (2018). Exocytosis-related genes and response to methylphenidate treatment in adults with adhd. *Mol Psychiatry* 23**,** 1446-1452.

Dai, D., Wang, J., Hua, J., & He, H. (2012). Classification of adhd children through multimodal magnetic resonance imaging. *Front Syst Neurosci* 6**,** 63.

Danielson, M.L., Bitsko, R.H., Ghandour, R.M., Holbrook, J.R., Kogan, M.D., & Blumberg, S.J. (2018). Prevalence of parent-reported adhd diagnosis and associated treatment among u.S. Children and adolescents, 2016. *J Clin Child Adolesc Psychol* 47**,** 199-212.

Davidson, R.J., Putnam, K.M., & Larson, C.L. (2000). Dysfunction in the neural circuitry of emotion regulation--a possible prelude to violence. *Science* 289**,** 591-594.

Deacon, T.W., Eichenbaum, H., Rosenberg, P., & Eckmann, K.W. (1983). Afferent connections of the perirhinal cortex in the rat. *J Comp Neurol* 220**,** 168-190.

Del Campo, N., Chamberlain, S.R., Sahakian, B.J., & Robbins, T.W. (2011). The roles of dopamine and noradrenaline in the pathophysiology and treatment of attention-deficit/hyperactivity disorder. *Biol Psychiatry* 69**,** e145-157.

Deng, L., & Platt, J. (2014). Ensemble deep learning for speech recognition. *Proc. Interspeech*.

Deshpande, G., Wang, P., Rangaprakash, D., & Wilamowski, B. (2015). Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data. *IEEE Trans Cybern* 45**,** 2668-2679.

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

Di, X., Kim, E.H., Chen, P., & Biswal, B.B. (2014). Lateralized resting-state functional connectivity in the task-positive and task-negative networks. *Brain Connect* 4**,** 641-648.

Domes, G., Schulze, L., Bottger, M., Grossmann, A., Hauenstein, K., Wirtz, P.H., Heinrichs, M., & Herpertz, S.C. (2010). The neural correlates of sex differences in emotional reactivity and emotion regulation. *Hum Brain Mapp* 31**,** 758-769.

Dos Santos Siqueira, A., Biazoli Junior, C.E., Comfort, W.E., Rohde, L.A., & Sato, J.R. (2014). Abnormal functional resting-state networks in adhd: Graph theory and pattern recognition analysis of fmri data. *Biomed Res Int* 2014**,** 380531.

Dosenbach, N.U., Fair, D.A., Cohen, A.L., Schlaggar, B.L., & Petersen, S.E. (2008). A dual-networks architecture of top-down control. *Trends Cogn Sci* 12**,** 99-105.

Dror, G., Koenigstein, N., Koren, Y., & Weimer, M. (2011). "The yahoo! Music dataset and kdd-cup'11", in: *Proceedings of the 2011 International Conference on KDD Cup 2011 - Volume 18.* JMLR.org).

Du, J., Wang, L., Jie, B., & Zhang, D. (2016). Network-based classification of adhd patients using discriminative subnetwork selection and graph kernel pca. *Comput Med Imaging Graph* 52**,** 82-88.

Dupaul, G.J., Kern, L., Belk, G., Custer, B., Daffner, M., Hatfield, A., & Peek, D. (2017). Face-to-face versus online behavioral parent training for young children at risk for adhd: Treatment engagement and outcomes. *J Clin Child Adolesc Psychol***,** 1-15.

Durston, S. (2003). A review of the biological bases of adhd: What have we learned from imaging studies? *Ment Retard Dev Disabil Res Rev* 9**,** 184-195.

Durston, S., Davidson, M.C., Mulder, M.J., Spicer, J.A., Galvan, A., Tottenham, N., Scheres, A., Castellanos, F.X., Van Engeland, H., & Casey, B.J. (2007). Neural and behavioral correlates of expectancy violations in attention-deficit hyperactivity disorder. *J Child Psychol Psychiatry* 48**,** 881-889.

Durston, S., Hulshoff Pol, H.E., Schnack, H.G., Buitelaar, J.K., Steenhuis, M.P., Minderaa, R.B., Kahn, R.S., & Van Engeland, H. (2004). Magnetic resonance imaging of boys with attention-deficit/hyperactivity disorder and their unaffected siblings. *J Am Acad Child Adolesc Psychiatry* 43**,** 332-340.

Durston, S., Mulder, M., Casey, B.J., Ziermans, T., & Van Engeland, H. (2006). Activation in ventral prefrontal cortex is sensitive to genetic vulnerability for attention-deficit hyperactivity disorder. *Biol Psychiatry* 60**,** 1062-1070.

Durston, S., Tottenham, N.T., Thomas, K.M., Davidson, M.C., Eigsti, I.M., Yang, Y., Ulug, A.M., & Casey, B.J. (2003). Differential patterns of striatal activation in young children with and without adhd. *Biol Psychiatry* 53**,** 871-878.

Efron, D., Jarman, F., & Barker, M. (1997). Methylphenidate versus dexamphetamine in children with attention deficit hyperactivity disorder: A double-blind, crossover trial. *Pediatrics* 100**,** E6.

Ellison-Wright, I., Ellison-Wright, Z., & Bullmore, E. (2008). Structural brain change in attention deficit hyperactivity disorder identified by meta-analysis. *BMC Psychiatry* 8**,** 51.

Eloyan, A., Muschelli, J., Nebel, M.B., Liu, H., Han, F., Zhao, T., Barber, A.D., Joel, S., Pekar, J.J., Mostofsky, S.H., & Caffo, B. (2012). Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging. *Front Syst Neurosci* 6**,** 61.

Epstein, J.N., Johnson, D., & Conners, C.K. (2006). Conners' adult adhd diagnostic interview for dsm-iv.

Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis. *Euler Archive - All Works* 53.

Fair, D.A., Cohen, A.L., Dosenbach, N.U., Church, J.A., Miezin, F.M., Barch, D.M., Raichle, M.E., Petersen, S.E., & Schlaggar, B.L. (2008). The maturing architecture of the brain's default network. *Proc Natl Acad Sci U S A* 105**,** 4028-4032.

Fair, D.A., Cohen, A.L., Power, J.D., Dosenbach, N.U., Church, J.A., Miezin, F.M., Schlaggar, B.L., & Petersen, S.E. (2009). Functional brain networks develop from a "local to distributed" organization. *PLoS Comput Biol* 5**,** e1000381.

Fair, D.A., Nigg, J.T., Iyer, S., Bathula, D., Mills, K.L., Dosenbach, N.U., Schlaggar, B.L., Mennes, M., Gutman, D., Bangaru, S., Buitelaar, J.K., Dickstein, D.P., Di Martino, A., Kennedy, D.N., Kelly, C., Luna, B., Schweitzer, J.B., Velanova, K., Wang, Y.F., Mostofsky, S., Castellanos, F.X., & Milham, M.P. (2012). Distinct neural signatures detected for adhd subtypes after controlling for micro-movements in resting state functional connectivity mri data. *Front Syst Neurosci* 6**,** 80.

Fair, D.A., Posner, J., Nagel, B.J., Bathula, D., Dias, T.G., Mills, K.L., Blythe, M.S., Giwa, A., Schmitt, C.F., & Nigg, J.T. (2010). Atypical default network connectivity in youth with attention-deficit/hyperactivity disorder. *Biol Psychiatry* 68**,** 1084-1091.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., Fox, P.T., Eickhoff, S.B., Yu, C., & Jiang, T. (2016). The human brainnetome atlas: A new brain atlas based on connectional architecture. *Cereb Cortex* 26**,** 3508-3526.

Farahani, F.V., Karwowski, W., & Lighthall, N.R. (2019). Application of graph theory for identifying connectivity patterns in human brain networks: A systematic review. *Front Neurosci* 13**,** 585.

Faraone, S.V., Biederman, J., & Mick, E. (2006). The age-dependent decline of attention deficit hyperactivity disorder: A meta-analysis of follow-up studies. *Psychol Med* 36**,** 159-165.

