Characterizing the Association between Brain Morphology and Behavioural Symptomatology in Autism Spectrum Disorder and Attention Deficit Hyperactivity Disorder

by

Sina Panahandeh

A thesis submitted in conformity with the requirements
for the degree of Masters of Applied Science
Institute of Biomaterials and Biomedical Engineering
University of Toronto

Abstract

Characterizing the Association between Brain Morphology and Behavioural Symptomatology in Autism Spectrum Disorder and Attention Deficit Hyperactivity Disorder

Sina Panahandeh
Masters of Applied Science
Institute of Biomaterials and Biomedical Engineering
University of Toronto
2019

Autism spectrum disorder (ASD) and attention deficit/hyperactivity disorder (ADHD) are prevalent and highly co-occurring neurodevelopmental disorders. The brain correlates of these disorders remain mostly unknown. This is partly due to the limitations of existing analytic methods in coping with the large within disorder variability and overlap between disorders. To address this challenges, we propose a new method called Bagged-Regression clustering for data-driven discovery of diagnosis-agnostic subgroups that may share brain-behaviour associations. This approach clusters the sample data into K groups, each with its own linear regression function. Using both simulated data and a real-dataset of brain-behaviour associations in ASD and ADHD, we show that the proposed method is able to recover multiple regression lines in the data.

Acknowledgements

I would like to sincerely thank my supervisor, Dr. Azadeh Kushki, for her cordial mentoring and guidance. I am immensely grateful for her active interest and support in the project. She made my experience as a graduate student enjoyable and memorable.

I would like to wholeheartedly thank my family for their invaluable love, support and encouragement.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Chapter Overview

We will define our research question and explain the rationale behind this study in section 1.2. In section 1.3, we introduce a possible solution to the analytical gaps in autism spectrum disorder (ASD) and attention deficit/hyperactivity disorder (ADHD) studies. Section 1.4 covers the contributions of this study and section 1.5 will give a brief overview of the futures chapters.

1.2 Rationale

Autism spectrum disorder (ASD) and attention deficit/hyperactivity disorder (ADHD) are complex neurodevelopmental disorders. ASD is primarily characterized by atypicalities in social communication function as well as the presence of repetitive behaviours and restricted interests [6], and affects 1.5-1.7% of children [12, 91]. ADHD is defined by the presence of one or more features of hyperactivity, inattention and impulsiveness that interfere with daily functions [6]. The prevalence of ADHD is estimated to be more than 7% [45, 120]. ASD and ADHD are defined qualitatively based on constellations of behaviours. Currently, there are no neurobiological markers for these disorders and their neurobiological underpinnings is poorly understood.

ASD and ADHD are highly co-morbid. For example, ADHD symptoms are present

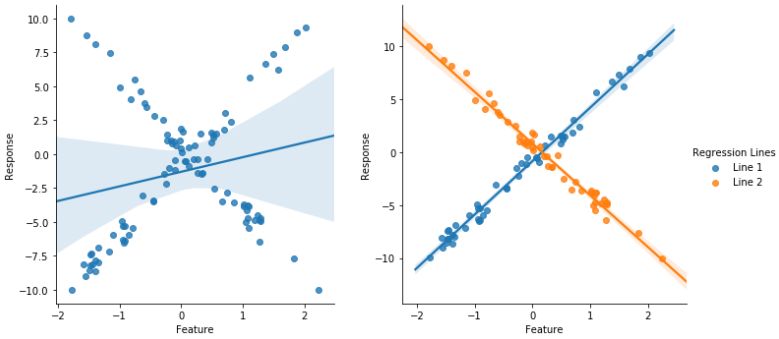
in 30-80% of ASD population; meanwhile, 20-50% of people diagnosed with ADHD show strong traits of ASD [102]. Looking at dimensional characterization of core-domain features of each disorder, there is also considerable overlap between ASD and ADHD. For example, social skills difficulties (core ASD feature) have been reported in samples of children with ADHD [115], and impulsivity and inattention (core ADHD feature) have also been reported in individuals with ASD [103]. Beyond phenotypic overlap [14, 8, 132, 123, 23, 46, 107], there is emerging evidence to support shared etiology [104, 114, 102, 74, 77, 76] and biology [5, 29, 88] in ASD and ADHD.

In addition to cross-diagnosis overlap, there is also large variability within ASD and ADHD. For example, a recent study showed that 26 mouse models of ASD could be clustered in three subgroups, each with different brain morphology characteristics, supporting the notion that ASD is not defined by a single neuroanatomical pattern [40]. Studies have also shown significant ethological heterogeneity in ADHD [48, 86, 73, 52, 109]. Neurobiological findings in ASD and ADHD have been highly variable in both the reported affected regions and the effect type [7, 40, 128, 72]. This variability poses a significant challenge to understanding of these disorders, especially when it comes to characterizing the neurobiology underlying the behavioural symptoms.

The cross-diagnosis overlap and within-diagnosis variability motivates a dimensional conceptualization of these conditions that goes beyond diagnostic labels to explain the underlying neurobiology of behaviours that cut across multiple disorders. However, very little is known about the neurobiological correlates of these difficulties. For example, it is unclear if social difficulties in ASD and ADHD share the same underlying neurobiology, or if these difficulties are associated with disorder-specific neurobiological features.

To answer the above question using traditional statistics, regression analyses can be used to characterize the association between brain and behavioural measures (e.g., cortical thickness and social function [37]). To investigate similarities and differences between diagnoses, a main effect of diagnosis or interaction effect involving diagnosis can be included in the model. For example, several studies have attempted to characterize neuropathology of social functions in ASD [66, 37] and ADHD [89] using traditional statistics. These approaches assume that diagnostic labels correspond to homogeneous groups, an assumption that is challenged by the presence of large within-diagnosis variability. In particular, linear regression analysis is not suitable for characterizing samples where multiple types of associations are present. This

challenge gives rise to a need for analytical methods that allow for different types of associations to emerge within different subgroups. Figure 1.1 is an example where traditional regression techniques relying on pre-defined diagnostic labels are not adequate to explain the data. As seen, the data are characterized by two regression lines with different slopes. In cases where these lines do not align with diagnostic labels, traditional regression analyses will fail to detect the two distinct associations.



(a) Single Regression (b) Multiple Regressions

Figure 1.1: Examples

There is already evidence to suggest that multiple types of brain-behaviour associations may exist in ASD [40, 60, 125] and ADHD [109, 9, 115], further highlighting the needs for new analytical approaches that can discover subgroups that may share similar patterns of brain-behaviour associations [4, 119, 131, 15].

1.3 Overview of the Proposed Approach

Motivated by the challenge of understanding neurobiological correlates of social difficulties across disorders, this thesis focuses on developing analytical techniques for discovering multiple types of linear associations in samples of data. To this end, I propose the use of machine learning techniques to model linear associations in data. The proposed approach builds on the concept of regression clustering (RC), an unsupervised learning method, for discovering clusters of associations. RC is similar to other centre-based clustering methods (e.g. K-Means), but instead of representing clusters by features in one domain, each cluster is represented by a regression function. In essence, the results of RC are not groups of participants that share brain or behaviour representations, but groups of participant who lie on the same regression line characterizing brain-behaviour associations. Applied to the problem of finding

brain correlates of social difficulties in ASD and ADHD, RC will result in subgroups of individuals who share similar brain-social behaviour associations.

Previously designed RC algorithms are limited when applied to data that has outlier points (eg. points that do not belong to any cluster) and have difficulties when two cluster intersect one another [10]. To address this gap, I introduce a novel RC algorithm that is based on multiple bagged robust regression models. Bagging enables us to find multiple correlates in the data and robust regressor allow the model to perform well in presence of outlier points. Our algorithm, called bagged RC (B-RC), is capable of characterizing the potentially complex relations (existence of multiple subtypes with distinct correlates) of the brain structure and social communication function. I apply this technique to multiple simulated data sets to examine its sensitivity characteristics; then, I use the same pipeline to analyze the brain-behaviour correlates in ASD and ADHD.

1.4 Contributions

The main technical contribution of this thesis is an analysis pipeline, B-RC, for discovering multiple regression lines that explain the association of brain-phenotype patterns across ASD and ADHD. Unlike traditional statistical methods, the proposed method allows for the existence of multiple subtypes with distinct brain-behaviour correlates. Moreover, the proposed method can detect clusters of simulated data in presence of outliers and when clusters intersect as we will show in future chapters.

The proposed pipeline was applied to examine associations between social communication abilities and cortical thickness in a sample of children with ASD and ADHD. Our results suggest that 1) multiple types of associations do exist in these data, and 2) the association subtypes discovered through our data-driven, diagnosis-agnostic approach do not align with existing diagnostic categories. To best of our knowledge, this is the first investigation of regression clustering in neurodevelopmental disorders.

In this thesis, I demonstrate the utility of the proposed pipeline for the specific case of social communication function/ cortical measure association. However, the pipeline is not limited to this type of data and can be applied to other data sets where discovery of multiple types of linear associations are of interest.

1.5 Thesis Overview

The rest of this document is structured as follows: Chapter 2 provides a review of relevant literature including existing studies of neuroanatomy in ASD, ADHD, as well as a review of machine learning concepts relevant to the proposed pipeline. Chapter 3 describes the proposed pipeline. Chapter 4 presents a detailed overview of the data sets used to examine the utility of the proposed pipeline. Experimental results are provided in chapter 5. In chapter 6, we will discuss the results and potential gaps of the proposed method. Chapter 7 summarizes contributions of this thesis and provides directions for future work.

Chapter 2

Background

2.1 Chapter Overview

This section provides a summary of the literature relevant to this thesis and an overview of background concepts. We begin by an outlook over social functions and its relevance in ASD and ADHD (section 2.2). Next, we will highlight previous findings regarding the neuroanatomy of ASD and ADHD with a focus on social communication processing (section 2.3). We provide a background on relevant machine learning topics (section 2.4) and discuss current challenges and gaps in the literature (section 2.5). Lastly, we explain the objectives of this study (section 2.6).

2.2 Social Difficulties in ASD and ADHD

Social perception and social communication difficulties are among the core symptoms of ASD [14, 100, 51]. According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), these include deficits in social-emotional reciprocity, nonverbal communication behaviour (e.g., gesture, eye contact), and difficulties in developing and maintaining relationships [6].

At the same time, these difficulties have been reported pervasively in other neurodevelopmental disorders, including ADHD [14, 123, 51, 84]. Children with ADHD are suggested to have poorer social and communication skills than those without [130].

Interestingly, the degree of social impairment is thought to increase with comorbidities such as oppositional defiant disorder or conduct disorder [130]. In many cases, children with ADHD may have difficulties with peer relationships, partly owning to difficulties with social exchanges such as sharing and turn taking [98, 130]. Difficulties with social perception have also been reported in ADHD [14].

Although social communication difficulties are associated with both ASD and ADHD, it remains unclear whether the neurobiology underlying these difficulties is shared across these disorder or if it is disorder specific.

2.3 Brain Correlates of Social Difficulties

Social function is modulated by a network of cortical and sub-cortical regions in the brain [119, 15, 65]. The social brain enables social cognition and ultimately social functions; hence, it is an integral part of neurological and psychiatric disorders [65, 15]. In this section, we review previous studies on brain correlates of the social behaviour and function.

Several studies have attempted to find the neurology behind social functions. Several cortical and sub-cortical regions including the amygdala, insula, temporal-parietal junction (TPJ), dorsal-medial, ventromedial prefrontal cortex (dM-PFC), the anterior and posterior cingulate cortex, the superior temporal sulcus/gyrus (STS/STG), retrosplenial Cortex, fusiform face area (FFA), and the temporal pole have all been reported to be involved in social communication function [119, 65, 27].

The anterior insula is thought to be involved with the conscious and unconscious social processes (social cognition) [18]. The amygdala is also involved in social cognition and responsible for generation emotions in social interactions. The anterior cingulate cortex (ACC) is involved with social cognition and regulation of emotion in social settings [18, 99, 41]. Posterior-STS is shown to be involved with face processing [18]. Ventral striatum are among the regions responsible for generating social related emotions. Regions of dorsal and ventral lateral PFC, Posterior medial PFC are reported to be involved with regulating emotions that arise from social interactions. TPJ, dM-PFC and precuneus have been identified to be correlated with deducing mental states and empathy [99]. For quantitative information (e.g. coordinates) on regions involved with the "social brain" you can refer to [2] where Alcal-Lpez et al.

has compiled a detail atlas of these regions (see page 6).

Previous studies show that the "social brain" consists of multiple regions as shown above. Multiple theories have been introduced to explain how a network of these regions work together to enable social cognition [2, 87, 18]. One such theory is called the "Social Information Processing Network" [87] which introduces a three-layered network of nodes. The "detection node" which is responsible for understanding the social aspects of a stimuli such as various motions involved with social interactions. This node involves regions such as the intraparietal sulcus, STS, FFA. Next, is the "affective node" which processes emotions involved in a social act and involves regions such as the amygdala, the ventral straitum, hypothalamus and orbitofrontal cortex. Lastly, the "cognitive regulatory node" which is tasked with understanding the psychological state of others in social interactions as well as generating goal oriented social behaviour.

2.3.1 ASD

In this section, we will begin by highlighting the previously found significant regions in ASD. Then, we will review previously suggested regions specifically related to social mechanisms in ASD.

Studies of brain morphology in ASD have revealed mixed and sometimes discrepant results. Volumetric structural MRI studies between typically developing children (TD) and children with ASD have reported brain differences in total brain volume, gray matter (GM) volume, and white matter (WM) volume [30, 32, 53, 108]. Several studies reported brain enlargements during infancy followed by a quiescent period during puberty in ASD groups compared to TD [39, 32, 58].

A few studies have suggested brain enlargements in adulthood [58, 44, 54]; in contrast, there have been a few papers to show reduced gray matter volume in regions such as medial temporal gyrus, fusiform gyrus, amygdala, medial frontal gyrus in adults with ASD [108]. Enlargements in pallidum and lateral ventricle volumes have been reported [122]. There is contradicting studies for atypical brain size in other brain regions such as the frontal cortex [63, 82], superior temporal sulcus [19], inferior parietal lobule [50] and cingulate [58, 22].

Cortical thickness and cortical surface area analysis of individuals with ASD has

shown an increase of gray matter compared to TD in frontal and temporal lobes; while a decrease in the gray matter has been reported in temporoparietal junction and cerebellum [38, 43]. Left lateralized cortical thickening in childhood (¿6 years) has also been reported with diminishing abnormalities in adulthood [66].

Impairments in the reported brain regions are often correlated with social mechanisms [108, 58, 66, 37]. While the neural correlates of social function in ASD are not fully understood, a network of regions including the amygdala, insula and cingulate are suggested to be involved [65]. Another study has reported atypical surface area in the right cingulate in ASD population and concluded enlargement in the insula and isthmus surface areas to be associated with poor social functions [37].

2.3.2 ADHD

In ADHD, reduced grey matter volume has been reported in ACC, basal ganglia, the dorsolateral prefrontal cortex, orbitofrontal cortex, inferior frontal cortex, medial prefrontal cortex [20, 33, 105, 59, 112, 111, 33], cortical thinning of the cortex in medial and superior prefrontal and precentral regions has also been reported in ADHD population when compared to TD [110].