Faraone, S.V., Perlis, R.H., Doyle, A.E., Smoller, J.W., Goralnick, J.J., Holmgren, M.A., & Sklar, P. (2005). Molecular genetics of attention-deficit/hyperactivity disorder. *Biol Psychiatry* 57**,** 1313-1323.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognit Lett* 27**,** 861-874.

First, M.B., Spitzer, R.L., Gibbon, M., & Williams, J. (2002). Structured clinical interview for dsm-iv-tr axis i disorders, research version. *Biometrics Research, New York State Psychiatric Institute, New York, NY.*

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A.M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* 33**,** 341-355.

Fornito, A., Zalesky, A., & Breakspear, M. (2013). Graph analysis of the human connectome: Promise, progress, and pitfalls. *Neuroimage* 80**,** 426-444.

Francx, W., Oldehinkel, M., Oosterlaan, J., Heslenfeld, D., Hartman, C.A., Hoekstra, P.J., Franke, B., Beckmann, C.F., Buitelaar, J.K., & Mennes, M. (2015). The executive control network and symptomatic improvement in attention-deficit/hyperactivity disorder. *Cortex* 73**,** 62-72.

Frank, S.M., & Greenlee, M.W. (2018). The parieto-insular vestibular cortex in humans: More than a single area? *J Neurophysiol* 120**,** 1438-1450.

Frodl, T., & Skokauskas, N. (2012). Meta-analysis of structural mri studies in children and adults with attention deficit hyperactivity disorder indicates treatment effects. *Acta Psychiatr Scand* 125**,** 114-126.

Froehlich, T.E., Anixt, J.S., Loe, I.M., Chirdkiatgumchai, V., Kuan, L., & Gilman, R.C. (2011a). Update on environmental risk factors for attention-deficit/hyperactivity disorder. *Curr Psychiatry Rep* 13**,** 333-344.

Froehlich, T.E., Epstein, J.N., Nick, T.G., Melguizo Castro, M.S., Stein, M.A., Brinkman, W.B., Graham, A.J., Langberg, J.M., & Kahn, R.S. (2011b). Pharmacogenetic predictors of methylphenidate dose-response in attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 50**,** 1129-1139 e1122.

Gehricke, J.G., Kruggel, F., Thampipop, T., Alejo, S.D., Tatos, E., Fallon, J., & Muftuler, L.T. (2017). The brain anatomy of attention-deficit/hyperactivity disorder in young adults - a magnetic resonance imaging study. *PLoS One* 12**,** e0175433.

Ghiassian, S., Greiner, R., Jin, P., & Brown, M.R. (2016). Using functional or structural magnetic resonance images and personal characteristic data to identify adhd and autism. *PLoS One* 11**,** e0166934.

Gizer, I.R., Ficks, C., & Waldman, I.D. (2009). Candidate gene studies of adhd: A meta-analytic review. *Hum Genet* 126**,** 51-90.

Gizer, I.R., & Waldman, I.D. (2012). Double dissociation between lab measures of inattention and impulsivity and the dopamine transporter gene (dat1) and dopamine d4 receptor gene (drd4). *J Abnorm Psychol* 121**,** 1011-1023.

Glover, V. (2014). Maternal depression, anxiety and stress during pregnancy and child outcome; what needs to be done. *Best Pract Res Clin Obstet Gynaecol* 28**,** 25-35.

Gornick, M.C., Addington, A., Shaw, P., Bobb, A.J., Sharp, W., Greenstein, D., Arepalli, S., Castellanos, F.X., & Rapoport, J.L. (2007). Association of the dopamine receptor d4 (drd4) gene 7-repeat allele with children with attention-deficit/hyperactivity disorder (adhd): An update. *Am J Med Genet B Neuropsychiatr Genet* 144B**,** 379-382.

Grahn, J.A., Parkinson, J.A., & Owen, A.M. (2008). The cognitive functions of the caudate nucleus. *Prog Neurobiol* 86**,** 141-155.

Graziano, P.A., & Garcia, A. (2016). Attention-deficit hyperactivity disorder and children's emotion dysregulation: A meta-analysis. *Clin Psychol Rev* 46**,** 106-123.

Greenstein, D., Malley, J.D., Weisinger, B., Clasen, L., & Gogtay, N. (2012). Using multivariate machine learning methods and structural mri to classify childhood onset schizophrenia and healthy controls. *Front Psychiatry* 3**,** 53.

Greven, C.U., Bralten, J., Mennes, M., O'dwyer, L., Van Hulzen, K.J., Rommelse, N., Schweren, L.J., Hoekstra, P.J., Hartman, C.A., Heslenfeld, D., Oosterlaan, J., Faraone, S.V., Franke, B., Zwiers, M.P., Arias-Vasquez, A., & Buitelaar, J.K. (2015). Developmentally stable whole-brain volume reductions and developmentally sensitive caudate and putamen volume alterations in those with attention-deficit/hyperactivity disorder and their unaffected siblings. *JAMA Psychiatry* 72**,** 490-499.

Grizenko, N., Fortier, M.E., Zadorozny, C., Thakur, G., Schmitz, N., Duval, R., & Joober, R. (2012). Maternal stress during pregnancy, adhd symptomatology in children and genotype: Gene-environment interaction. *J Can Acad Child Adolesc Psychiatry* 21**,** 9-15.

Grizenko, N., Shayan, Y.R., Polotskaia, A., Ter-Stepanian, M., & Joober, R. (2008). Relation of maternal stress during pregnancy to symptom severity and response to treatment in children with adhd. *J Psychiatry Neurosci* 33**,** 10-16.

Haghshenas, S., Hosseini, M.S., & Aminjan, A.S. (2014). A possible correlation between vestibular stimulation and auditory comprehension in children with attention-deficit/hyperactivity disorder. *Psychol Neurosci* 7**,** 159-162.

Hagmann, P., Kurant, M., Gigandet, X., Thiran, P., Wedeen, V.J., Meuli, R., & Thiran, J.P. (2007). Mapping human whole-brain structural networks with diffusion mri. *PLoS One* 2**,** e597.

Halperin, J.M., & Schulz, K.P. (2006). Revisiting the role of the prefrontal cortex in the pathophysiology of attention-deficit/hyperactivity disorder. *Psychol Bull* 132**,** 560-581.

Hanley, J.A., & Mcneil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology* 143**,** 29-36.

Hannestad, J., Gallezot, J.D., Planeta-Wilson, B., Lin, S.F., Williams, W.A., Van Dyck, C.H., Malison, R.T., Carson, R.E., & Ding, Y.S. (2010). Clinically relevant doses of methylphenidate significantly occupy norepinephrine transporters in humans in vivo. *Biol Psychiatry* 68**,** 854-860.

Hansen, L.K., & Salamon, P. (1990). Neural network ensembles. *IEEE Trans. Pattern Anal. Mach. Intell.* 12**,** 993-1001.

Harkness, A.R., Reynolds, S.M., & Lilienfeld, S.O. (2014). A review of systems for psychology and psychiatry: Adaptive systems, personality psychopathology five (psy-5), and the dsm-5. *J Pers Assess* 96**,** 121-139.

Hart, H., Chantiluke, K., Cubillo, A.I., Smith, A.B., Simmons, A., Brammer, M.J., Marquand, A.F., & Rubia, K. (2014). Pattern classification of response inhibition in adhd: Toward the development of neurobiological markers for adhd. *Hum Brain Mapp* 35**,** 3083-3094.

Hart, H., Radua, J., Nakao, T., Mataix-Cols, D., & Rubia, K. (2013). Meta-analysis of functional magnetic resonance imaging studies of inhibition and attention in attention-deficit/hyperactivity disorder: Exploring task-specific, stimulant medication, and age effects. *JAMA Psychiatry* 70**,** 185-198.

He, Y., Chen, Z.J., & Evans, A.C. (2007). Small-world anatomical networks in the human brain revealed by cortical thickness from mri. *Cereb Cortex* 17**,** 2407-2419.

He, Y., & Evans, A. (2010). Graph theoretical modeling of brain connectivity. *Curr Opin Neurol* 23**,** 341-350.

Hilgetag, C.C., & Kaiser, M. (2004). Clustered organization of cortical connectivity. *Neuroinformatics* 2**,** 353-360.

Hodgkins, P., Shaw, M., Coghill, D., & Hechtman, L. (2012). Amfetamine and methylphenidate medications for attention-deficit/hyperactivity disorder: Complementary treatment options. *Eur Child Adolesc Psychiatry* 21**,** 477-492.

Hoerl, A.E., & Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12**,** 55-67.

Iannaccone, R., Hauser, T.U., Ball, J., Brandeis, D., Walitza, S., & Brem, S. (2015). Classifying adolescent attention-deficit/hyperactivity disorder (adhd) based on functional and structural imaging. *Eur Child Adolesc Psychiatry* 24**,** 1279-1289.

Iraji, A., Calhoun, V.D., Wiseman, N.M., Davoodi-Bojd, E., Avanaki, M.R.N., Haacke, E.M., & Kou, Z. (2016). The connectivity domain: Analyzing resting state fmri data using feature-based data-driven and model-based methods. *Neuroimage* 134**,** 494-507.

Jarbo, K., & Verstynen, T.D. (2015). Converging structural and functional connectivity of orbitofrontal, dorsolateral prefrontal, and posterior parietal cortex in the human striatum. *J Neurosci* 35**,** 3865-3878.

Johansson, S., Halleland, H., Halmoy, A., Jacobsen, K.K., Landaas, E.T., Dramsdahl, M., Fasmer, O.B., Bergsholm, P., Lundervold, A.J., Gillberg, C., Hugdahl, K., Knappskog, P.M., & Haavik, J. (2008). Genetic analyses of dopamine related genes in adult adhd patients suggest an association with the drd5-microsatellite repeat, but not with drd4 or slc6a3 vntrs. *Am J Med Genet B Neuropsychiatr Genet* 147B**,** 1470-1475.

Johnston, B.A., Mwangi, B., Matthews, K., Coghill, D., Konrad, K., & Steele, J.D. (2014). Brainstem abnormalities in attention deficit hyperactivity disorder support high accuracy individual diagnostic classification. *Hum Brain Mapp* 35**,** 5179-5189.

Kambeitz, J., Romanos, M., & Ettinger, U. (2014). Meta-analysis of the association between dopamine transporter genotype and response to methylphenidate treatment in adhd. *Pharmacogenomics J* 14**,** 77-84.

Karalunas, S.L., Fair, D., Musser, E.D., Aykes, K., Iyer, S.P., & Nigg, J.T. (2014). Subtyping attention-deficit/hyperactivity disorder using temperament dimensions: Toward biologically based nosologic criteria. *JAMA Psychiatry* 71**,** 1015-1024.

Katzman, M.A., Bilkey, T.S., Chokka, P.R., Fallu, A., & Klassen, L.J. (2017). Adult adhd and comorbid disorders: Clinical implications of a dimensional approach. *BMC Psychiatry* 17**,** 302.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (k-sads-pl): Initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 36**,** 980-988.

Klein, M., Onnink, M., Van Donkelaar, M., Wolfers, T., Harich, B., Shi, Y., Dammers, J., Arias-Vasquez, A., Hoogman, M., & Franke, B. (2017). Brain imaging genetics in adhd and beyond - mapping pathways from gene to disorder at different levels of complexity. *Neurosci Biobehav Rev* 80**,** 115-155.

Konrad, K., Neufang, S., Hanisch, C., Fink, G.R., & Herpertz-Dahlmann, B. (2006). Dysfunctional attentional networks in children with attention deficit/hyperactivity disorder: Evidence from an event-related functional magnetic resonance imaging study. *Biol Psychiatry* 59**,** 643-651.

Kuang, D., & He, L. (2014). Classification on adhd with deep learning. *2014 International Conference on Cloud Computing and Big Data***,** 6.

Lahey, B.B., D'onofrio, B.M., & Waldman, I.D. (2009). Using epidemiologic methods to test hypotheses regarding causal influences on child and adolescent mental disorders. *J Child Psychol Psychiatry* 50**,** 53-62.

Lam, L., & Suen, C.Y. (1997). Application of majority voting to pattern recognition: An analysis of its behavior and performance. *IEEE T Syst Man Cy B* 27**,** 16.

Lasky-Su, J., Banaschewski, T., Buitelaar, J., Franke, B., Brookes, K., Sonuga-Barke, E., Ebstein, R., Eisenberg, J., Gill, M., Manor, I., Miranda, A., Mulas, F., Oades, R.D., Roeyers, H., Rothenberger, A., Sergeant, J., Steinhausen, H.C., Taylor, E., Zhou, K., Thompson, M., Asherson, P., & Faraone, S.V. (2007). Partial replication of a drd4 association in adhd individuals using a statistically derived quantitative trait for adhd in a family-based association test. *Biol Psychiatry* 62**,** 985-990.

Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Phys Rev Lett* 87**,** 198701.

Laucht, M., Hohm, E., Esser, G., Schmidt, M.H., & Becker, K. (2007). Association between adhd and smoking in adolescence: Shared genetic, environmental and psychopathological factors. *J Neural Transm (Vienna)* 114**,** 1097-1104.

Li, X., Branch, C., De La Fuente, A., & Xia, S. (2013). Role of pulvinar-cortical functional brain pathways in attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 52**,** 756-758.

Li, X., Jiang, J., Zhu, W., Yu, C., Sui, M., Wang, Y., & Jiang, T. (2007). Asymmetry of prefrontal cortical convolution complexity in males with attention-deficit/hyperactivity disorder using fractal information dimension. *Brain Dev* 29**,** 649-655.

Li, X., Sroubek, A., Kelly, M.S., Lesser, I., Sussman, E., He, Y., Branch, C., & Foxe, J.J. (2012a). Atypical pulvinar-cortical pathways during sustained attention performance in children with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 51**,** 1197-1207 e1194.

Li, X., Xia, S., Bertisch, H.C., Branch, C.A., & Delisi, L.E. (2012b). Unique topology of language processing brain network: A systems-level biomarker of schizophrenia. *Schizophr Res* 141**,** 128-136.

Liao, X., Vasilakos, A.V., & He, Y. (2017). Small-world human brain networks: Perspectives and challenges. *Neurosci Biobehav Rev* 77**,** 286-300.

Lifford, K.J., Harold, G.T., & Thapar, A. (2009). Parent-child hostility and child adhd symptoms: A genetically sensitive and longitudinal analysis. *J Child Psychol Psychiatry* 50**,** 1468-1476.

Lilienfeld, S.O., & Treadway, M.T. (2016). Clashing diagnostic approaches: Dsm-icd versus rdoc. *Annu Rev Clin Psychol* 12**,** 435-463.

Lim, L., Marquand, A., Cubillo, A.A., Smith, A.B., Chantiluke, K., Simmons, A., Mehta, M., & Rubia, K. (2013). Disorder-specific predictive classification of adolescents with attention deficit hyperactivity disorder (adhd) relative to autism using structural magnetic resonance imaging. *PLoS One* 8**,** e63660.

Lin, H.Y., Tseng, W.Y., Lai, M.C., Matsuo, K., & Gau, S.S. (2015). Altered resting-state frontoparietal control network in children with attention-deficit/hyperactivity disorder. *J Int Neuropsychol Soc* 21**,** 271-284.

Liu, T., Chen, Y., Lin, P., & Wang, J. (2015). Small-world brain functional networks in children with attention-deficit/hyperactivity disorder revealed by eeg synchrony. *Clin EEG Neurosci* 46**,** 183-191.

Loney, J., & Milich, R. (1982). Hyperactivity, inattention and aggression in clinical practice,. In wolraich m. & r. D. (eds.). *Advances in Developmental and Behavioral pediatrics* 2**,** 113-147.