2.3.3 Cross-Diagnosis Studies

Few studies have analyzed these neurodevelopmental disorders in pairs. ASD-ADHD studies show reduced grey matter volume in temporal lobe and increased grey matter volume in inferior parietal cortex [26]. Studies have also shown smaller grey matter cerebellum in ADHD and ASD populations [36, 75]. Neuroimaging investigations of ASD and ADHD groups have shown common abnormalities in left inferior frontal gyrus, frontal cortex, caudate nucleus and amygdala [58, 47, 90, 89].

2.4 Machine Learning Preliminaries

In this section, we review a few mathematical and machine learning concepts that we use in our analysis pipeline.

2.4.1 Symptom Severity Prediction from Brain Morphology in ASD and ADHD

Few recent studies have attempted to predict ASD severity scores based on structural magnetic resonance imaging (MRI) measurements by performing regression analysis [66, 109, 85, 60]. The regressor predicts a behavioural score for every participant; the score is bounded in a range of possible values dependent on the behavioural questionnaire used in the study. In the following, we will review these papers in greater depth.

In [66], the authors first performed statistical analysis to identify regions with significant differences in cortical thickness in ASD and then applied regression analysis to show the relationship between cortical thickness abnormalities and ADOS severity scores in ASD population. Likewise, in [109], authors performed regression and statistical analysis to better understand the volumetric brain measurements and behavioural ratings in ADHD population. In [85], the authors performed support vector regression (SVR) and cross-validation to obtain a symptom severity predictor. The authors were able to achieve a mean absolute error (MAE) of 1.34 in predicting Autism Diagnostic Observation Schedule (ADOS) severity scores. Hong et al. [60] introduce a different approach to prediction of ASD symptom severity; the authors used structural MRI and resting state MRI (rsMRI) to extract cortical thickness, intensity contrast on white and gray matter boundary, cortical surface area and geodesic distance from each participant. Next, they normalized the ASD population data against the distribution of TD participants and performed a hierarchical clustering to divide the participants into three distinct clusters: ASD-I: cortical thickening, increased surface area, tissue blurring; ASD-II: cortical thinning, decreased distance; ASD-III: increased distance. After dividing the participants into three subtypes, they applied regression by gradient boosting to predict ADOS based severity scores and reported MAEs of 2.080.14, 3.260.31 and 2.710.17 for each cluster respectively.

2.4.2 Regression Clustering

Clustering is an unsupervised learning method that can be generally defined as a way to find groups of samples in the data that share a common pattern. There are multiple ways of achieving this task (e.g. hierarchical clustering, K-Means) [70]. A branch of

clustering approaches are known as the centre-based clustering algorithms where each cluster is represented by a single point (e.g. K-means). As described by [133], any centre-based clustering algorithm that represents each cluster by a regression function is part of the "Regression Clustering" (RC) family of learning algorithms. These set of algorithms have also been referred to by other names such as "Clusterwise Linear Regression" [118, 34, 117, 116, 56, 55, 57, 133, 10, 11, 35].

RC is used when the nature of data is continuous and multiple coexisting regression patterns are present in the data. RC generally involves the following steps (a detailed version can be found in [133]):

Step1: Pick K regression functions on subsets of data from the entire set (in case where the subsets are overlapping, we can have a soft clustering mechanism where each point can be assigned to multiple clusters).

Step2: Iterate over all points and obtain a prediction error from every cluster-centre (regressor). Then move each point to that cluster that produced the lowest residual for that point.

Step3: Update the regression function of each cluster since the cluster memberships have been altered in the previous step.

Step4: Repeat *steps 2 and 3* until no change in cluster membership is detected.

Most centre-based clustering algorithms are known to be sensitive to their initialization steps (e.g. K-means) [133]. A popular way to combat this gap is to perform multiple random initializations and report the "best" results or compare the outcomes to observe if any common pattern arises [133, 127, 10, 78].

2.4.3 Bagging

Bootstrap aggregating or "bagging" is concept frequently used in ensemble learning algorithms. The idea is to use "bags" of sub-sampled data from the original data set in order to train multiple predictors [24]. In ensemble methods such as random forest, the outcome of multiple predictors are averaged for the final prediction [25]. We will describe in chapter 3 how we employ this idea for our analysis.

2.4.4 RANSAC Regression

The random sample consensus (RANSAC) is a popular regression method that has the ability to interpret the data that contains gross errors (outlier points) [42]. The RANSAC algorithm is an iterative model where a random minimal sample set (MSS) (eg. MSS=2 when finding the line of best fit) of data is chosen at every turn in search of the best consensus set. To find the best consensus set, the algorithm fits a model to the MSS and examines the distance (using a user defined loss function) of all point in the data set to the fitted line; using a user defined distance threshold the model then counts the number of "inliers" (number of points within the defined thresholds) and selects the model with the highest number of inliers to interpret the data and the inliers are also referred to the best consensus set [42].

In the context of this paper, we chose to use RANSAC regression because of its robustness towards outlier points. Regression method is suited because the nature of our real world data is a continuous spectrum rather than discrete measures (we will discuss the data properties in chapter 4). Robustness is a desired property since the brain-behaviour is known to be a complex problem as explained in previous sections of this chapter.

2.4.5 Spectral Clustering

Spectral clustering techniques use a similarity matrix is to partition the data into distinct groups where the data points inside a cluster are highly similar to each other. Similarity matrix describes the "similarity" of any pair of data points. The "similarity" measure can be defined differently depending on the application. We will describe our definition of a "similarity" score and how we calculate the similarity matrix in chapter 3. One of the ways to perform spectral clustering is defined in [113]; where the authors use the eigenvectors of the similarity matrix to define the data set in a new space and use k-means to cluster the points in the newly defined space. Please see [113, 126] for a more detailed explanation. We chose spectral clustering because it has been shown to perform better than other clustering methods such as k-means [126].

R-Squared

The coefficient of determination or R-Squared (R^2) is a well-known goodness of fit measure in regression problems. Past studies have introduced multiple definitions for this measure [69]. Assuming that there is a set of predictor variables $X = \{x_i | i = 1, ..., m\}$ which contains information about a dependent variable $Y = \{y_i | i = 1, ..., n\}$; or in other words, a set of data points with m features with information about a response Y on n number of samples. Let $\hat{Y}$ be the prediction of Y obtained from a regressing function. Additionally, let us define $\bar{\hat{Y}}$ and $\bar{Y}$ as the mean of $\hat{Y}$ and Y distributions respectively. In the context of this thesis, we use the following definition of R^2 [69, 71]:

$$R^2 = 1 - \Sigma(Y - \hat{Y})^2 / \Sigma(Y - \bar{Y})^2 \tag{2.1}$$

This definition of R^2 provides insight into the amount of variation in the data [(Y) around its mean $(\bar{Y})$] that the regressor model can explain. As it can be seen from 2.1, the R^2 is a dimensionless measure and its value approaches the higher bound $R^2 \leq 1$ for better predictions. On the other hand, the value of R^2 decreases to near zero (linear cases) or negative (in nonlinear cases or with presence of outliers) [69, 71].

2.5 Challenges and Gaps

Structural and functional neuroimaging techniques have enabled researchers to capture and hypothesize brain networks related to social functions. Despite the advances in imaging and our understanding of the "social brain", it remains unclear whether or not there is a shared social brain network in neurodevelopmental conditions such as ASD and ADHD that can explain the social communication deficits observed in these diagnostic categories [4]. There is increasing evidence to suggest a common pattern in symptoms and biology of these disorders. As a result, there has been a call for research in neuropathology of these conditions that span across multiple diagnostic categories [131].

Furthermore, most relational studies on brain scans and social symptom severity scores target a single diagnostic category (ASD or ADHD) and use traditional sta-

tistical methods which are limited in the type of associations they can find. These methods do not allow for presence of multiple types of brain-behaviour associations (subtypes).

As we discussed, there has been an effort by Hong et al. [60] to allow for multiple subtypes of ASD while exploring the brain regions that are correlated with symptom severity score; however, the clustering (subtyping) and symptom score prediction (regression) are performed separately of one another. Therefore, there is no feed back from the regression function to affect the subtypes. Subtypes obtained in this manner may not be a direct outcome of the brain-behaviour associations.

We believe regression clustering has the potential to address this analytical gap. By applying regression clustering to data from all diagnostic categories we hope to address these gaps. However, RC algorithms proposed in [133, 10] have been shown to be limited on contaminated data (presence of "outliers"). Furthermore, the traditional RC algorithms are sensitive to the initialization stage and the number of clusters need to be defined before analysis. To address these gaps, we will introduce a new novel regression clustering that is capable of robust clusterwise regression. We will introduce a new algorithm for estimating the number of clusters.

2.6 Objectives

In this study, we propose to use machine learning tools to characterize the links between structural brain morphology and social communication function in ASD and ADHD. We expect to find 1) a significant correlation between brain morphology and social communication deficit severity scores and 2) a subset of brain regions contributing to this relation. We also expect that there will be collections of individuals who do not follow the model, reflecting different etiologies or interactions of comorbid symptomatology with social function. Our objectives are four-folds:

1. Allowing for multiple subtypes in our analysis;
2. Coupling of clustering and regression;
3. Determining brain features that best predict social function severity scores in ASD and ADHD;

4. Determine if the potential subtypes are separable by traditional diagnostic categories.

Chapter 3

Research Methods

3.1 Chapter Overview

Figure 3.1 illustrates an overview of our analysis pipeline. We will explain each step in this chapter. We will propose a regression clustering algorithm based on random sample consensus (RANSAC) and spectral clustering to characterize linear brain-behaviour correlations (section 3.2). The proposed pipeline is implemented in Python (version 3.6) and takes advantage of open-source software packages provided by the SciPy community (Scientific Computing in Python) [96, 67, 62, 92, 83, 64, 95] and others [129, 31]. Lastly, we describe our validation measures to examine cluster stability and analyze significant features (section 3.3).

3.2 Bagged Regression Clustering

We propose a regression clustering algorithm based on bagged regression and spectral clustering, which we call bagged regression clustering (B-RC) to follow the naming convection used in previous literature [133]. The input to the algorithm two vectors of response (eg. social communication score) and feature (eg. measurements of a brain region). The first step is to build a similarity matrix that characterizes the similarity between two data points (eg. participants) with respect to a regression line (3.2.1). The next steps are to estimate the number of clusters in the data and group (or cluster) the data points based on their spectrum of similarity scores (3.2.2). The

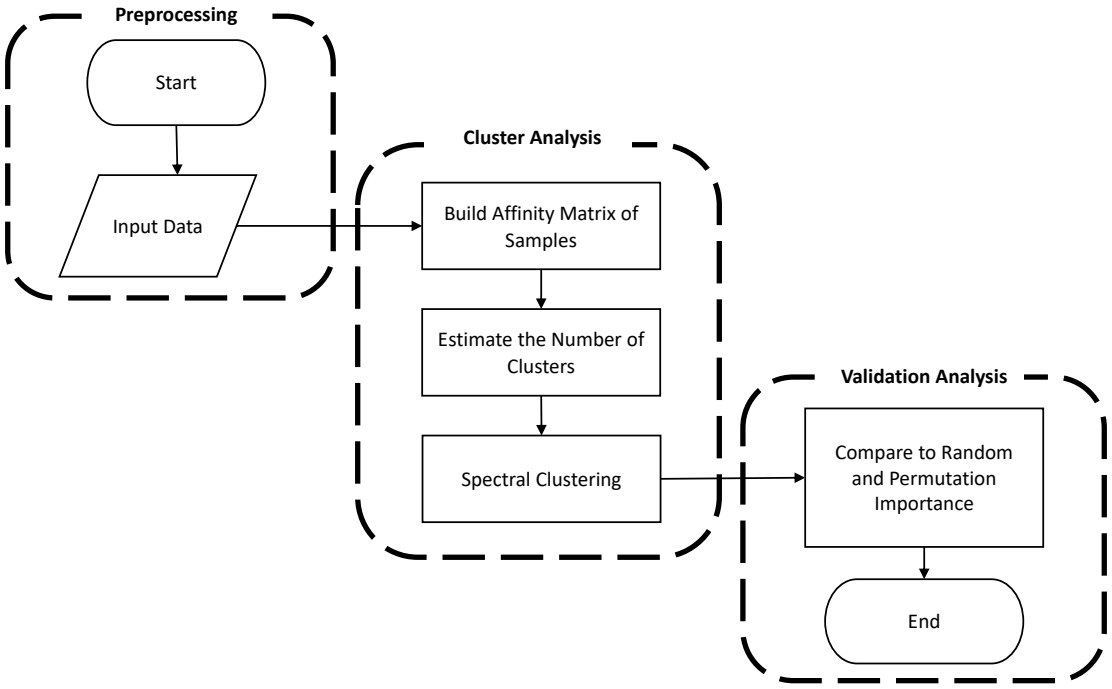


Figure 3.1: Analysis Pipeline Overview

output of the algorithm is cluster assignments for each data point (or participant in the case of our real world data set). A general overview of the processing procedure is visualized in Figure 3.1. We introduce a novel way of building a similarity matrix based on regression outcomes from random subset of samples (bags) that allows us to map the similarities of participants to each other which enables finding multiple regression lines in the data. Additionally, we will discuss a scatter measure to estimate the number of clusters based on within and between cluster similarity scores.

3.2.1 Building the Affinity Matrix

The goal is to build a similarity matrix $\boldsymbol{S} := (s_{i,j})$, where $s_{i,j}$ denotes the similarity between two data points i and j ($i, j \in \{1, 2, .., N\}$; N : number of points in the data set). In case of regression clustering, the similarity between two points is expected to be high when both samples belong to the same hyper-plane in any given space. In this report, we will limit our analysis to the two-dimensional space defined by a feature (e.g., cortical thickness) and the response (e.g., measure of social function). Therefore, if there exists a linear correlation between the feature and the response, samples

that are in close proximity of the line are thought as "similar" to each other. To compute the affinity matrix, several steps are taken as described in Figure 3.2. These steps include bagging, fitting a regression line to samples in each bag, finding the distance of each point to the line, transforming distances to similarity values, computing affinity values, and updating the affinity matrix. These steps are described in detail in what follows.

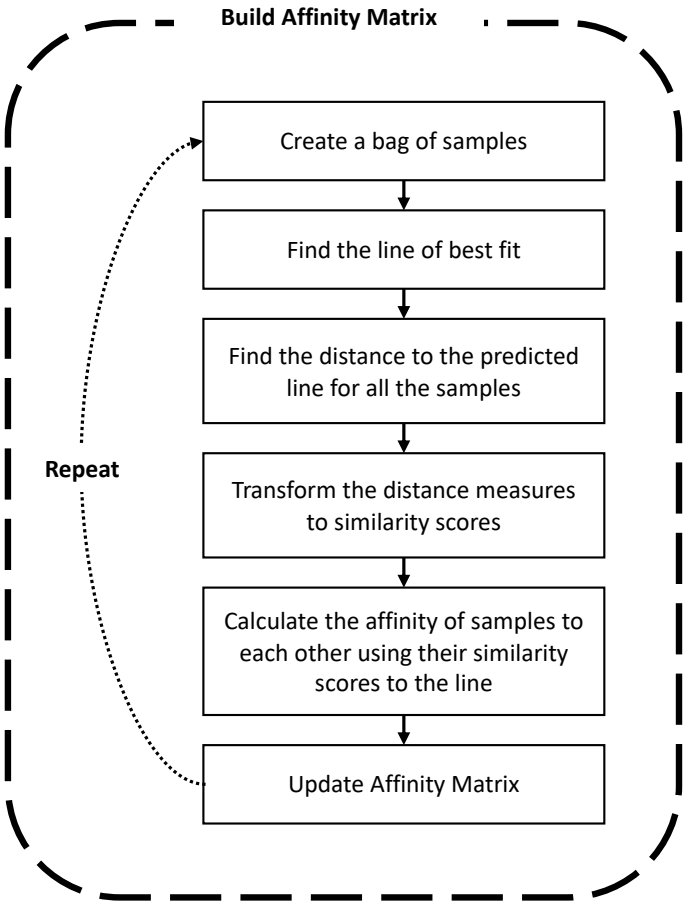


Figure 3.2: Overview of building the affinity matrix

Bagged Regression

Bagging for multiple iterations allows us to look at small portions (individual bags) of the data set and examine if the relation between those samples (if any) can generalize to the entire data set. If we perform regression on the entire data set we could be missing correlations if the data set is made up of multiple correlations as we explained

in previous chapters. In our analysis pipeline, bagging prior to regression allows us to examine the existence of multiple correlations in the data set. We see if there is any linear correlation in each bag and if there is, we then look at the entire data set to examine if the linear relation also generalized to other samples in the data set. Bagging has been applied to ensemble regression models such as random forest [25] however, this is the first time that bagging is applied to regression clustering to the best of our knowledge.

Bagging will create random bags of samples without replacement (K iterations in total). The bag size is set by the user as a percentage of the entire sample size ($x\%$ of the entire set).

Finding Line of Best Fit

At each iteration we perform a RANSAC regression on the given bag of samples. It is not guaranteed that the RANSAC algorithm will find a consensus set in every bag; therefore, we do not expect the regressor to converge at every iteration. In other words for K iterations the total number of regression lines will be $m \leq K$. We expect to have participants in our real world data that do not belong to any cluster; such data points are often referred to as "outlier" points in context of regression. We chose the RANSAC regressor of its robustness to outliers [42].

Distances to the Line

At each iteration, once a regression line is found, we calculate the prediction residuals (distance measures) from the regressor for all samples - these distances are used to build a similarity measure. The distance, for the i^{th} sample at iteration k, is defined as:

$$r_{i,k} = y_i - \hat{y}_{i,k}, \tag{3.1}$$

where y_i is the actual value of response and $\hat{y}_{i,k}$ is the predicted response value for the i^{th} sample in the k^{th} iteration.

Similarity Scores

For each bag, the residual ($r_{i,k}$) between the predicted value using the regression line is computed using Equation 3.1. The distance measure gives information about the the generalizability of the line in the data set since points that are far from the would have large residuals. The the generalizability of the line in the data set increases as the number of points with negligible residuals increases. We transform the distance into a similarity measure using a Gaussian kernel as defined in Equation 3.2.

$$a_{i,k} = \alpha e^{\beta(\frac{r_{i,k}-\bar{r_k}}{\sigma_{r_k}})^2}, \tag{3.2}$$

where, $r_{i,k}$ is the residual for the i^{th} sample at iteration k, and $\bar{r_k}$ and σ_{r_k} are the the mean and standard deviation of the "inliers" within the bag as determined by the RANSAC regressor. This normalization allows us to determine how the residual r_i compares to the residuals from the data points that define the regressor (inliers).

There are many ways to convert a distance measure to a similarity score; we use a Gaussian kernel to be able to control the smoothness of the transformation (at which point we want to reach a low similarity score). The two variables of α and β in Equation 3.2 are design parameters used to control the height and width of the Gaussian kernel and can be selected based on the application based on a desired similarity score for when a residual reaches a threshold. In our case, these parameters were chosen so that the function yields a similarity score of 1 for residuals that are 2 standard deviation away from the mean ($\bar{r_k}$) and points that are farther would have a similarity score of < 1. This design constraint resulted in values of $\alpha = 10$ and $\beta = 0.6$.

Given $a_{i,k}, a_{j,k}$, we use Equation 3.3 to describe the similarity of the points to each other ($s_{i,j,k}$) as follows:

$$s_{i,j,k} = a_{i,k}.a_{j,k} \tag{3.3}$$

Using Equation 3.3, similarity of two points would be highest only when both points have a high similarity score to the line (small distance, small prediction residual). In turn, the similarity score of two points would be lower if only one point has a high

similarity score to the line and it would be lowest when both point are far from the line.

Calculating the Affinity Matrix

From Equation 3.3, we can build the an affinity matrix $\boldsymbol{S_k} := (s_{i,j,k})$. Next, we average all affinity matrices over K iterations to build the final affinity matrix as described in Equation 3.4. The obtained affinity matrix is then used for clustering as detailed in the next section.

$$\boldsymbol{S} = \frac{\sum_{k=1}^{K} \boldsymbol{S_k}}{K} \tag{3.4}$$

3.2.2 Clustering

Several methods exist for clustering an affinity matrix and among these we picked spectral clustering. After obtaining the affinity matrix which describes the similarity of samples to each other, we use spectral clustering to obtain clusters of samples. Clustering allows us to create groups of samples that are similar to each other as defined by Equation 3.3. If these are any linear relationships between the input and the response for a group of samples, we expect that it would be shown in the obtained clusters.

Spectral clustering algorithm requires the number of clusters to be defined before performing clustering. To estimate the number of clusters, we first assume that there are n clusters in a data set. For each cluster we define a within-to-between scatter ratio:

$$\mu_{C_\theta} = \frac{Median(Within_{C_\theta})}{Median(Between_{C_\theta})}. \tag{3.5}$$

Equation (3.5) is the ration of the similarity of data points within a given cluster (within cluster similarity) to similarity of the data points within that cluster to all data points outside the given cluster (between cluster similarity). In Equation 3.5, we use the median similarity scores between the data points assigned to the same cluster

to estimate the "within cluster similarity" ($Median(Within_{C_\theta})$); similarly, we use the median similarity score among the points within the cluster and points outside of that cluster to estimate the "between cluster similarity" ($Median(Between_{C_\theta})$). We chose to represent the within and between cluster similarities with the median measurement since we did not want to make assumptions about the distribution of the score and the median is considered a robust measurement to represent the data even under skewed conditions. The ratio in Equation 3.5 increases as the data points in a cluster become more similar to each other and/or dissimilar (low similarity score) to data points outside their cluster.

Since it is desired to obtain differentiated clusters, we seek to find the best number of clusters (n) to maximize Equation 3.5 for all clusters. To account for clusters with different number of points, we use the following:

$$\phi_n = \sum_{\theta=1}^{n} \mu_{C_\theta} . \frac{N_{C_\theta}}{N_\theta}, \tag{3.6}$$

where N_{C_θ} is the number of samples in cluster θ and N_θ is the total number of samples in the data set.

Equation 3.6 uses the weighted average of all individual cluster scores obtained from Equation 3.5 to build a score for when the clustering is performed at a defined number of clusters. The weight of each score (μ_{C_θ}) in Equation 3.6 is based on the number of samples within a cluster since we value larger clusters more than clusters with fewer samples. Large sample size in a cluster is an indication of its generalizability in the entire data set.

In summary, the number of clusters is obtained as follows: cluster the samples multiple times where each time the clustering is performed at a different number of clusters (e.g., 2 to 10 clusters); use the labels from each clustering and the affinity matrix to calculate a ratio of within to between cluster affinity for each cluster as defined by Equation 3.5; then, calculate the weighted average scores of these ratios as defined in Equation 3.6 for each time the pipeline is initialized with a different number of clusters; Last, compare the scatter ratios for different number of clusters (ϕ_2 to ϕ_{10}) and use the first local maximum score as an estimation for the total number of clusters in a given data set. We choose the first local maximum as opposed to the maximum score since the ratio in Equation 3.6 can artificially increase for high number

of clusters. By increasing the number of clusters we expect to see smaller clusters with higher within to between similarity scores(best case is when every point is its own cluster since the similarity of a point to itself is the highest). The process is summarized in Figure 3.3.

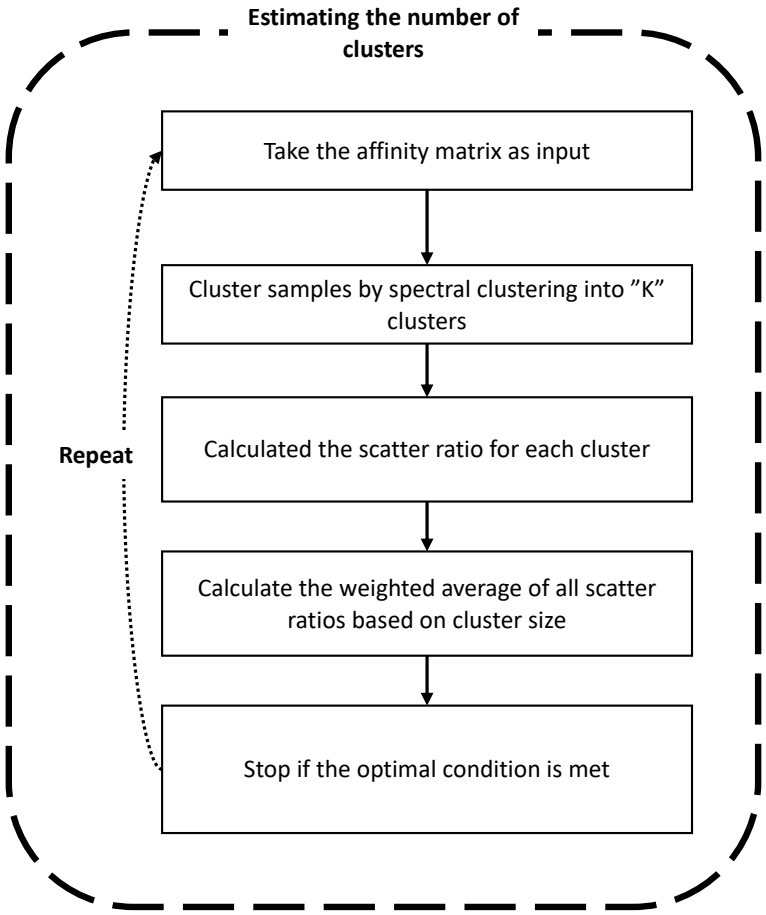


Figure 3.3: Overview of the process for estimating the number of clusters.

3.3 Performance Evaluation

3.3.1 Known Labels

We use the adjusted rand index (ARI) [61] to measure the performance of a clustering algorithm if the data has known labels (clusters are known prior to clustering). An ARI score of +1 means the expected labels match the predicted labels from the

pipeline perfectly and a score of 0 suggests that the agreement between points as defined by predicted labels are not different from the agreement between points as defined by randomly picked labels [61].

3.3.2 Unknown Labels

The previous sections described the B-RC pipeline for discovering multiple types of association in samples of data. In the remaining of this section, we discuss method that can be used to determine the significance of the obtained clusters when the true labels are unknown.

Assume that the data points for a given $Y = f(X)$ relationship are clustered into n clusters; where Y is a response and X is an input (feature). The first goal is to understand how different this clustering is compared to random grouping of the data points. To evaluate if the discovered clustering is significantly different than chance, we propose a permutation test to compare the scatter ratio obtained from the data to that obtained by chance. To this end, we proposed two methods based on how "chance" is defined (Sections 3.3.2 and 3.3.3). Second, if the clustering is significantly different ($p < 0.05$) from chance, we would like to know if clusters are characterized by a significant linear correlation between X and Y (Section 3.3.4).

Permutation Test

To this end, we build a distribution of random ϕ_n scores at different levels of n and compare the ϕ score from the pipeline to the distribution to calculate a p-value for a give clustering on X.

To create the random distribution, we randomize the relation between the feature vector (X) and the response vector (y) by permuting the feature vector to create another feature vector X'. Next, we perform clustering on (X', Y) via the proposed pipeline: use the permuted feature to compute $\boldsymbol{S'_i}$ (Equation 3.4). We repeat this procedure t times with different permutations of the original feature ($1 < i \leq t$) to obtain the set $\boldsymbol{S'} = \{\boldsymbol{S'_1}, ..., \boldsymbol{S'_t}\}$. Finally, for a given number of clusters (n) we will perform spectral clustering on all $\boldsymbol{S'_i} \in \boldsymbol{S'}$ and use the $\phi'_{n,i}$ values to build a distribution of random scatter scores (ϕ'_n). This method has the advantage that it

does not make any assumptions about the distribution of the response and feature vectors. We use the following equation to obtain a significance p-value for ϕ_n:

$$p = \sum_{i=1}^{t} [\phi'_{n,i} > \phi_n] \tag{3.7}$$

All generated p-values from this step are corrected for false discovery (FDR) using the Benjamini-Hochberg method [16].

3.3.3 Random Inputs

In this section, we describe how a randomly generated data set can be used in the same manner as section 3.3.2 to compare the performance of a clustering against clustering to random data with a pre-defined distribution. Given a random data set (X', Y') we perform the same steps as in section 3.3.2 to obtain a set of affinity matrices $(\boldsymbol{S'})$ based on random data and build a random scatter score distribution (ϕ'_n). Then, we can use Equation 3.7 to calculate a p-value that describes how significant is the difference between a given clustering score (ϕ_n) compared to a clustering with the same number of clusters (n) on randomly generated data. In this study we used uniform distributions to build our random data set. We only report the results from this test to compliment to results from section 3.3.2 since we do not know the true distribution of our real world data. As before, the generated p-values from this step are FDR-corrected using the Benjamini-Hochberg method in [16].

3.3.4 Testing Significance of Associations

This section explains the procedure of understanding which clusters contain significant linear correlations after detection of a significant clustering ($p < 0.05$ form Section 3.3.2). To do this, we perform a two-step process to determine 1) if the variance in identified cluster can be explained by a linear model as measured by R^2, and 2) the type and strength of the linear correlation between the feature and the response.

Adjusted R^2

We use data within each cluster to perform an ordinary least squares (OLS) regression and obtain the "cluster's adjusted R^2". This R^2 measure is calculated using the predicted response values (Y'_{C_θ}) and the true response values (Y_{C_θ}) for samples in cluster θ. This allows us to understand how much of the variance in the cluster can be explained by a linear regressor. [71].

Type of correlation

If the obtained $Adj.R^2$ is above the threshold of 0.3, we will use the coefficient value in the OLS model as an estimate of the correlation strength between the input the response as explained by the samples of that specific cluster. We consider a linear correlation to be strong when the correlation coefficient is $R > 0.5$ (since $0 \leq R \leq 1$ where the value of 1 represents a prefect correlation and the value of 0 represents no correlation [71]). For linear models, coefficient of determination (R^2) is equivalent to square of the coefficient of correlation; hence the R^2 threshold of $\lceil (0.5)^2 \rceil = 0.3$.