Luo, Y., Alvarez, T.L., Halperin, J.M., & Li, X. (2020a). Multimodal neuroimaging-based prediction of adult outcomes in childhood-onset adhd using ensemble learning techniques. *Neuroimage Clin* 26**,** 102238.

Luo, Y., Halperin, J.M., & Li, X. (2020b). Anatomical substrates of symptom remission and persistence in young adults with childhood attention deficit/hyperactivity disorder. *Eur Neuropsychopharmacol* 33**,** 117-125.

Luo, Y., Schulz, K.P., Alvarez, T.L., Halperin, J.M., & Li, X. (2018). Distinct topological properties of cue-evoked attention processing network in persisters and remitters of childhood adhd. *Cortex* 109**,** 234-244.

Luo, Y., Weibman, D., Halperin, J.M., & Li, X. (2019). A review of heterogeneity in attention deficit/hyperactivity disorder (adhd). *Front Hum Neurosci* 13.

Mahone, E.M., Ranta, M.E., Crocetti, D., O'brien, J., Kaufmann, W.E., Denckla, M.B., & Mostofsky, S.H. (2011). Comprehensive examination of frontal regions in boys and girls with attention-deficit/hyperactivity disorder. *J Int Neuropsychol Soc* 17**,** 1047-1057.

Makris, N., Buka, S.L., Biederman, J., Papadimitriou, G.M., Hodge, S.M., Valera, E.M., Brown, A.B., Bush, G., Monuteaux, M.C., Caviness, V.S., Kennedy, D.N., & Seidman, L.J. (2008). Attention and executive systems abnormalities in adults with childhood adhd: A dt-mri study of connections. *Cereb Cortex* 18**,** 1210-1220.

Mao, A.R., & Findling, R.L. (2014). Comorbidities in adult attention-deficit/hyperactivity disorder: A practical guide to diagnosis in primary care. *Postgrad Med* 126**,** 42-51.

Martinussen, R., Hayden, J., Hogg-Johnson, S., & Tannock, R. (2005). A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 44**,** 377-384.

Martinussen, R., & Tannock, R. (2006). Working memory impairments in children with attention-deficit hyperactivity disorder with and without comorbid language learning disorders. *J Clin Exp Neuropsychol* 28**,** 1073-1094.

Mattfeld, A.T., Gabrieli, J.D., Biederman, J., Spencer, T., Brown, A., Kotte, A., Kagan, E., & Whitfield-Gabrieli, S. (2014). Brain differences between persistent and remitted attention deficit hyperactivity disorder. *Brain* 137**,** 2423-2428.

Mawjee, K., Woltering, S., Lai, N., Gotlieb, H., Kronitz, R., & Tannock, R. (2017). Working memory training in adhd: Controlling for engagement, motivation, and expectancy of improvement (pilot study). *J Atten Disord* 21**,** 956-968.

Max, J.E., Fox, P.T., Lancaster, J.L., Kochunov, P., Mathews, K., Manes, F.F., Robertson, B.A., Arndt, S., Robin, D.A., & Lansing, A.E. (2002). Putamen lesions and the development of attention-deficit/hyperactivity symptomatology. *J Am Acad Child Adolesc Psychiatry* 41**,** 563-571.

Mccracken, J.T., Smalley, S.L., Mcgough, J.J., Crawford, L., Del'homme, M., Cantor, R.M., Liu, A., & Nelson, S.F. (2000). Evidence for linkage of a tandem duplication polymorphism upstream of the dopamine d4 receptor gene (drd4) with attention deficit hyperactivity disorder (adhd). *Mol Psychiatry* 5**,** 531-536.

Meunier, D., Fonlupt, P., Saive, A.L., Plailly, J., Ravel, N., & Royet, J.P. (2014). Modular structure of functional networks in olfactory memory. *Neuroimage* 95**,** 264-275.

Meunier, D., Lambiotte, R., & Bullmore, E.T. (2010). Modular and hierarchically modular organization of brain networks. *Front Neurosci* 4**,** 200.

Mick, E., Biederman, J., Prince, J., Fischer, M.J., & Faraone, S.V. (2002). Impact of low birth weight on attention-deficit hyperactivity disorder. *J Dev Behav Pediatr* 23**,** 16-22.

Moffitt, T.E., Houts, R., Asherson, P., Belsky, D.W., Corcoran, D.L., Hammerle, M., Harrington, H., Hogan, S., Meier, M.H., Polanczyk, G.V., Poulton, R., Ramrakha, S., Sugden, K., Williams, B., Rohde, L.A., & Caspi, A. (2015). Is adult adhd a childhood-onset neurodevelopmental disorder? Evidence from a four-decade longitudinal cohort study. *Am J Psychiatry* 172**,** 967-977.

Molina, B.S., Hinshaw, S.P., Swanson, J.M., Arnold, L.E., Vitiello, B., Jensen, P.S., Epstein, J.N., Hoza, B., Hechtman, L., Abikoff, H.B., Elliott, G.R., Greenhill, L.L., Newcorn, J.H., Wells, K.C., Wigal, T., Gibbons, R.D., Hur, K., Houck, P.R., & Group, M.T.a.C. (2009). The mta at 8 years: Prospective follow-up of children treated for combined-type adhd in a multisite study. *J Am Acad Child Adolesc Psychiatry* 48**,** 484-500.

Nagel, B.J., Bathula, D., Herting, M., Schmitt, C., Kroenke, C.D., Fair, D., & Nigg, J.T. (2011). Altered white matter microstructure in children with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 50**,** 283-292.

Nakao, T., Radua, J., Rubia, K., & Mataix-Cols, D. (2011). Gray matter volume abnormalities in adhd: Voxel-based meta-analysis exploring the effects of age and stimulant medication. *Am J Psychiatry* 168**,** 1154-1163.

Neale, B.M., Medland, S.E., Ripke, S., Asherson, P., Franke, B., Lesch, K.P., Faraone, S.V., Nguyen, T.T., Schafer, H., Holmans, P., Daly, M., Steinhausen, H.C., Freitag, C., Reif, A., Renner, T.J., Romanos, M., Romanos, J., Walitza, S., Warnke, A., Meyer, J., Palmason, H., Buitelaar, J., Vasquez, A.A., Lambregts-Rommelse, N., Gill, M., Anney, R.J., Langely, K., O'donovan, M., Williams, N., Owen, M., Thapar, A., Kent, L., Sergeant, J., Roeyers, H., Mick, E., Biederman, J., Doyle, A., Smalley, S., Loo, S., Hakonarson, H., Elia, J., Todorov, A., Miranda, A., Mulas, F., Ebstein, R.P., Rothenberger, A., Banaschewski, T., Oades, R.D., Sonuga-Barke, E., Mcgough, J., Nisenbaum, L., Middleton, F., Hu, X., Nelson, S., & Psychiatric, G.C.a.S. (2010). Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 49**,** 884-897.

Neuman, R.J., Lobos, E., Reich, W., Henderson, C.A., Sun, L.W., & Todd, R.D. (2007). Prenatal smoking exposure and dopaminergic genotypes interact to cause a severe adhd subtype. *Biol Psychiatry* 61**,** 1320-1328.

Newcorn, J.H., Kratochvil, C.J., Allen, A.J., Casat, C.D., Ruff, D.D., Moore, R.J., Michelson, D., & Atomoxetine/Methylphenidate Comparative Study, G. (2008). Atomoxetine and osmotically released methylphenidate for the treatment of attention deficit hyperactivity disorder: Acute comparison and differential response. *Am J Psychiatry* 165**,** 721-730.

Nigg, J., Nikolas, M., & Burt, S.A. (2010). Measured gene-by-environment interaction in relation to attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 49**,** 863-873.

Nigg, J.T. (2008). Adhd, lead exposure and prevention: How much lead or how much evidence is needed? *Expert Rev Neurother* 8**,** 519-521.

Nigg, J.T., Stavro, G., Ettenhofer, M., Hambrick, D.Z., Miller, T., & Henderson, J.M. (2005). Executive functions and adhd in adults: Evidence for selective effects on adhd symptom domains. *J Abnorm Psychol* 114**,** 706-717.

Noordermeer, S.D.S., Luman, M., Greven, C.U., Veroude, K., Faraone, S.V., Hartman, C.A., Hoekstra, P.J., Franke, B., Buitelaar, J.K., Heslenfeld, D.J., & Oosterlaan, J. (2017). Structural brain abnormalities of attention-deficit/hyperactivity disorder with oppositional defiant disorder. *Biol Psychiatry* 82**,** 642-650.

Ogdie, M.N., Fisher, S.E., Yang, M., Ishii, J., Francks, C., Loo, S.K., Cantor, R.M., Mccracken, J.T., Mcgough, J.J., Smalley, S.L., & Nelson, S.F. (2004). Attention deficit hyperactivity disorder: Fine mapping supports linkage to 5p13, 6q12, 16p13, and 17p11. *Am J Hum Genet* 75**,** 661-668.

Onnink, A.M., Zwiers, M.P., Hoogman, M., Mostert, J.C., Kan, C.C., Buitelaar, J., & Franke, B. (2014). Brain alterations in adult adhd: Effects of gender, treatment and comorbid depression. *Eur Neuropsychopharmacol* 24**,** 397-409.

Pakkenberg, B., Pelvig, D., Marner, L., Bundgaard, M.J., Gundersen, H.J., Nyengaard, J.R., & Regeur, L. (2003). Aging and the human neocortex. *Exp Gerontol* 38**,** 95-99.

Park, H.J., & Friston, K. (2013). Structural and functional brain networks: From connections to cognition. *Science* 342**,** 1238411.

Pastor, P., Reuben, C., Duran, C., & Hawkins, L. (2015). Association between diagnosed adhd and selected characteristics among children aged 4-17 years: United states, 2011-2013. *NCHS Data Brief***,** 201.

Pelham, W.E., Jr., Fabiano, G.A., Waxmonsky, J.G., Greiner, A.R., Gnagy, E.M., Pelham, W.E., 3rd, Coxe, S., Verley, J., Bhatia, I., Hart, K., Karch, K., Konijnendijk, E., Tresco, K., Nahum-Shani, I., & Murphy, S.A. (2016). Treatment sequencing for childhood adhd: A multiple-randomization study of adaptive medication and behavioral interventions. *J Clin Child Adolesc Psychol* 45**,** 396-415.

Peng, X., Lin, P., Zhang, T., & Wang, J. (2013). Extreme learning machine-based classification of adhd using brain structural mri data. *PLoS One* 8**,** e79476.

Perennou, D.A., Leblond, C., Amblard, B., Micallef, J.P., Rouget, E., & Pelissier, J. (2000). The polymodal sensory cortex is crucial for controlling lateral postural stability: Evidence from stroke patients. *Brain Res Bull* 53**,** 359-365.

Perou, R., Bitsko, R.H., Blumberg, S.J., Pastor, P., Ghandour, R.M., Gfroerer, J.C., Hedden, S.L., Crosby, A.E., Visser, S.N., Schieve, L.A., Parks, S.E., Hall, J.E., Brody, D., Simile, C.M., Thompson, W.W., Baio, J., Avenevoli, S., Kogan, M.D., Huang, L.N., Centers for Disease, C., & Prevention (2013). Mental health surveillance among children--united states, 2005-2011. *MMWR Suppl* 62**,** 1-35.

Peterson, D.J., Ryan, M., Rimrodt, S.L., Cutting, L.E., Denckla, M.B., Kaufmann, W.E., & Mahone, E.M. (2011). Increased regional fractional anisotropy in highly screened attention-deficit hyperactivity disorder (adhd). *J Child Neurol* 26**,** 1296-1302.

Pheula, G.F., Rohde, L.A., & Schmitz, M. (2011). Are family variables associated with adhd, inattentive type? A case-control study in schools. *Eur Child Adolesc Psychiatry* 20**,** 137-145.

Pineda, D.A., Palacio, L.G., Puerta, I.C., Merchan, V., Arango, C.P., Galvis, A.Y., Gomez, M., Aguirre, D.C., Lopera, F., & Arcos-Burgos, M. (2007). Environmental influences that affect attention deficit/hyperactivity disorder: Study of a genetic isolate. *Eur Child Adolesc Psychiatry* 16**,** 337-346.

Pliszka, S.R. (2007). Pharmacologic treatment of attention-deficit/hyperactivity disorder: Efficacy, safety and mechanisms of action. *Neuropsychol Rev* 17**,** 61-72.

Pliszka, S.R., Glahn, D.C., Semrud-Clikeman, M., Franklin, C., Perez, R., 3rd, Xiong, J., & Liotti, M. (2006). Neuroimaging of inhibitory control areas in children with attention deficit hyperactivity disorder who were treatment naive or in long-term treatment. *Am J Psychiatry* 163**,** 1052-1060.

Polanczyk, G., De Lima, M.S., Horta, B.L., Biederman, J., & Rohde, L.A. (2007). The worldwide prevalence of adhd: A systematic review and metaregression analysis. *Am J Psychiatry* 164**,** 942-948.

Posner, J., Park, C., & Wang, Z. (2014). Connecting the dots: A review of resting connectivity mri studies in attention-deficit/hyperactivity disorder. *Neuropsychol Rev* 24**,** 3-15.

Posner, M.I., & Rothbart, M.K. (2009). Toward a physical basis of attention and self regulation. *Phys Life Rev* 6**,** 103-120.

Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., & Petersen, S.E. (2012). Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *Neuroimage* 59**,** 2142-2154.

Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., & Petersen, S.E. (2011). Functional network organization of the human brain. *Neuron* 72**,** 665-678.

Proal, E., Reiss, P.T., Klein, R.G., Mannuzza, S., Gotimer, K., Ramos-Olazagasti, M.A., Lerch, J.P., He, Y., Zijdenbos, A., Kelly, C., Milham, M.P., & Castellanos, F.X. (2011). Brain gray matter deficits at 33-year follow-up in adults with attention-deficit/hyperactivity disorder established in childhood. *Arch Gen Psychiatry* 68**,** 1122-1134.

Qiu, M.G., Ye, Z., Li, Q.Y., Liu, G.J., Xie, B., & Wang, J. (2011). Changes of brain structure and function in adhd children. *Brain Topogr* 24**,** 243-252.

Quinn, P.O., & Madhoo, M. (2014). A review of attention-deficit/hyperactivity disorder in women and girls: Uncovering this hidden diagnosis. *Prim Care Companion CNS Disord* 16.

Qureshi, M.N., Min, B., Jo, H.J., & Lee, B. (2016). Multiclass classification for the differential diagnosis on the adhd subtypes using recursive feature elimination and hierarchical extreme learning machine: Structural mri study. *PLoS One* 11**,** e0160697.

Qureshi, M.N.I., Oh, J., Min, B., Jo, H.J., & Lee, B. (2017). Multi-modal, multi-measure, and multi-class discrimination of adhd with hierarchical feature extraction and extreme learning machine using structural and functional brain mri. *Front Hum Neurosci* 11**,** 157.

Rajendran, K., Trampush, J.W., Rindskopf, D., Marks, D.J., O'neill, S., & Halperin, J.M. (2013). Association between variation in neuropsychological development and trajectory of adhd severity in early childhood. *Am J Psychiatry* 170**,** 1205-1211.

Ring, H.A., & Serra-Mestres, J. (2002). Neuropsychiatry of the basal ganglia. *J Neurol Neurosurg Psychiatry* 72**,** 12-21.

Rissman, J., Gazzaley, A., & D'esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage* 23**,** 752-763.

Rogers, M., Hwang, H., Toplak, M., Weiss, M., & Tannock, R. (2011). Inattention, working memory, and academic achievement in adolescents referred for attention deficit/hyperactivity disorder (adhd). *Child Neuropsychol* 17**,** 444-458.