Chapter 4

Data Sets

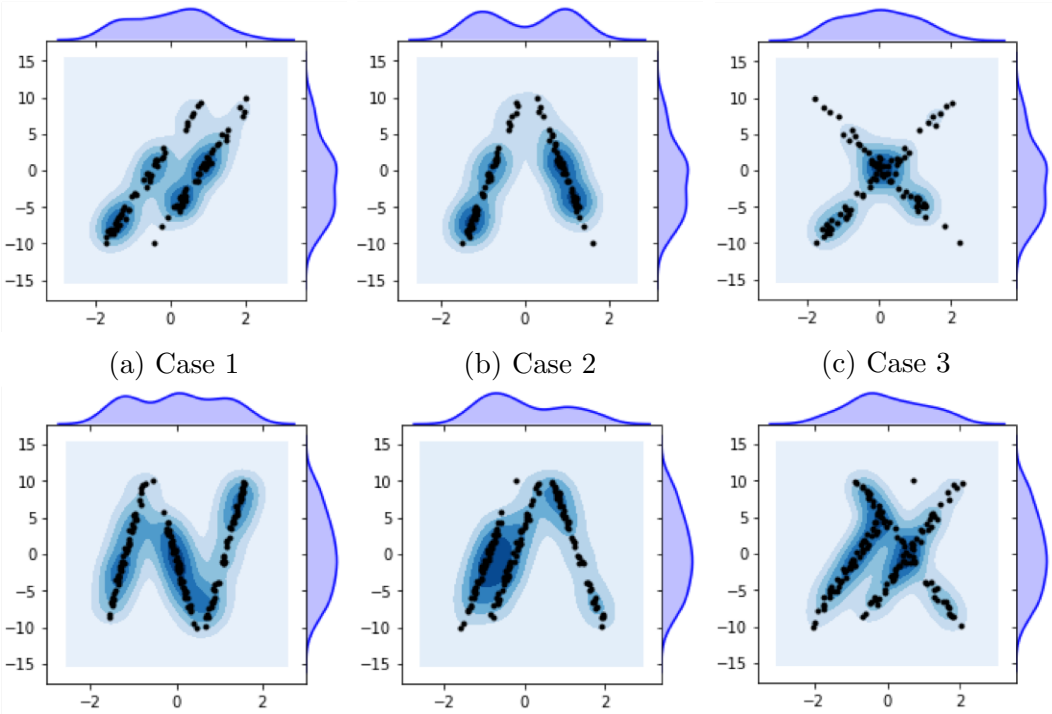
4.1 Chapter Overview

This chapter describes the data sets used to demonstrate the function of the B-RC algorithm. First, we use a simulated data set to validate the algorithm and examine the effect of algorithm parameters (section 4.2). Second, we use a data set from the Province of Ontario Neurodevelopmental Disorders (POND) Network to illustrate how the algorithm can help in discovering brain-behaviour associations (section 4.3).

4.2 Simulated Data

We created a set of simulated data to validate the proposed algorithm. The purpose of this set is to quantify the behaviour of the B-RC algorithm on a set with known cluster labels, and to explore the effect of algorithm parameters on clustering performance.

Our simulated data set contained 6 clustering scenarios shown in Figure 4.1. These cases were designed to contain similar types as associations reported in previous neurodevelopmental studies; we designed the clusters to have positive and negative linear relationships to the response which are inspired by reports of increase or decrease in certain regions of the brain as we described in chapter 2. We considered cases with 2 or 3 clusters based on previous literature reports that have examined clustering in ASD disorders [60, 40].



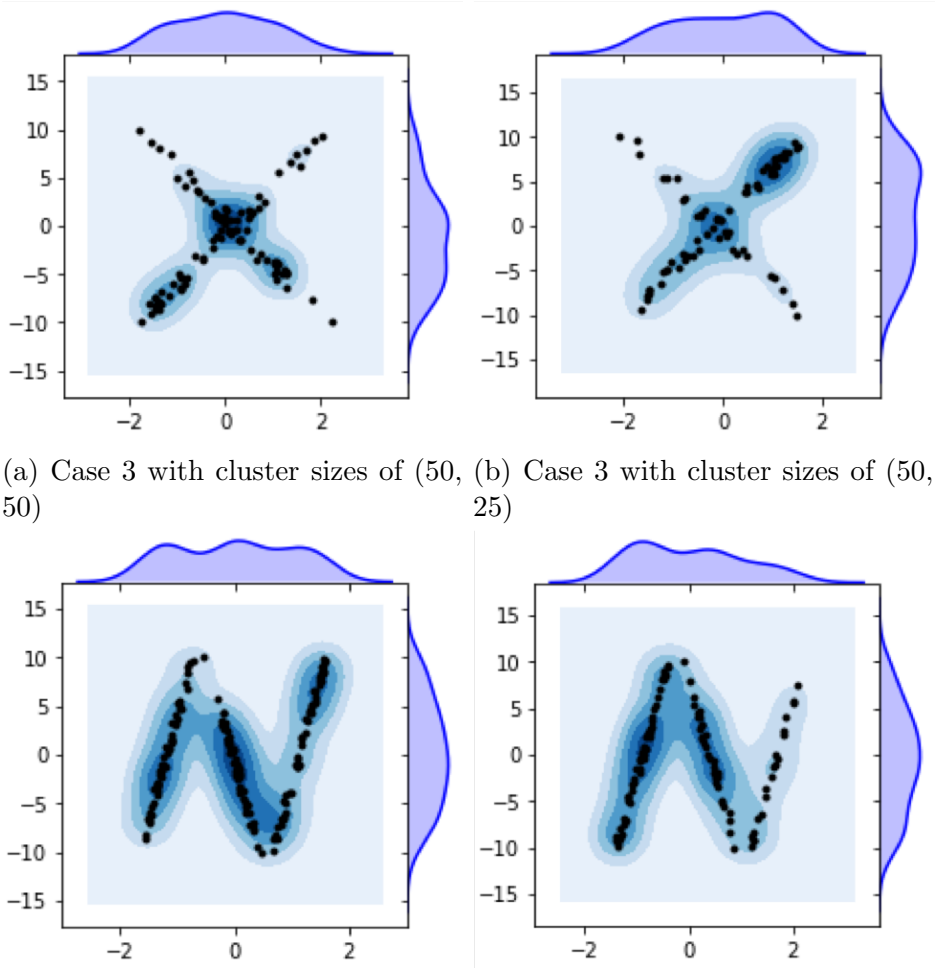
(a) Case 1 (b) Case 2 (c) Case 3

(d) Case 4 (e) Case 5 (f) Case 6

Figure 4.1: Six distinct patterns were used to generate the simulated data sets.

To examine the effect of cluster size, we examined two variations of each of the six cases above where the portion of samples contained in each cluster were varied (number of samples in two-cluster cases: (50, 50), (50, 25); number of samples in three-cluster cases: (50, 50, 50), (50, 35, 20)). Examples of these sets are shown in Figure 4.2. We chose these sampling sizes so that the total number of samples be close to the actual size of our real world data set (171 samples); we picked the smallest cluster sizes be close to 20 samples since of the smallest found in a previous cluster analysis in ASD was 19 samples [60]. We did not consider other constraints and the rest were picked arbitrary.

We also varied the distribution of points within each cluster. We randomly sampled one cluster from the Gaussian distribution and one cluster from the uniform distribution for cases with 2 clusters; cases with 3 clusters have one cluster that is sampled from the Gaussian distribution and two clusters that are sampled from a uniform distribution. For example, if a cluster is randomly sampled from the Gaussian distribution, it means that the response for the points in that cluster are randomly picked from a Gaussian distribution. Since we do not know the actual population distribution of our real world data, we used the two commonly known distributions



(a) Case 3 with cluster sizes of (50, 50)

(b) Case 3 with cluster sizes of (50, 25)

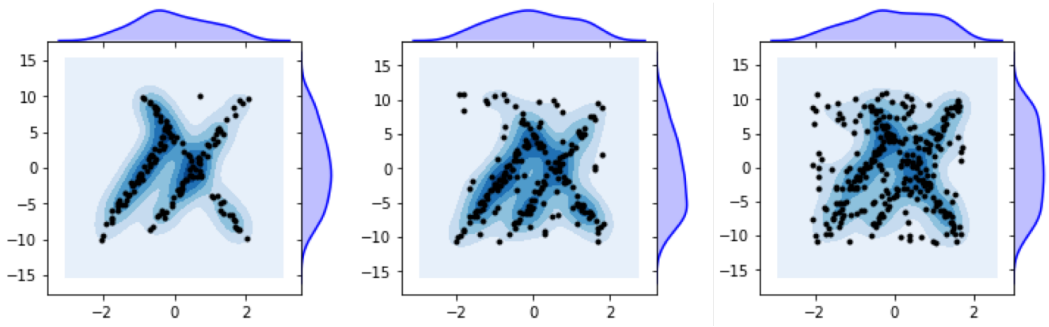
(c) Case 4 with cluster sizes of (50, 50, 50)

(d) Case 4 with cluster sizes of (50, 35, 20)

Figure 4.2: Proportion of number of samples in each cluster was varied to generated additional simulated sets.

of Gaussian and uniform. The simulated features were normalized to have zero mean and unit variance since the real world data are also preprocessed in the same manner. We will describe the real world preprocessing in a future section.

Finally, to examine the effect of noise on algorithm performance, we considered the above cases under three noise ratios: namely, 0%, 25% and 50%. This was simulated by adding uniformly distributed random noise with the specified power to the base cases. Examples of sets with different noise ratios are shown in Figure 4.3. For example, 25% noise in a data set means that 25% of the entire samples size is consisted of random points. In this report, we sometimes refer to noise as "outlier rate".



(a) Case 6 at 0% noise (b) Case 6 at 25% noise (c) Case 6 at 50% noise

Figure 4.3: Test cases with three different outlier rates were considered.

We also defined a vector of true labels for each set such that points belonging to the same cluster share the same label. Additionally, noise is defined as a separate cluster and all added random points shared a similar label in data sets with noise rate of greater than 0%. The above variations produced a total of 36 simulated test sets. Table 4.1 provides details of each case.

	2 Clusters	3 Clusters
Number of samples per cluster	(50, 50), (50, 25)	(50, 50, 50), (50, 35, 20)
Cluster distribution	(Uniform, Gaussian)	(Uniform, Gaussian, Uniform)
Noise	0%, 25%, 50%	0%, 25%, 50%

Table 4.1: Simulated Data Characteristics

4.3 POND Data

The second data used to illustrate the performance of the B-RC algorithm included brain-behaviour data from a subset of participants who took part in POND Network studies. POND is a research collaboration among multiple centres in Ontario, Canada, aimed at characterizing a sample of children with neurodevelopmental disorders on several measures ranging from genetics to neuroimaging to various phenotypic measures.

4.3.1 Participants

Data from a sample of 171 participants from POND were used for this study. The included participants were 5-20 years old, had sufficient English comprehension to complete the testing protocols, and did not have contraindications for MRI. Participants were recruited through and tested at Holland Bloorview Kids Rehabilitation Hospital and the Hospital for Sick Children. A primary diagnosis of ASD or ADHD was required for inclusion in our data set. Diagnoses for the ASD groups was supported by the Autism Diagnostic Observation Schedule-2 (ADOS) and the Autism Diagnostic InterviewRevised (ADI-R). ADHD diagnosis was supported by the Kiddie-Schedule for Affective Disorders and Schizophrenia (K-SADS) and the Parent Interview for Child Symptoms (PICS). Participants with missing imaging or behavioural data were excluded from the analyses. Participants needed to have answered all items in the Social Communication Questionnaire (SCQ) to be included in the study.

The research ethics boards at Holland Bloorview Kids Rehabilitation Hospital and the Hospital for Sick Children approved the study. Participants who had capacity to consent provided informed consent. For others, consent was obtained from guardians and assent was obtained from the participants. Participants characteristics is reported in Table 4.2.

	ASD(n=121)	ADHD(n=50)	Total(n=171)
Age	11.90 ± 3.58	10.83 ± 2.34	11.59 ± 3.30
Sex (M:F)	98 : 23	42 : 8	140 : 31
Full-scale IQ	91.16 ± 24.52 (n=115)	98.78 ± 16.65 (n=37)	93.01 ± 23.04 (n=152)
SCQ Total	19.86 ± 7.71	8.20 ± 5.58	16.45 ± 8.90
SCQ Soc/Com	13.21 ± 6.07	5.80 ± 4.60	11.05 ± 6.60
SCQ RRB	5.13 ± 2.13	1.74 ± 1.86	4.14 ± 2.57

Table 4.2: Characteristics of Participants

4.3.2 Measures

The utility of the proposed algorithm was demonstrated by examining brain-behaviour associations across the two diagnostic groups of ASD and ADHD. In particular, we

considered the association between symptom severity in the social communication domain and cortical measures of volume, surface area, and thickness.

Social Communication Measure

We used the Social Communication Questionnaire (SCQ) - Life Time, to quantify symptom severity in the social communication domain. The SCQ is a dimensional measure of ASD symptomatology for individuals over 4 years of age. The SCQ is a parent/caregiver questionnaire and consists of 40 yes/no questions probing ASD-like symptomatology. This questions can be grouped into three categories of reciprocal social Interaction, communication, and restricted, repetitive, and stereotyped behaviours [106]. For this study, We used the sum of the scores across the social (15 items) and communication (13 items) domains (Figure 4.4). The SCQ has been shown to have good psychometric properties (internal consistency: .84-0.93, strong correlation with the autism diagnostic interview total score, high discriminative validity [106]).

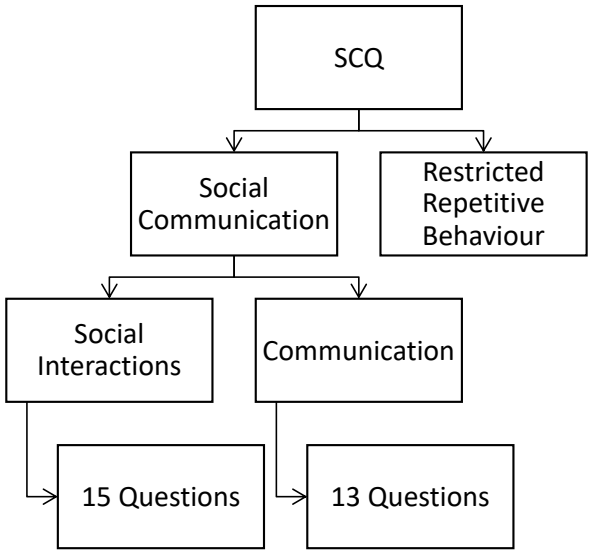


Figure 4.4: The SCQ score break down for the proposed analyses.

Imaging Data

Structural MRI data was acquired for all participants at the Hospital for Sick Children (Toronto, Canada). The majority of scans (n=134) were captured using a 3-Tesla Siemens Trio TIM scanner and the remaining scans (n=37) were obtained from a Siemens Prisma scanner after a scanner upgrade in June of 2016.

All images were processed with the fully automated CIVET pipeline (version 2.1.0) [134, 1]. The Montreal Neurologic Institute (MNI ICBM152 - version 2009) template was used as target surface registration. Brain tissues were classified to white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) based on T1-weighted images. A surface diffusion kernel was applied and the images were mapped to the automated anatomical labelling atlas (AAL). Cortical thickness was measured as the distance between WM and GM surfaces; the area and volume are vertex-based measurements from the local variations in area/volume contraction and expansion [17]. Scans were quality controlled (QC) for motion artifacts and vetted by the CIVET's QC pipeline. Flagged scans were manually reviewed by an expert and excluded if needed.

This procedure results in 76 cortical regions. Corticometric, morphometric and volumetric features were calculated for every region (total=228) and used as inputs to the analysis pipeline.

We statistically corrected features for age, sex, scanner effects, and total gray matter volume prior to analysis [60]. We further mean centered and standardized each brain region. The overall process used to obtain the brain data is outlined in Figure 4.5.

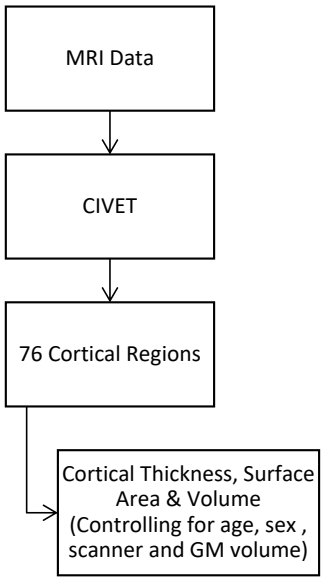


Figure 4.5: Brain measures for the proposed analyses.

Chapter 5

Results

5.1 Chapter Overview

This chapter details the results of the experiments performed to evaluate the proposed B-RC algorithm. Section 5.2 provides results of experiments on simulated data, and section 5.4 illustrates how the proposed method can be applied to discover brain-behaviour associations using data from POND.

5.2 Simulated Data

This section details the results of the proposed regression clustering pipeline on simulated data outlined in the previous chapter. To this end, we evaluated the performance of the algorithm on the six cases discussed previously, and examined the algorithm's sensitivity to choices of parameters and noise conditions. Lastly, examples illustrating cases where the algorithm performed well and cases that it failed to identify the correct clusters are presented. We study each case (base cases illustrated in Figure 4.1) separately where the predictor is uni-dimensional and it is possible to visualize the results in a 2-dimensional space.

Given that the number of iterations the algorithm can be run is limited by computation time, we set the number of iterations (K) to 100,000 and we repeated each experiment 10 times to characterize performance variability. Each run provided us

with an affinity matrix and scatter score (ϕ_n) estimation for different number of initialized clusters. The number of clusters are estimated using the average of these scatter scores. The obtained affinity matrices from theses runs were averaged to perform spectral clustering using the estimated number of clusters. The predicted labels obtained from spectral clustering were compared with true labels to calculate an ARI score for each experiment.

5.2.1 Sensitivity to Bag Size

In this section, we characterize the effect of bag size on performance of the B-RC algorithm. The bag size refers to the number of samples chosen in each bag at each iteration of the algorithm. Bag sizes of 5%, 33% and 63.2% were chosen for these experiments. We picked this range of bag size to examine the algorithm on relatively small, medium, and large bagging rates. We based the smallest bag size on the smallest simulated data set with 75 samples and the constraint of the minimum sample set (MSS) from the RANSAC algorithm. The latter (63.2%) is commonly used in bagging as this bag size increases the odds of selecting at least one sample from each of the clusters [24]. This guarantee, however, is not significant for the proposed algorithm since having multiple points from the same cluster does not ensure that the RANSAC regressor would find that cluster.

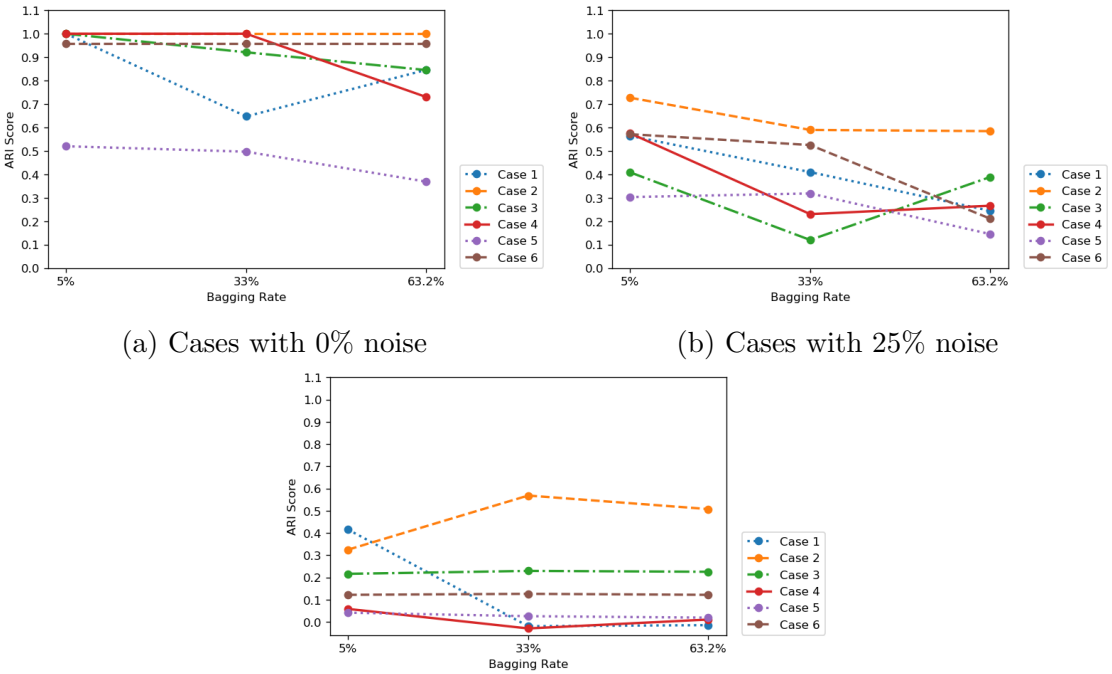
Figure 5.1 shows the effect of bag size on the ARI scores at different noise levels. As seen, the algorithm can be sensitive to the choice of bag size and the bagging rate of 5% out performed higher rates in majority of the comparisons.

We observe that at 0% noise level and 5% bag size (Figure 5.1a), almost all cases (except for case 5) show a high ARI score (> 0.9) and the number of cases with lower ARI scores increases as the bag size increases. These cases include cases 1 and 5 for 33% bag size and cases 1, 3, 4 and 5 for 63.2%.

This observation also holds true when looking at 25% noise level (Figure 5.1b); the number of cases with ARI scores of higher than 0.5 drops as the bag size increases. These include cases 3 and 5 at 5%, cases 1, 3, 4, 5 at 33% and almost all cases (except for case 2) at 63.2%. The proposed pipeline obtains low ARI scores on all data sets at the noise level of 50% as shown in Figure 5.1c.

Of the several cases tested, in one case (Case 2 in Figure 5.1c) the bag sizes of 33%

and 63.2% resulted in higher ARI scores compared to a 5% rate. We further examined the clustering visualizations of these cases (Figures 5.8c and 5.8d) for 33% and 63.2% bagging rate with Figure 5.2). The comparison shows that in this instance bag sizes of 33% and 63.2% outperform the 5%.



(a) Cases with 0% noise

(b) Cases with 25% noise

(c) Cases with 50% noise

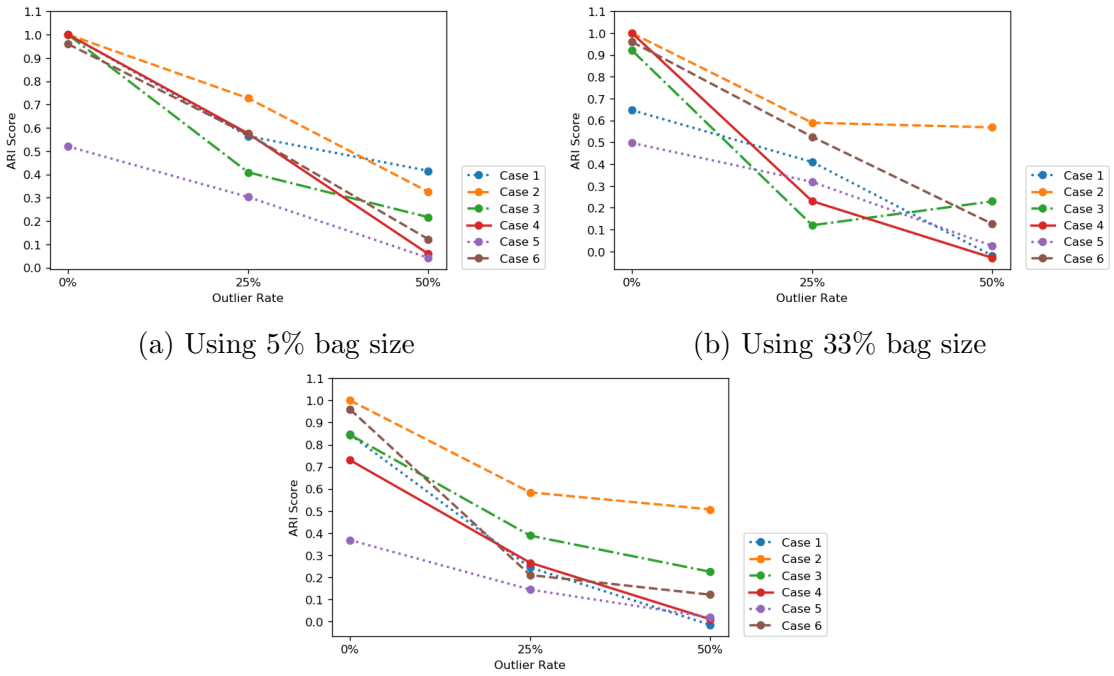
Figure 5.1: Effect of bag size on algorithm performance.

Figure 5.2: Case 2, 50% noise, 5% bag size, balanced

5.2.2 Algorithm Sensitivity To Different Noise Levels

This section characterizes the effect of noise on the proposed algorithm's performance. As we discussed before, the noise level determines the amount of random data in a simulated set compared to the true data. Figure 5.3 shows the results of the algorithm at noise levels of 0%, 25% and 50%.

As expected, the performance of the algorithm degrades with increasing noise, with the lowest ARIs at noise level of 50% in most of the tested cases. Figure 5.3a shows that the ARI score drops for all cases as the noise (random points) in the data set become more prominent. The same observation can be made about Figures 5.3b and 5.3c at 33% and 63.2% bag sizes.



(a) Using 5% bag size

(b) Using 33% bag size

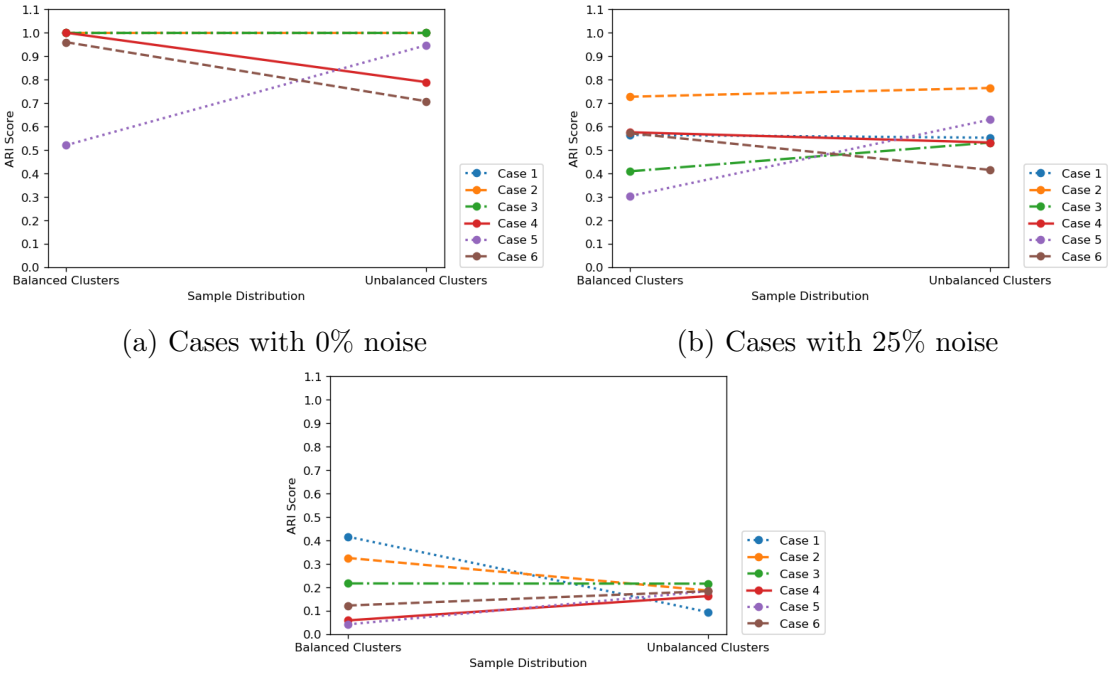
(c) Using 63.2% bag size

Figure 5.3: Effect of noise on algorithm performance.

5.2.3 Detecting Variable Size Clusters

In this section, we examine the performance of the algorithm when the number of points in different clusters is different. To this end, the ARI values for the unbalanced data set described in chapter 4 are reported. The bag size is set to 5% for the experiments in this section.

Figure 5.4 shows the ARI results for balanced and unbalanced clusters for noise levels of 0%, 25% and 50%. As seen, the effect of cluster balance seems to vary across cases. For example, for the no noise condition (Figure 5.4a), unbalanced clusters resulted in decreased ARI for two cases (4,6), increased ARI for one case (5), and no change for three cases (1,2,3). The effect also seem to vary depending on the noise level.



(a) Cases with 0% noise

(b) Cases with 25% noise

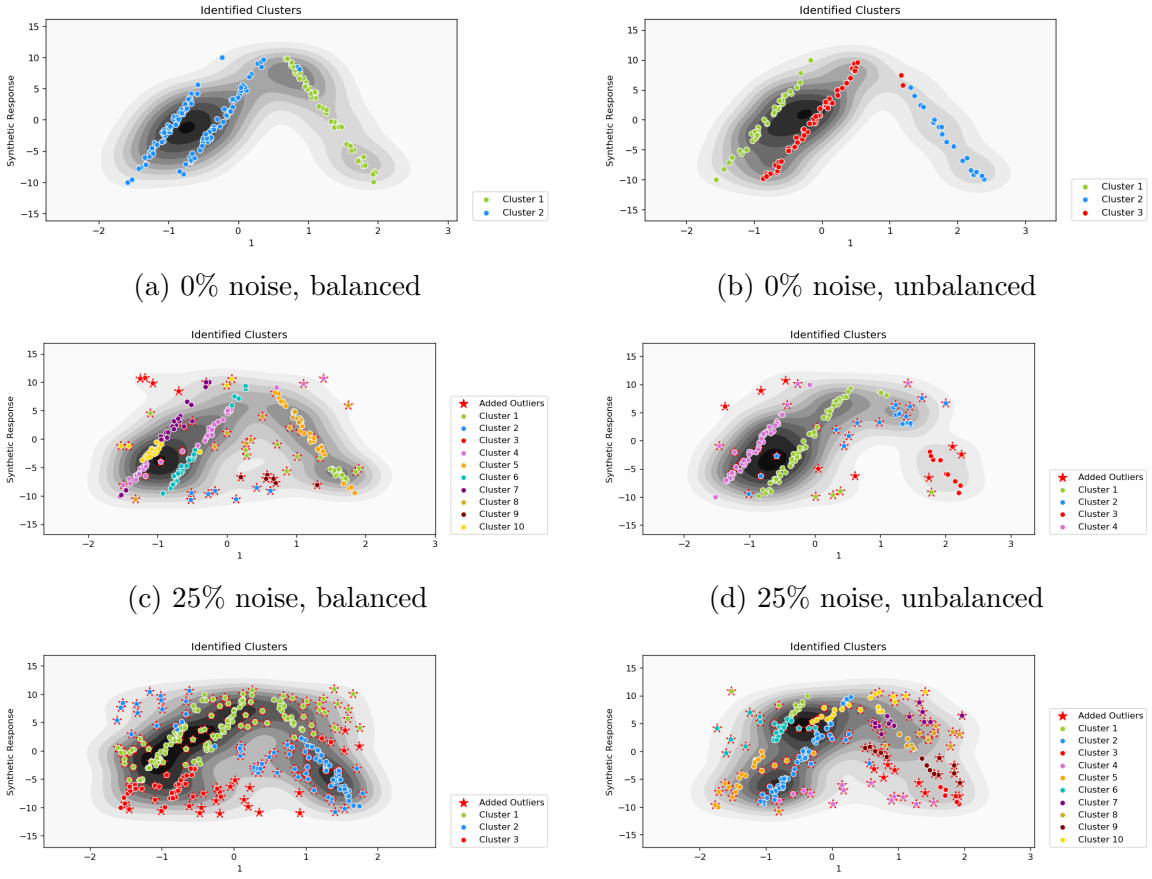
(c) Cases with 50% noise

Figure 5.4: Effect of cluster balance on performance. Results provided for bag size set to 5% of points.