Rosas, H.D., Liu, A.K., Hersch, S., Glessner, M., Ferrante, R.J., Salat, D.H., Van Der Kouwe, A., Jenkins, B.G., Dale, A.M., & Fischl, B. (2002). Regional and progressive thinning of the cortical ribbon in huntington's disease. *Neurology* 58**,** 695-701.

Rosler, M., Retz, W., Fischer, R., Ose, C., Alm, B., Deckert, J., Philipsen, A., Herpertz, S., & Ammer, R. (2010). Twenty-four-week treatment with extended release methylphenidate improves emotional symptoms in adult adhd. *World J Biol Psychiatry* 11**,** 709-718.

Rubia, K., Criaud, M., Wulff, M., Alegria, A., Brinson, H., Barker, G., Stahl, D., & Giampietro, V. (2019). Functional connectivity changes associated with fmri neurofeedback of right inferior frontal cortex in adolescents with adhd. *Neuroimage* 188**,** 43-58.

Rubia, K., Overmeyer, S., Taylor, E., Brammer, M., Williams, S.C., Simmons, A., & Bullmore, E.T. (1999). Hypofrontality in attention deficit hyperactivity disorder during higher-order motor control: A study with functional mri. *Am J Psychiatry* 156**,** 891-896.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* 52**,** 1059-1069.

Ruta, D., & Gabrys, B. (2005). Classifier selection for majority voting. *Inform Fusion* 6**,** 19.

Saez, M., Barcelo, M.A., Farrerons, M., & Lopez-Casasnovas, G. (2018). The association between exposure to environmental factors and the occurrence of attention-deficit/hyperactivity disorder (adhd). A population-based retrospective cohort study. *Environ Res* 166**,** 205-214.

Salvador, R., Suckling, J., Coleman, M.R., Pickard, J.D., Menon, D., & Bullmore, E. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb Cortex* 15**,** 1332-1342.

Samu, D., Seth, A.K., & Nowotny, T. (2014). Influence of wiring cost on the large-scale architecture of human cortical connectivity. *PLoS Comput Biol* 10**,** e1003557.

Santosa, F., & Symes, W.W. (1986). Linear inversion of band-limited reflection seismograms. *SIAM J Sci and Stat Comput* 7**,** 1307-1330.

Schapire, R.E. (1990). The strength of weak learnability. *Machine Learning* 5**,** 197-227.

Schneider, M.F., Krick, C.M., Retz, W., Hengesch, G., Retz-Junginger, P., Reith, W., & Rosler, M. (2010). Impairment of fronto-striatal and parietal cerebral networks correlates with attention deficit hyperactivity disorder (adhd) psychopathology in adults - a functional magnetic resonance imaging (fmri) study. *Psychiatry Res* 183**,** 75-84.

Schulz, K.P., Li, X., Clerkin, S.M., Fan, J., Berwid, O.G., Newcorn, J.H., & Halperin, J.M. (2017). Prefrontal and parietal correlates of cognitive control related to the adult outcome of attention-deficit/hyperactivity disorder diagnosed in childhood. *Cortex* 90**,** 1-11.

Schulz, K.P., Newcorn, J.H., Fan, J., Tang, C.Y., & Halperin, J.M. (2005). Brain activation gradients in ventrolateral prefrontal cortex related to persistence of adhd in adolescent boys. *J Am Acad Child Adolesc Psychiatry* 44**,** 47-54.

Schwenke, E., Fasching, P.A., Faschingbauer, F., Pretscher, J., Kehl, S., Peretz, R., Keller, A., Haberle, L., Eichler, A., Irlbauer-Muller, V., Dammer, U., Beckmann, M.W., & Schneider, M. (2018). Predicting attention deficit hyperactivity disorder using pregnancy and birth characteristics. *Arch Gynecol Obstet*.

Seidman, L.J., Biederman, J., Liang, L., Valera, E.M., Monuteaux, M.C., Brown, A., Kaiser, J., Spencer, T., Faraone, S.V., & Makris, N. (2011). Gray matter alterations in adults with attention-deficit/hyperactivity disorder identified by voxel based morphometry. *Biol Psychiatry* 69**,** 857-866.

Seidman, L.J., Valera, E.M., Makris, N., Monuteaux, M.C., Boriel, D.L., Kelkar, K., Kennedy, D.N., Caviness, V.S., Bush, G., Aleardi, M., Faraone, S.V., & Biederman, J. (2006). Dorsolateral prefrontal and anterior cingulate cortex volumetric abnormalities in adults with attention-deficit/hyperactivity disorder identified by magnetic resonance imaging. *Biol Psychiatry* 60**,** 1071-1080.

Sen, B., Borle, N.C., Greiner, R., & Brown, M.R.G. (2018). A general prediction model for the detection of adhd and autism using structural and functional mri. *PLoS One* 13**,** e0194856.

Sergeant, J. (2000). The cognitive-energetic model: An empirical approach to attention-deficit hyperactivity disorder. *Neurosci Biobehav Rev* 24**,** 7-12.

Shaffer, D., Fisher, P., Piacentini, J., Schwab-Stone, M., & Wicks, J. (1989). Diagnostic interview schedule for children-parent version (disc-2.1p).

Shang, C.Y., Wu, Y.H., Gau, S.S., & Tseng, W.Y. (2013). Disturbed microstructural integrity of the frontostriatal fiber pathways and executive dysfunction in children with attention deficit hyperactivity disorder. *Psychol Med* 43**,** 1093-1107.

Shaw, P., Gogtay, N., & Rapoport, J. (2010). Childhood psychiatric disorders as anomalies in neurodevelopmental trajectories. *Hum Brain Mapp* 31**,** 917-925.

Shaw, P., Lerch, J., Greenstein, D., Sharp, W., Clasen, L., Evans, A., Giedd, J., Castellanos, F.X., & Rapoport, J. (2006). Longitudinal mapping of cortical thickness and clinical outcome in children and adolescents with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* 63**,** 540-549.

Shaw, P., Malek, M., Watson, B., Greenstein, D., De Rossi, P., & Sharp, W. (2013). Trajectories of cerebral cortical development in childhood and adolescence and adult attention-deficit/hyperactivity disorder. *Biol Psychiatry* 74**,** 599-606.

Shaw, P., Stringaris, A., Nigg, J., & Leibenluft, E. (2014). Emotion dysregulation in attention deficit hyperactivity disorder. *American J Psychiatry* 171**,** 276-293.

Shaw, P., Sudre, G., Wharton, A., Weingart, D., Sharp, W., & Sarlls, J. (2015). White matter microstructure and the variable adult outcome of childhood attention deficit hyperactivity disorder. *Neuropsychopharmacology* 40**,** 746-754.

Shehzad, Z., Kelly, A.M., Reiss, P.T., Gee, D.G., Gotimer, K., Uddin, L.Q., Lee, S.H., Margulies, D.S., Roy, A.K., Biswal, B.B., Petkova, E., Castellanos, F.X., & Milham, M.P. (2009). The resting brain: Unconstrained yet reliable. *Cereb Cortex* 19**,** 2209-2229.

Shum, S.B., & Pang, M.Y. (2009). Children with attention deficit hyperactivity disorder have impaired balance function: Involvement of somatosensory, visual, and vestibular systems. *J Pediatr* 155**,** 245-249.

Sibley, M.H., Kuriyan, A.B., Evans, S.W., Waxmonsky, J.G., & Smith, B.H. (2014). Pharmacological and psychosocial treatments for adolescents with adhd: An updated systematic review of the literature. *Clin Psychol Rev* 34**,** 218-232.

Sibley, M.H., Rohde, L.A., Swanson, J.M., Hechtman, L.T., Molina, B.S.G., Mitchell, J.T., Arnold, L.E., Caye, A., Kennedy, T.M., Roy, A., Stehli, A., & Multimodal Treatment Study of Children With, A.C.G. (2018). Late-onset adhd reconsidered with comprehensive repeated assessments between ages 10 and 25. *Am J Psychiatry* 175**,** 140-149.