We can observe that in Figure 5.4a cases 4 and 6 obtained a lower ARI scores on the unbalanced cluster condition while the clustering performance in case 5 improves under uneven cluster conditions. The results were mixed for higher noise levels as illustrated in Figures 5.4b and 5.4c.

Although the algorithm performance was generally better for cases with balanced clusters, there were instances (eg. case 5 in Figures 5.4a ,5.4b and 5.4c) that the proposed algorithm obtained a higher ARI score for cases with unbalanced clusters.

We further investigated these results by looking at the actual clusters; the most prominent differentiating factor was the estimated number of clusters. Figure 5.5 shows the differentiation in the number of clusters and the obtained clusters for case 5. The experiment in Figure 5.4 suggest that sensitivity in number of clusters estimation could be an important factor in the performance of the overall system. In the next section, we investigate the number of clusters estimation.



(a) 0% noise, balanced

(b) 0% noise, unbalanced

(c) 25% noise, balanced

(d) 25% noise, unbalanced

(e) 50% noise, balanced

(f) 50% noise, unbalanced

Figure 5.5: Illustration of Case 5 with 5% bag size

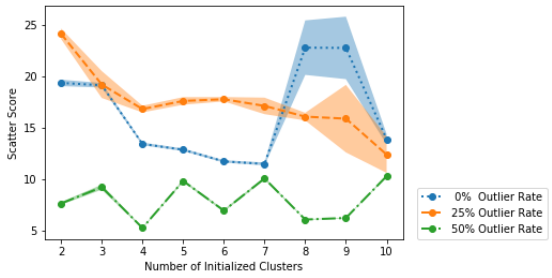
5.2.4 Scatter Scores

As discussed in chapter 3, we use Equation 3.6 to estimate the number of clusters in a data set. In this section, we examine the accuracy of this method for determining the number of clusters using our simulated data set.

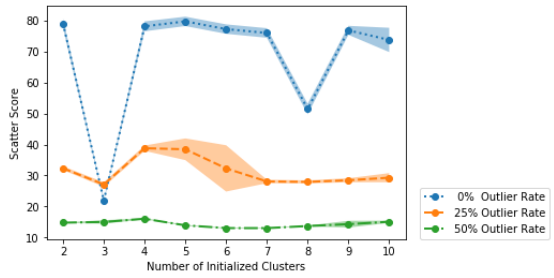
Figures 5.6 and 5.7 show the proposed scatter ratio for each of the simulated cases at different noise levels and cluster balance conditions. For each case, the bag size was set to 5% of the data. In particular, Figure 5.6 demonstrates the result for cases with 2 clusters and Figure 5.6 includes the results for cases with 3 clusters. Note that the results are based on 10 runs of the pipeline, allowing the estimation of confidence intervals around each scatter line. As seen in Figures 5.6 and 5.7, out of the 36 scatter score lines, we can observe that the first local maximum occurs at the correct number of clusters 19 times. Additionally, in 10 instances the estimated number of clusters

falls within ± 1 of the actual number of clusters. The algorithm fails to detect the correct number of clusters 7 times ($n_{estimated} > n_{actual} + 2$ or $n_{estimated} < n_{actual} - 2$).

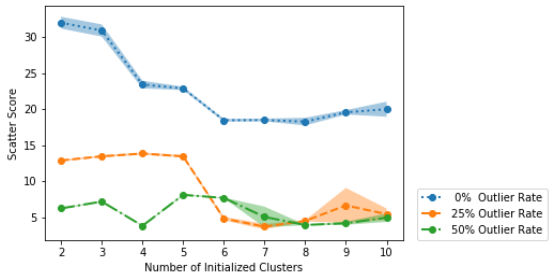
In some instances such as case 3 in Figures 5.6e and 5.6f (first local maximum at 2 clusters) our expectations were met for different noise levels and balanced and unbalanced clusters. Although there were cases that showed mixed results such as case 1 in Figures 5.6a and 5.6b (expected first local maximum at 2 clusters) and case 4 in Figures 5.7a and 5.7b (expected first local maximum at 3 clusters), we can observe that the scatter scores (ϕ_n) become frailer as noise increases. Additionally, case 6 in Figures 5.7e and 5.7f we can observe that the same scatter scores also become weaker when we compare balanced cluster to unbalanced sets.



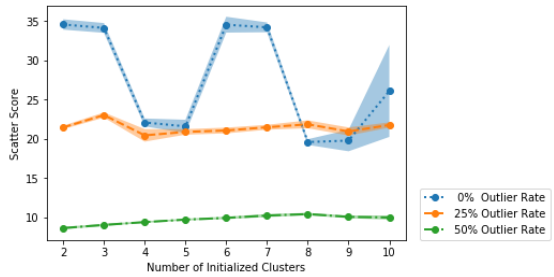
(a) Case 1 - balanced set



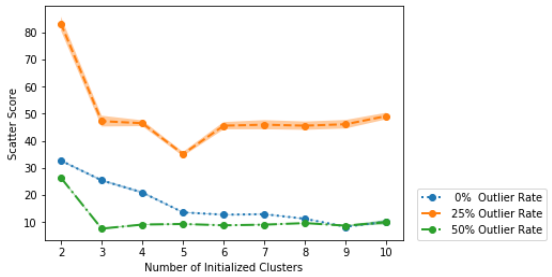
(b) Case 1 - unbalanced set



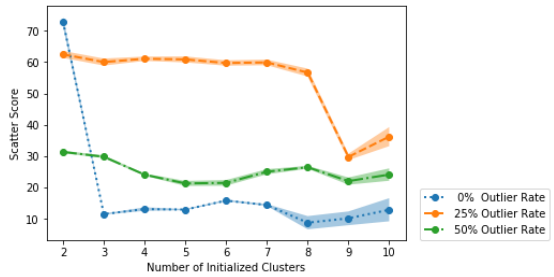
(c) Case 2 - balanced set



(d) Case 2 - unbalanced set

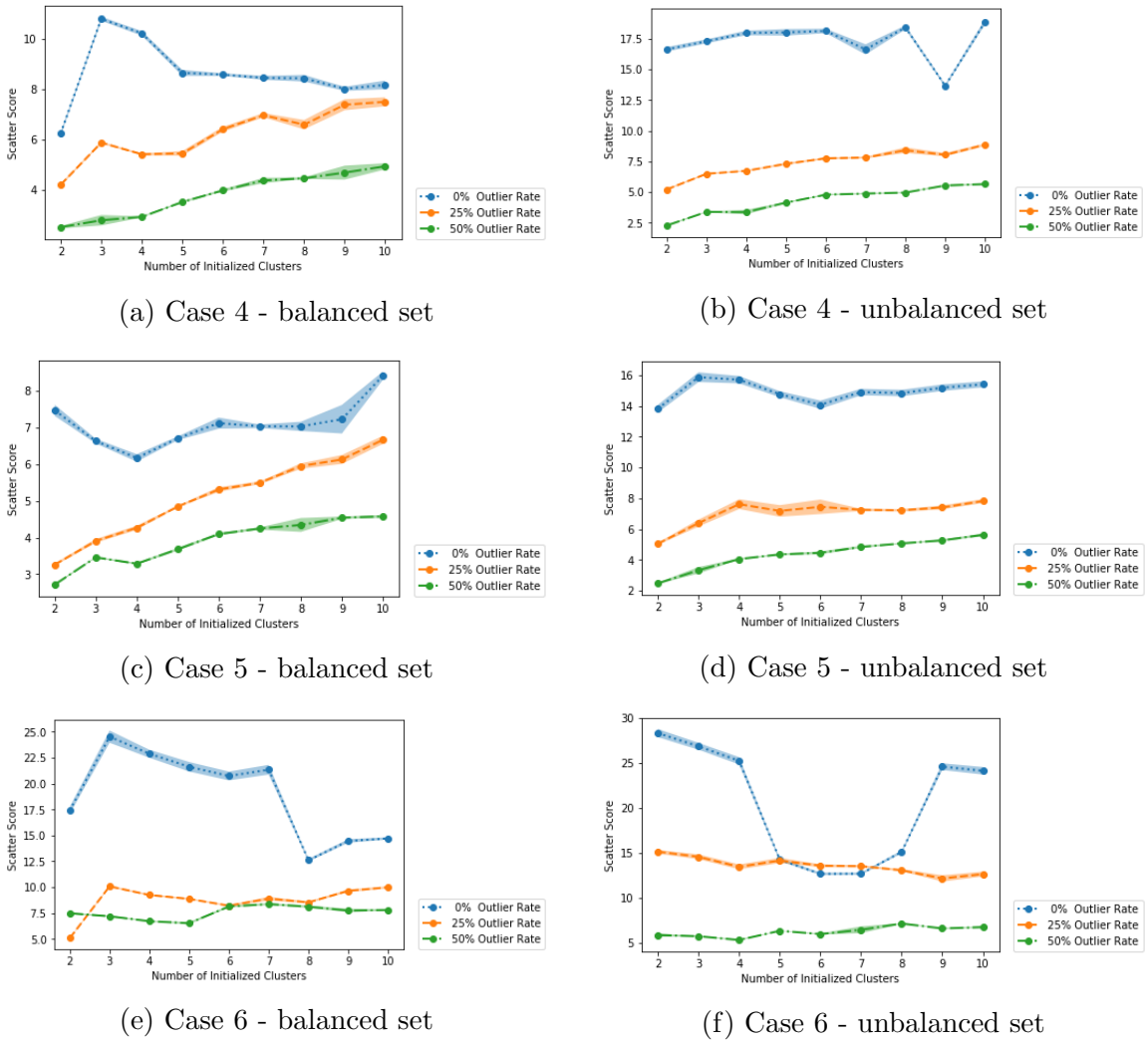


(e) Case 3 - balanced set



(f) Case 3 - unbalanced set

Figure 5.6: The scatter ratios under different conditions in sets with 2 clusters



(a) Case 4 - balanced set

(b) Case 4 - unbalanced set

(c) Case 5 - balanced set

(d) Case 5 - unbalanced set

(e) Case 6 - balanced set

(f) Case 6 - unbalanced set

Figure 5.7: The scatter ratios under different conditions in sets with 3 clusters

5.2.5 Clustering Result Illustrations

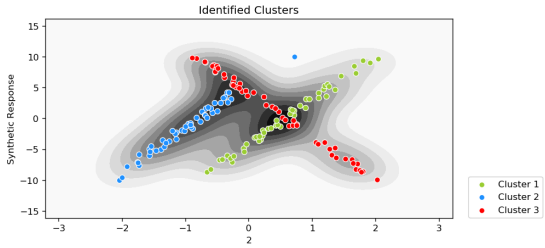
In this section, we include example visualizations of clustering results to better explain demonstrate how the proposed pipeline works. Examples from both successful and unsuccessful cases are chosen to highlight the strengths and weaknesses of the algorithm.

Examples of successful clustering

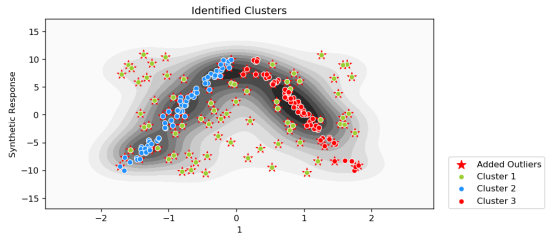
Figure 5.8 illustrates example of cases where the algorithm successfully recovered data clusters. Figures 5.8a and 5.8b demonstrate retrieval of 2 and 3 clusters under similar

conditions of 0% noise and 5% bag size for balanced and unbalanced clusters. On the other hand, Figures 5.8c and 5.8d show retrieval of two clusters using bag sizes of 33% and 63.2% when the clusters are balanced and 50% of the data set is consisted of random points.

(a) Case 3, 0% noise, 5% bag size, unbalanced



(b) Case 6, 0% noise, 5% bag size, balanced



(c) Case 2, 50% noise, 33% bag size, balanced

(d) Case 2, 50% noise, 63.2% bag size, balanced

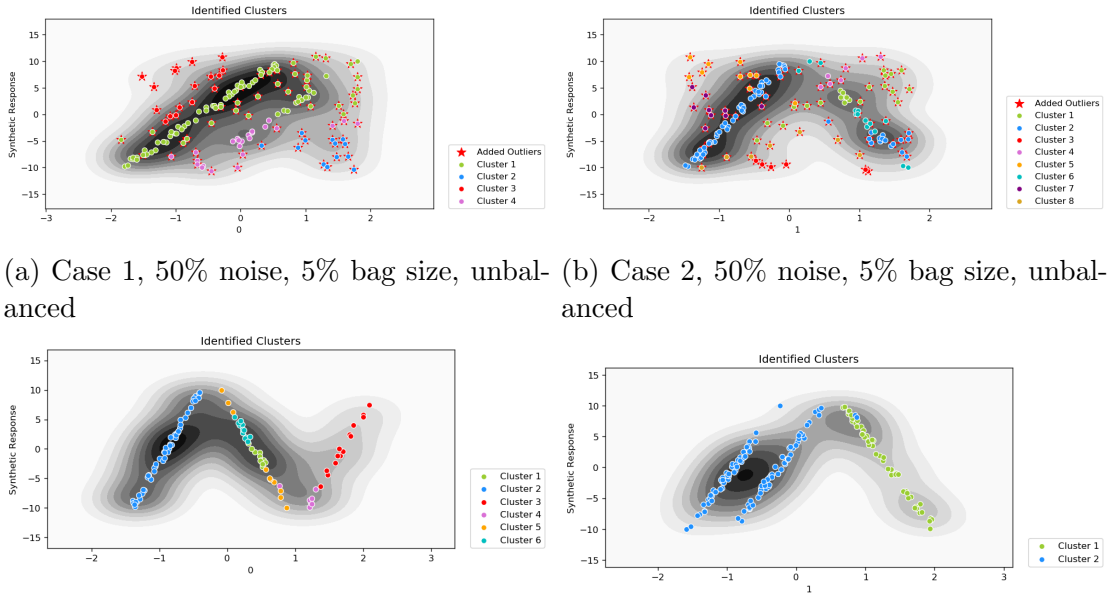
Figure 5.8: Successful Clustering Illustrations

Examples of unsuccessful clustering

Figure 5.9 visualizes examples where the proposed pipeline fails to recover the data clusters. Figures 5.9a and 5.9b are example cases with 50% noise and bag size of 5% when the clusters are unbalanced. Figures 5.9c and 5.9d illustrate failed cases at 0% noise level, 5% bagging rate for 3 unbalanced and balanced clusters respectively.

5.3 Method Comparison

We compared the performance of B-RC to RansaCov [80, 79] on simulated data with balanced clusters. B-RC bag size was set to 5% and ran for 100,000 iterations. RansaCov's threshold (threshold for detecting inliers) was empirically optimized on case 1 at 0% noise and kept the same in all experiments (similar approach to [80]).



(a) Case 1, 50% noise, 5% bag size, unbalanced

(b) Case 2, 50% noise, 5% bag size, unbalanced

(c) Case 4, 0% noise, 5% bag size, unbalanced

(d) Case 5, 0% noise, 5% bag size, balanced

Figure 5.9: Failed Clustering Illustrations

Both methods automatically estimated the number of clusters (B-RC: explained in section 3.2.2 and RansaCov: "Set Cover" approach explained in [79]). Our method performed better (as measured by adjusted rand index (ARI)) under the tested scenarios as shown in Tables 5.1 and 5.2.

0% noise	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
B-RC	**1**	**1**	**1**	**1**	0.52	**0.96**
RansaCov	**1**	0.88	0.73	0.74	**0.78**	0.65

Table 5.1: Performance Comparison at 0% noise on balanced clusters

25% noise	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
B-RC	**0.56**	**0.73**	**0.41**	**0.58**	**0.30**	**0.57**
RansaCov	0.14	0.27	0.15	0.35	0.22	0.18

Table 5.2: Performance Comparison at 25% noise on balanced clusters

5.4 POND Data

In this section, we report on the results of applying the proposed pipeline to the POND data set described in Chapter 4. Each of the 228 brain features were run individually through the pipeline.

As with the simulated set, we ran each feature 10 times (each for 100,000 iterations - K) through the pipeline. Each run provided us with an affinity matrix and scatter score (ϕ_n) estimation for different number of initialized clusters. The number of clusters are estimated using the average of these scatter scores. We averaged the obtained affinity matrices from theses runs to perform spectral clustering using the estimated number of clusters. The results reported in the remainder of this section are generated using 5% for the bag size. We picked the 5% bag size because of the obtained results on simulated data.

5.4.1 Significant Brain Regions

We used the concept of permutation importance to determine which of the brain features results in significant clusters [3]. To this end, we compared the optimum ϕ_n score (ϕ_n score for the estimated number of clusters) of each feature to a distribution of 2280 ϕ'_n scores obtained from randomly permuted features. We permuted each brain feature 10 times (for 100,000 iterations each) then calculated an affinity matrix for each which resulted in 2280 random affinity matrices. We performed spectral clustering using 2 to 10 initialized clusters which resulted in 9 random ϕ' score distributions. We examined the optimum ϕ_n score of each feature against the corresponding (obtained from the same number of clusters as the feature) random ϕ'_n score distribution.

Additionally, we ran 100 randomly generated features through the pipeline which provided us with random ϕ' score distribution. The affinity matrices were averaged over 10 runs (100,000 iterations at each run) for each feature; each affinity matrix generated 9 different ϕ' scores at different number of clusters. We examined the brain ϕ_n scores to these random ϕ'_n score distribution. The comparison of the ϕ_n scores provided us with an additional significance measure. This measure is reported as complimentary consideration to previous permutation test results. For more details on these significant tests please visit chapter 3.3.2.

Of the 228 brain features, 12 had scatter ratios that were significantly different than chance after FDR correction as determined by the permutation test and contained clusters with $Adj.R^2 > 0.3$. Table 5.3 shows a list of these features. Of the 12 significant features, 4 were surface area features, 6 were cortical volume features, and 2 were cortical thickness features.

	Feature	Regions	ϕ_n	Adj. p
1	Lobe Area	Left Middle Frontal Gyrus	5.6243	0.0429
2	Lobe Area	Left Precuneus*	6.1461	0.0000
3	Lobe Area	Right Inferior Frontal Gyrus Triangular part	5.3737	0.0000
4	Lobe Area	Right Insula	5.8284	0.0200
5	Lobe Volume	Left Inferior Occipital Gyrus*	6.1566	0.0200
6	Lobe Volume	Right Gyrus Rectus*	5.7259	0.0000
7	Lobe Volume	Right Inferior Frontal Gyrus Orbital part*	6.2704	0.0000
8	Lobe Volume	Right Lingual Gyrus*	5.5909	0.0200
9	Lobe Volume	Right Heschl Gyrus*	6.3542	0.0000
10	Lobe Volume	Right Anterior Cingulate and Paracingulate Gyri*	6.1364	0.0200
11	Lobe Thickness	Left Calcarine Fissure and surrounding Cortex	5.5804	0.0429
12	Lobe Thickness	Right Anterior Cingulate and Paracingulate Gyri	5.9995	0.0333

Table 5.3: Brain regions with significant brain-behaviour correlates (*also had adj. $p < 0.01$ when compared to randomly generated data)

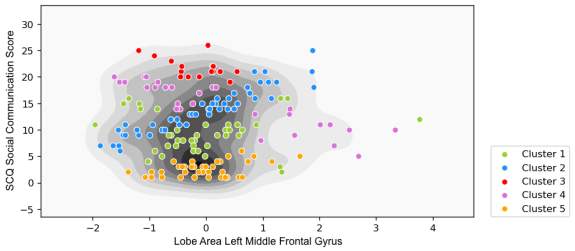
5.4.2 Cluster characteristics

For each significant features, cluster characteristics were examine by fitting a traditional linear model to the points in that cluster. We also characterized the clusters based on SCQ scores, IQ scores, and proportion of diagnostic labels within each cluster.

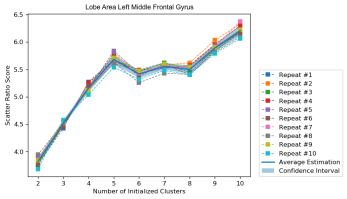
We performed OLS (ordinary least square) regression analysis on each of the identified clusters to identify the robustness and significance of each cluster compared to other clusters found in a single brain region. We report on 4 different statistics of each cluster. First, the population of each cluster which is the number of samples assigned to a specific cluster. Second, the cluster adj. R^2 which is obtained from an OLS analysis and describes the variance explained in the response by the input. Third,

the coefficient of the input from the OLS analysis which is an indicator of the type of response-input relations as well as its strength. Please visit chapter 3 for greater details on these measurements. The measurements are reported in Figures 5.10h, 5.11h, 5.12h, 5.13h, 5.14h, 5.15h, 5.16h, 5.17h, 5.18h, 5.19h, 5.20h, 5.21h. The number of discovered clusters ranged from 4 to 9 (2 to 7 clusters with adj. R^2 of higher than 0.3) and the majority of the discovered clusters did not align with diagnostic labels. There was no effect of FSIQ between any of the cluster, but in some regions there was an effect of between the SCQ scores which is included in the description of results for each region. We used the Kruskal-Wallis [68] to determine group effects and the MannWhitney [81] to test pairwise analysis and corrected for multiple comparisons by the Bonferroni criteria [21].

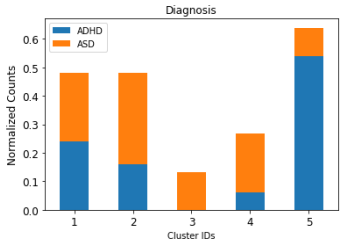
Figure 5.10 shows a summary of the identified clusters in the area of the left middle frontal gyrus. There are 2 clusters (# 2&4) with Adj. R^2 of greater than 0.3. There are no significant effects between the SCQ scores of clusters 2 and 4.



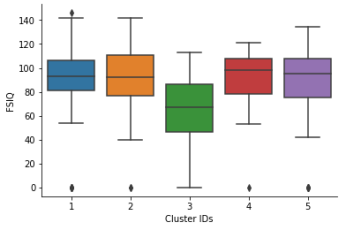
(a) Clusters and Sample Density



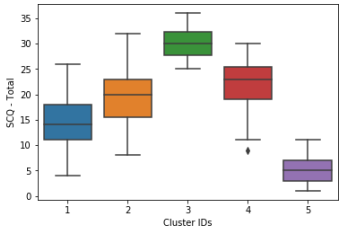
(b) Scatter Score Lines



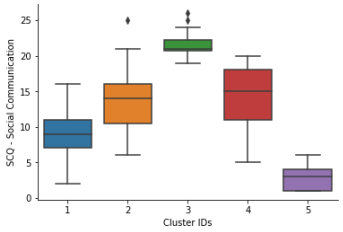
(c) Clinical Diagnosis



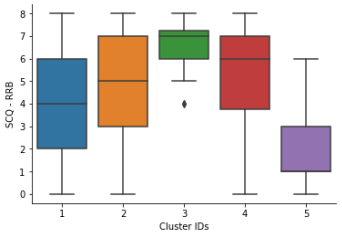
(d) FSIQ Distributions



(e) SCQ Totoal Score



(f) SCQ Social Communication Sub-score



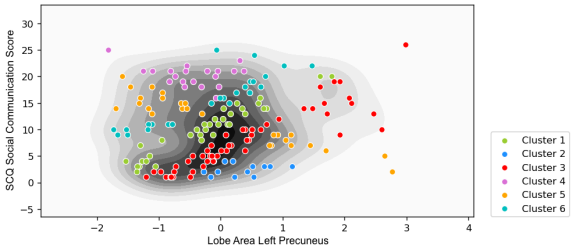
(g) SCQ RRB. Sub-score

Cluster ID	Population	Cluster Adj. R^2	Coefficient
1	41	-0.026	0.0226
2	47	0.878	3.9508
3	16	-0.26	-1.9387
4	28	0.765	-2.6334
5	39	0.228	1.067

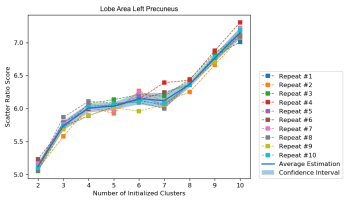
(h) Cluster Statistics in Lobe Area Left Middle Frontal Gyrus

Figure 5.10: Cluster Characteristics in Lobe Area Left Middle Frontal Gyrus

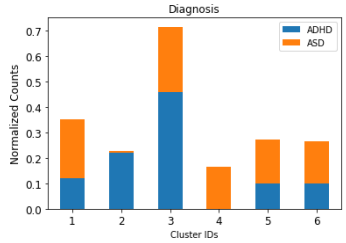
Figure 5.11 shows a summary of the identified clusters in the area of the left precuneus. There are 4 clusters (# 1,3,5&6) with Adj. R^2 of greater than 0.3. Clusters 3 and 6 have significantly ($p < 0.001$) different total SCQ scores.



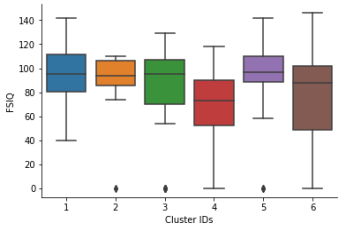
(a) Clusters and Sample Density



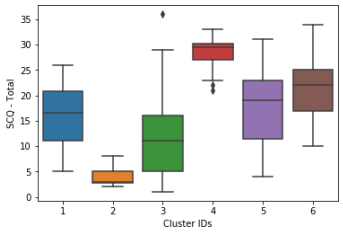
(b) Scatter Score Lines



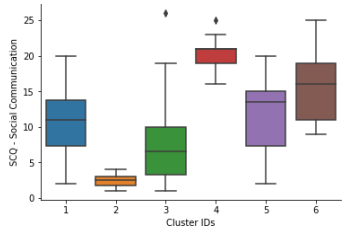
(c) Clinical Diagnosis



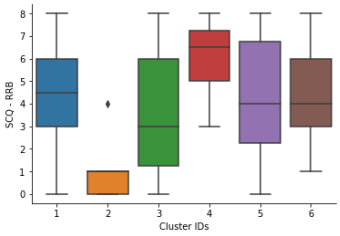
(d) FSIQ Distributions



(e) SCQ Totoal Score



(f) SCQ Social Communication Sub-score



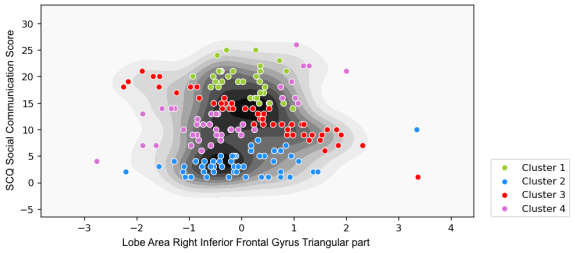
(g) SCQ RRB. Sub-score

Cluster ID	Population	Cluster Adj. R^2	Coefficient
1	34	0.923	5.4704
2	12	-0.017	0.7001
3	54	0.828	4.8451
4	20	0.041	-1.1032
5	26	0.884	-3.3701
6	25	0.769	4.7328

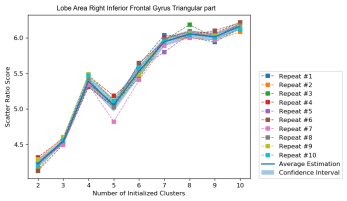
(h) Cluster Statistics in Lobe Area Left Precuneus

Figure 5.11: Cluster Characteristics in Lobe Area Left Precuneus

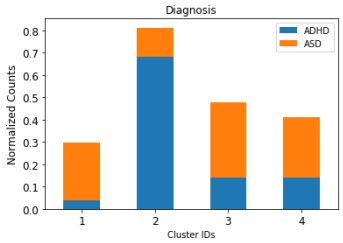
Figure 5.12 shows a summary of the identified clusters in the area of the right inferior frontal gyrus. There are 2 clusters (# 3&4) with Adj. R^2 of greater than 0.3. There are no significant effects between the SCQ scores of clusters 3 and 4.