Silk, T., Vance, A., Rinehart, N., Egan, G., O'boyle, M., Bradshaw, J.L., & Cunnington, R. (2005). Fronto-parietal activation in attention-deficit hyperactivity disorder, combined type: Functional magnetic resonance imaging study. *Br J Psychiatry* 187**,** 282-283.

Silva, D., Colvin, L., Hagemann, E., & Bower, C. (2014). Environmental risk factors by gender associated with attention-deficit/hyperactivity disorder. *Pediatrics* 133**,** e14-22.

Simone, A.N., Marks, D.J., Bedard, A.C., & Halperin, J.M. (2018). Low working memory rather than adhd symptoms predicts poor academic achievement in school-aged children. *J Abnorm Child Psychol* 46**,** 277-290.

Smith, A.B., Taylor, E., Brammer, M., Toone, B., & Rubia, K. (2006). Task-specific hypoactivation in prefrontal and temporoparietal brain regions during motor inhibition and task switching in medication-naive children and adolescents with attention deficit hyperactivity disorder. *Am J Psychiatry* 163**,** 1044-1051.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., & Matthews, P.M. (2004). Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage* 23 Suppl 1**,** S208-219.

Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., & De Stefano, N. (2002). Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage* 17**,** 479-489.

Solanto, M.V. (2002). Dopamine dysfunction in ad/hd: Integrating clinical and basic neuroscience research. *Behav Brain Res* 130**,** 65-71.

Sonuga-Barke, E.J. (2002). Psychological heterogeneity in ad/hd--a dual pathway model of behaviour and cognition. *Behav Brain Res* 130**,** 29-36.

Sonuga-Barke, E.J. (2003). The dual pathway model of ad/hd: An elaboration of neuro-developmental characteristics. *Neurosci Biobehav Rev* 27**,** 593-604.

Spencer, T., Biederman, J., Wilens, T., Harding, M., O'donnell, D., & Griffin, S. (1996). Pharmacotherapy of attention-deficit hyperactivity disorder across the life cycle. *J Am Acad Child Adolesc Psychiatry* 35**,** 409-432.

Sporns, O., Chialvo, D.R., Kaiser, M., & Hilgetag, C.C. (2004). Organization, development and function of complex brain networks. *Trends Cogn Sci* 8**,** 418-425.

Sporns, O., Honey, C.J., & Kotter, R. (2007). Identification and classification of hubs in brain networks. *PLoS One* 2**,** e1049.

Sprich, S., Biederman, J., Crawford, M.H., Mundy, E., & Faraone, S.V. (2000). Adoptive and biological families of children and adolescents with adhd. *J Am Acad Child Adolesc Psychiatry* 39**,** 1432-1437.

Stevens, S.E., Kumsta, R., Kreppner, J.M., Brookes, K.J., Rutter, M., & Sonuga-Barke, E.J. (2009). Dopamine transporter gene polymorphism moderates the effects of severe deprivation on adhd symptoms: Developmental continuities in gene-environment interplay. *Am J Med Genet B Neuropsychiatr Genet* 150B**,** 753-761.

Suades-Gonzalez, E., Forns, J., Garcia-Esteban, R., Lopez-Vicente, M., Esnaola, M., Alvarez-Pedrerol, M., Julvez, J., Caceres, A., Basagana, X., Lopez-Sala, A., & Sunyer, J. (2017). A longitudinal study on attention development in primary school children with and without teacher-reported symptoms of adhd. *Front Psychol* 8**,** 655.

Sun, D., Van Erp, T.G., Thompson, P.M., Bearden, C.E., Daley, M., Kushan, L., Hardt, M.E., Nuechterlein, K.H., Toga, A.W., & Cannon, T.D. (2009). Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: Classification analysis using probabilistic brain atlas and machine learning algorithms. *Biol Psychiatry* 66**,** 1055-1060.

Supekar, K., Musen, M., & Menon, V. (2009). Development of large-scale functional brain networks in children. *PLoS Biol* 7**,** e1000157.

Suskauer, S.J., Simmonds, D.J., Fotedar, S., Blankner, J.G., Pekar, J.J., Denckla, M.B., & Mostofsky, S.H. (2008). Functional magnetic resonance imaging evidence for abnormalities in response selection in attention deficit hyperactivity disorder: Differences in activation associated with response inhibition but not habitual motor response. *J Cogn Neurosci* 20**,** 478-493.

Szekely, E., Sudre, G.P., Sharp, W., Leibenluft, E., & Shaw, P. (2017). Defining the neural substrate of the adult outcome of childhood adhd: A multimodal neuroimaging study of response inhibition. *Am J Psychiatry* 174**,** 867-876.

Tajima-Pozo, K., Yus, M., Ruiz-Manrique, G., Lewczuk, A., Arrazola, J., & Montanes-Rada, F. (2018). Amygdala abnormalities in adults with adhd. *J Atten Disord* 22**,** 671-678.

Tannock, R., Ickowicz, A., & Schachar, R. (1995). Differential effects of methylphenidate on working memory in adhd children with and without comorbid anxiety. *J Am Acad Child Adolesc Psychiatry* 34**,** 886-896.

Tenev, A., Markovska-Simoska, S., Kocarev, L., Pop-Jordanov, J., Muller, A., & Candrian, G. (2014). Machine learning approach for classification of adhd adults. *Int J Psychophysiol* 93**,** 162-166.

Thakur, G.A., Grizenko, N., Sengupta, S.M., Schmitz, N., & Joober, R. (2010). The 5-httlpr polymorphism of the serotonin transporter gene and short term behavioral response to methylphenidate in children with adhd. *BMC Psychiatry* 10**,** 50.

Thapar, A., Cooper, M., Eyre, O., & Langley, K. (2013). What have we learnt about the causes of adhd? *J Child Psychol Psychiatry* 54**,** 3-16.

Thapar, A., Cooper, M., Jefferies, R., & Stergiakouli, E. (2012). What causes attention deficit hyperactivity disorder? *Arch Dis Child* 97**,** 260-265.

Thapar, A., & Rutter, M. (2009). Do prenatal risk factors cause psychiatric disorder? Be wary of causal claims. *Br J Psychiatry* 195**,** 100-101.

Thulborn, K.R., Waterton, J.C., Matthews, P.M., & Radda, G.K. (1982). Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochim Biophys Acta* 714**,** 265-270.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J R Stat Soc B* 58**,** 267-288.

Tomasi, D., Wang, R., Wang, G.J., & Volkow, N.D. (2014). Functional connectivity and brain activation: A synergistic approach. *Cereb Cortex* 24**,** 2619-2629.

Uchida, M., Spencer, T.J., Faraone, S.V., & Biederman, J. (2018). Adult outcome of adhd: An overview of results from the mgh longitudinal family studies of pediatrically and psychiatrically referred youth with and without adhd of both sexes. *J Atten Disord* 22**,** 523-534.

Ustinova, K.I., Chernikova, L.A., Ioffe, M.E., & Sliva, S.S. (2001). Impairment of learning the voluntary control of posture in patients with cortical lesions of different locations: The cortical mechanisms of pose regulation. *Neurosci Behav Physiol* 31**,** 259-267.

Vaidya, C.J., Bunge, S.A., Dudukovic, N.M., Zalecki, C.A., Elliott, G.R., & Gabrieli, J.D. (2005). Altered neural substrates of cognitive control in childhood adhd: Evidence from functional magnetic resonance imaging. *Am J Psychiatry* 162**,** 1605-1613.

Van Batenburg-Eddes, T., Brion, M.J., Henrichs, J., Jaddoe, V.W., Hofman, A., Verhulst, F.C., Lawlor, D.A., Davey Smith, G., & Tiemeier, H. (2013). Parental depressive and anxiety symptoms during pregnancy and attention problems in children: A cross-cohort consistency study. *J Child Psychol Psychiatry* 54**,** 591-600.

Van Den Heuvel, M.P., & Sporns, O. (2013). Network hubs in the human brain. *Trends Cogn Sci* 17**,** 683-696.

Van Den Heuvel, M.P., Stam, C.J., Boersma, M., & Hulshoff Pol, H.E. (2008). Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *Neuroimage* 43**,** 528-539.

Van Dessel, J., Sonuga-Barke, E., Mies, G., Lemiere, J., Van Der Oord, S., Morsink, S., & Danckaerts, M. (2018). Delay aversion in attention deficit/hyperactivity disorder is mediated by amygdala and prefrontal cortex hyper-activation. *J Child Psychol Psychiatry* 59**,** 888-899.

Van Dessel, J., Sonuga-Barke, E., Moerkerke, M., Van Der Oord, S., Lemiere, J., Morsink, S., & Danckaerts, M. (2019). The amygdala in adolescents with attention-deficit/hyperactivity disorder: Structural and functional correlates of delay aversion. *World J Biol Psychiatry***,** 1-12.

Vance, A., Silk, T.J., Casey, M., Rinehart, N.J., Bradshaw, J.L., Bellgrove, M.A., & Cunnington, R. (2007). Right parietal dysfunction in children with attention deficit hyperactivity disorder, combined type: A functional mri study. *Mol Psychiatry* 12**,** 826-832, 793.

Vertes, P.E., Alexander-Bloch, A.F., Gogtay, N., Giedd, J.N., Rapoport, J.L., & Bullmore, E.T. (2012). Simple models of human brain functional networks. *Proc Natl Acad Sci U S A* 109**,** 5868-5873.

Victor, M.M., Grevet, E.H., Salgado, C.A., Silva, K.L., Sousa, N.O., Karam, R.G., Vitola, E.S., Picon, F.A., Zeni, G.D., Contini, V., Rohde, L.A., Belmonte-De-Abreu, P., & Bau, C.H. (2009). Reasons for pretreatment attrition and dropout from methylphenidate in adults with attention-deficit/hyperactivity disorder: The role of comorbidities. *J Clin Psychopharmacol* 29**,** 614-616.

Visser, S.N., Danielson, M.L., Bitsko, R.H., Holbrook, J.R., Kogan, M.D., Ghandour, R.M., Perou, R., & Blumberg, S.J. (2014). Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United states, 2003-2011. *J Am Acad Child Adolesc Psychiatry* 53**,** 34-46 e32.

Volkow, N.D., Wang, G.J., Fowler, J.S., Logan, J., Franceschi, D., Maynard, L., Ding, Y.S., Gatley, S.J., Gifford, A., Zhu, W., & Swanson, J.M. (2002). Relationship between blockade of dopamine transporters by oral methylphenidate and the increases in extracellular dopamine: Therapeutic implications. *Synapse* 43**,** 181-187.

Wang, G., Hao, J., Ma, J., & Jiang, H. (2011). A comparative assessment of ensemble learning for credit scoring. *Expert Syst Appl* 38**,** 8.

Wang, L., Zhu, C., He, Y., Zang, Y., Cao, Q., Zhang, H., Zhong, Q., & Wang, Y. (2009). Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder. *Hum Brain Mapp* 30**,** 638-649.

Wang, X., Jiao, Y., Tang, T., Wang, H., & Lu, Z. (2013). Altered regional homogeneity patterns in adults with attention-deficit hyperactivity disorder. *Eur J Radiol* 82**,** 1552-1557.

Watts, D.J., & Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393**,** 440-442.

Wig, G.S., Laumann, T.O., Cohen, A.L., Power, J.D., Nelson, S.M., Glasser, M.F., Miezin, F.M., Snyder, A.Z., Schlaggar, B.L., & Petersen, S.E. (2014). Parcellating an individual subject's cortical and subcortical brain structures using snowball sampling of resting-state correlations. *Cereb Cortex* 24**,** 2036-2054.

Wilens, T.E., Biederman, J., Faraone, S.V., Martelon, M., Westerberg, D., & Spencer, T.J. (2009). Presenting adhd symptoms, subtypes, and comorbid disorders in clinically referred adults with adhd. *J Clin Psychiatry* 70**,** 1557-1562.

Wilens, T.E., Morrison, N.R., & Prince, J. (2011). An update on the pharmacotherapy of attention-deficit/hyperactivity disorder in adults. *Expert Rev Neurother* 11**,** 1443-1465.

Willcutt, E.G., Doyle, A.E., Nigg, J.T., Faraone, S.V., & Pennington, B.F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biol Psychiatry* 57**,** 1336-1346.

Winsberg, B.G., & Comings, D.E. (1999). Association of the dopamine transporter gene (dat1) with poor methylphenidate response. *J Am Acad Child Adolesc Psychiatry* 38**,** 1474-1477.

Wolpert, D.H. (1992). Stacked generalization. *Neural Networks* 5**,** 19.

Xia, S., Foxe, J.J., Sroubek, A.E., Branch, C., & Li, X. (2014). Topological organization of the "small-world" visual attention network in children with attention deficit/hyperactivity disorder (adhd). *Front Hum Neurosci* 8**,** 162.

Xia, S., Li, X., Kimball, A.E., Kelly, M.S., Lesser, I., & Branch, C. (2012). Thalamic shape and connectivity abnormalities in children with attention-deficit/hyperactivity disorder. *Psychiatry Res* 204**,** 161-167.

Xu, G., Strathearn, L., Liu, B., Yang, B., & Bao, W. (2018). Twenty-year trends in diagnosed attention-deficit/hyperactivity disorder among us children and adolescents, 1997-2016. *JAMA Netw Open* 1**,** e181471.

Yang, H., Wu, Q.Z., Guo, L.T., Li, Q.Q., Long, X.Y., Huang, X.Q., Chan, R.C., & Gong, Q.Y. (2011). Abnormal spontaneous brain activity in medication-naive adhd children: A resting state fmri study. *Neurosci Lett* 502**,** 89-93.

Yang, L., Wang, Y.F., Li, J., & Faraone, S.V. (2004). Association of norepinephrine transporter gene with methylphenidate response. *J Am Acad Child Adolesc Psychiatry* 43**,** 1154-1158.

Yasumura, A., Omori, M., Fukuda, A., Takahashi, J., Yasumura, Y., Nakagawa, E., Koike, T., Yamashita, Y., Miyajima, T., Koeda, T., Aihara, M., Tachimori, H., & Inagaki, M. (2017). Applied machine learning method to predict children with adhd using prefrontal cortex activity: A multicenter study in japan. *J Atten Disord***,** 1087054717740632.

Yoav Freund, & Schapire, R.E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55**,** 21.

Yoshimasu, K., Kiyohara, C., Minami, T., Yoshikawa, N., Kihira, S., Toyonaga, K., Yamamoto, A., Shinosaki, K., Yamashita, H., Miyashita, K., & Wakayama, A.S.G. (2009). Maternal smoking during pregnancy and offspring attention-deficit/hyperactivity disorder: A case-control study in japan. *Atten Defic Hyperact Disord* 1**,** 223-231.

Yu, S., Huang, D., Singer, W., & Nikolic, D. (2008). A small world of neuronal synchrony. *Cereb Cortex* 18**,** 2891-2901.

Zhang-James, Y., Helminen, E.C., Liu, J., Franke, B., Hoogman, M., & Faraone, S.V. (2019). Machine learning classification of attention-deficit/hyperactivity disorder using structural mri data. *bioRxiv***,** 546671.

Zhong, S., He, Y., & Gong, G. (2015). Convergence and divergence across construction methods for human brain white matter networks: An assessment based on individual differences. *Hum Brain Mapp* 36**,** 1995-2013.

Zhu, C.Z., Zang, Y.F., Cao, Q.J., Yan, C.G., He, Y., Jiang, T.Z., Sui, M.Q., & Wang, Y.F. (2008). Fisher discriminative analysis of resting-state brain function for attention-deficit/hyperactivity disorder. *Neuroimage* 40**,** 110-120.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *J R Stat Soc B* 67**,** 301-320.

Zou, L., Zheng, J., Miao, C., Mckeown, M.J., & Wang, Z.J. (2017). 3d cnn based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural mri. *IEEE Access* 5.