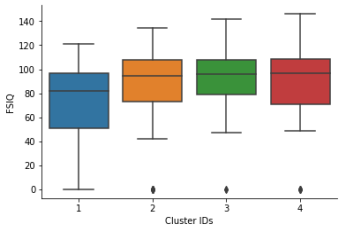
(a) Clusters and Sample Density



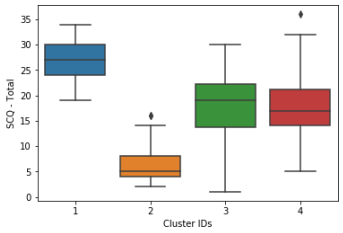
(b) Scatter Score Lines



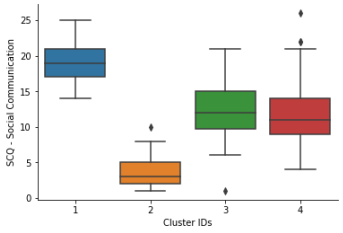
(c) Clinical Diagnosis



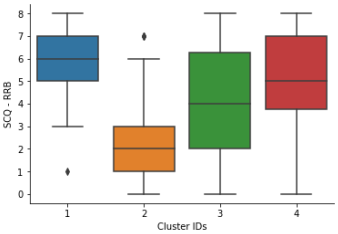
(d) FSIQ Distributions



(e) SCQ Totoal Score



(f) SCQ Social Communication Sub-score



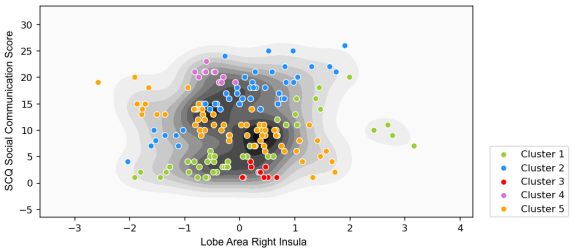
(g) SCQ RRB. Sub-score

Cluster ID	Population	Cluster Adj. R^2	Coefficient
1	33	0.12	-2.2909
2	50	0.15	0.9258
3	48	0.91	-3.2315
4	40	0.52	3.521

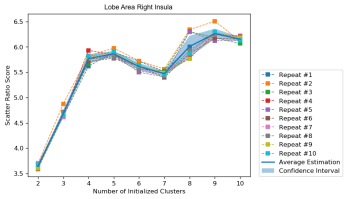
(h) Cluster Statistics in Lobe Area Right Inferior Frontal Gyrus Triangular part

Figure 5.12: Cluster Characteristics in Lobe Area Right Inferior Frontal Gyrus Triangular part

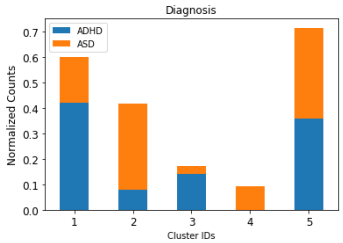
Figure 5.13 shows a summary of the identified clusters in the area of the right insula. There are 3 clusters (# 1&2&5) with Adj. R^2 of greater than 0.3. All three clusters were significantly different from each other ($p < 0.01$) in the three report categories of the SCQ score (except for the clusters 1&5 in restricted repetitive behaviour sub score of the SCQ).



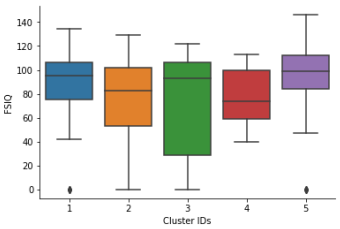
(a) Clusters and Sample Density



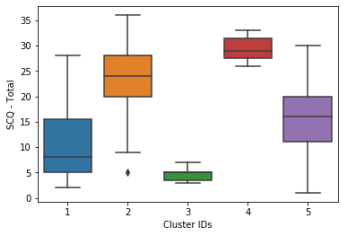
(b) Scatter Score Lines



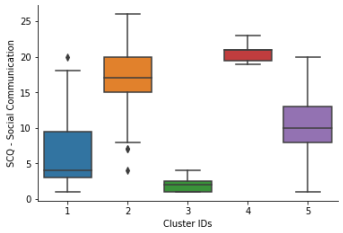
(c) Clinical Diagnosis



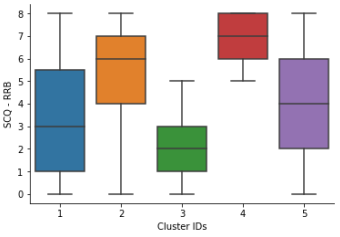
(d) FSIQ Distributions



(e) SCQ Totoal Score



(f) SCQ Social Communication Sub-score



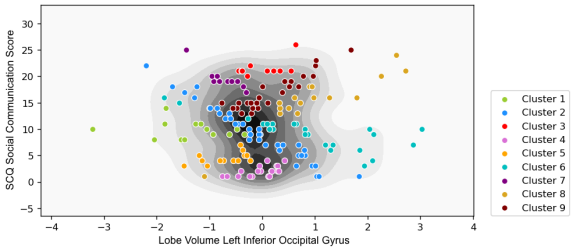
(g) SCQ RRB. Sub-score

Cluster ID	Population	Cluster Adj. R^2	Coefficient
1	43	0.55	3.0441
2	45	0.71	4.6445
3	11	-0.08	-0.9684
4	11	0.273	-3.2974
5	61	0.703	-3.2108

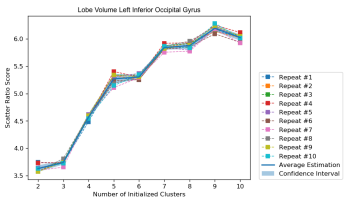
(h) Cluster Statistics in Lobe Area Right Insula

Figure 5.13: Cluster Characteristics in Lobe Area Right Insula

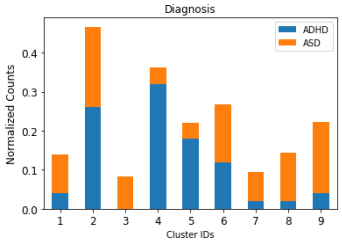
Figure 5.14 shows a summary of the identified clusters in the volume of the left inferior occipital gyrus. There are 5 clusters (# 2,5,7,8&9) with Adj. R^2 of greater than 0.3. Clusters 2 and 5 were significantly different ($p < 0.001$) from clusters 5,7,8&9 and 2,7,8&9 in total SCQ scores and the social communication sub scores. Clusters 2 and 9 had also significantly ($p < 0.001$) different restricted repetitive behaviour sub scores.



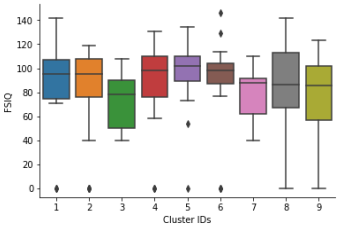
(a) Clusters and Sample Density



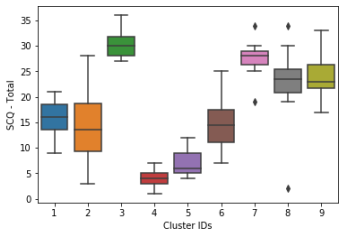
(b) Scatter Score Lines



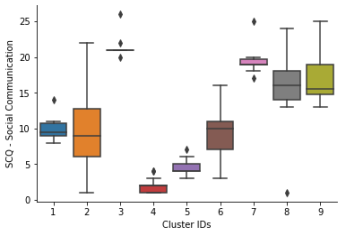
(c) Clinical Diagnosis



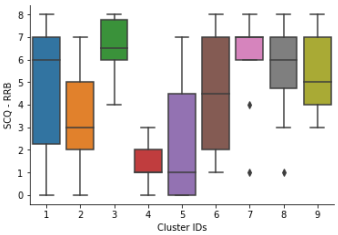
(d) FSIQ Distributions



(e) SCQ Totoal Score



(f) SCQ Social Communication Sub-score



(g) SCQ RRB. Sub-score

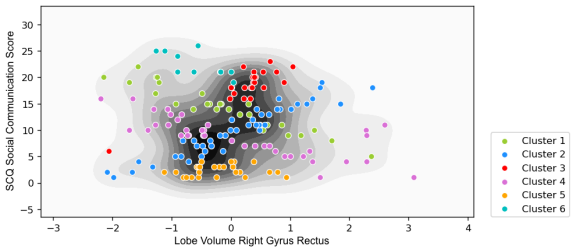
Cluster ID	Population	Cluster Adj. R^2	Coefficient
1	14	-0.05	-0.3611
2	38	0.95	-5.6137
3	10	0.185	2.1211
4	21	0.19	2.1211
5	14	0.31	1.7141
6	24	0.23	1.4555
7	10	0.56	-1.8912
8	16	0.79	-5.6144
9	24	0.77	4.5247

(h) Cluster Statistics in Lobe Volume Left Inferior Occipital Gyrus

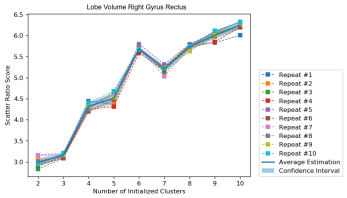
Figure 5.14: Cluster Characteristics in Lobe Volume Left Inferior Occipital Gyrus

Figure 5.15 shows a summary of the identified clusters in the volume of the right gyrus rectus. There are 5 clusters (# 1,2,3,4&6) with Adj. R^2 of greater than 0.3. Total SCQ scores were significantly ($p < 0.001$) between cluster 6 and clustes 1,2,3&4; additionally clusters 2 and 4 were different from clusters 3 and 1. There were also

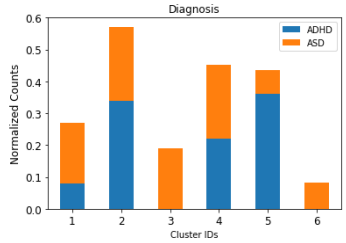
significant effects between the groups in social communication sub scores; cluster 6 was significantly different from clusters 4,2&1; clusters 4 and 2 were significantly different from clusters 3 and 1; clusters 1 and 3 were also significantly different from each other. There were no significant effects in the restricted repetitive behaviour sub scores.



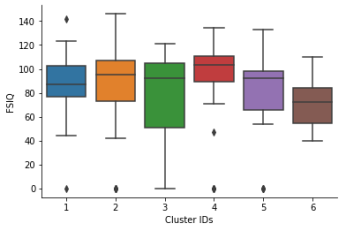
(a) Clusters and Sample Density



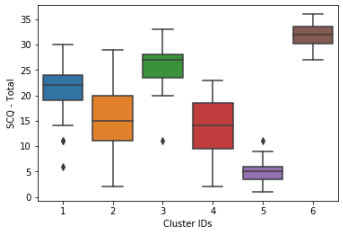
(b) Scatter Score Lines



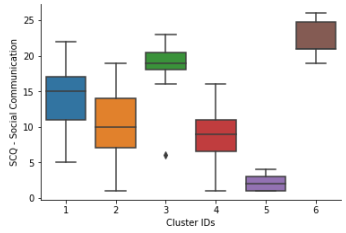
(c) Clinical Diagnosis



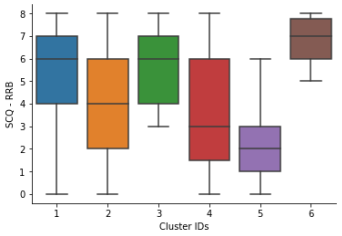
(d) FSIQ Distributions



(e) SCQ Totoal Score



(f) SCQ Social Communication Sub-score



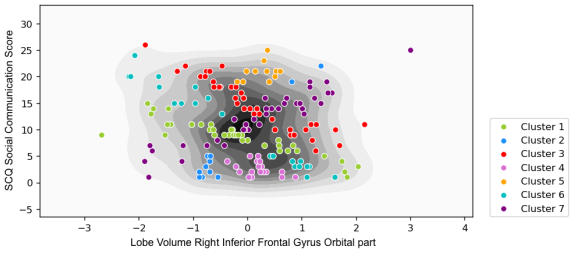
(g) SCQ RRB. Sub-score

Cluster ID	Population	Cluster Adj. R^2	Coefficient
1	27	0.902	-3.6107
2	45	0.902	4.412
3	23	0.76	5.1261
4	39	0.545	-2.02
5	27	-0.04	-0.0258
6	10	0.44	-3.5482

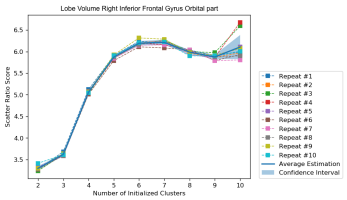
(h) Cluster Statistics in Lobe Volume Right Gyrus Rectus

Figure 5.15: Cluster Characteristics in Lobe Volume Right Gyrus Rectus

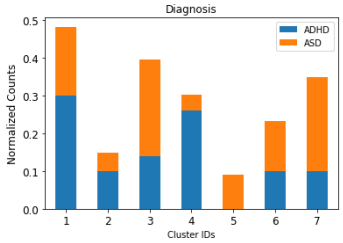
Figure 5.16 shows a summary of the identified clusters in the volume of the right inferior frontal gyrus orbital part. There are 5 clusters (# 1,2,3,6&7) with Adj. R^2 of greater than 0.3. Cluster 3 was significantly ($p < 0.001$) different from clusters 1 and 2 in the total SCQ score and the social communication sub score. Furthermore, cluster 7 was significantly different from cluster 1 in social communication sub score.



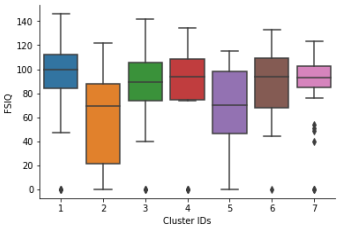
(a) Clusters and Sample Density



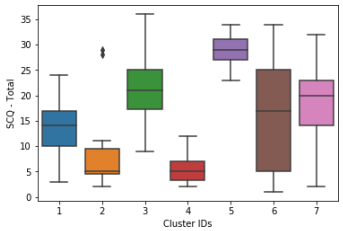
(b) Scatter Score Lines



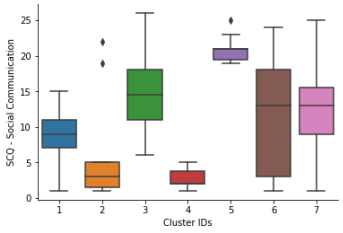
(c) Clinical Diagnosis



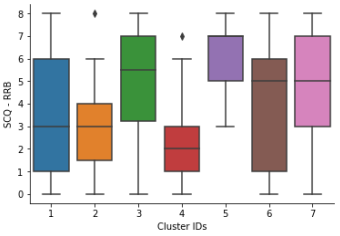
(d) FSIQ Distributions



(e) SCQ Totoal Score



(f) SCQ Social Communication Sub-score



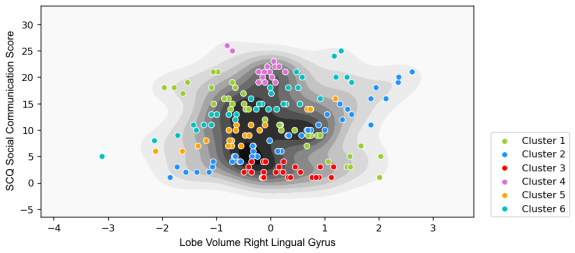
(g) SCQ RRB. Sub-score

Cluster ID	Population	Cluster Adj. R^2	Coefficient
1	37	0.784	-2.4372
2	11	0.945	9.4992
3	38	0.844	-4.9769
4	18	0.107	-1.7503
5	11	0.007	2.2312
6	21	0.955	-5.9071
7	35	0.89	4.2547

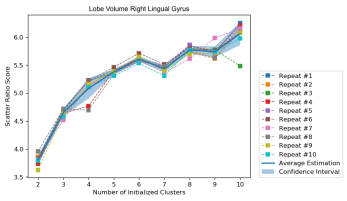
(h) Cluster Statistics in Lobe Volume Right Inferior Frontal Gyrus Orbital part

Figure 5.16: Cluster Characteristics in Lobe Volume Right Inferior Frontal Gyrus Orbital part

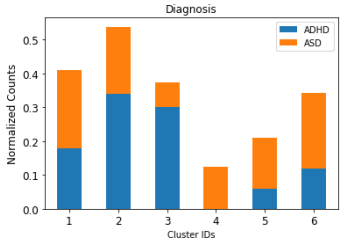
Figure 5.17 shows a summary of the identified clusters in the volume of the right right lingual gyrus. There are 5 clusters (# 1,2,4,5&6) with Adj. R^2 of greater than 0.3. Cluster 4 had significantly ($p < 0.001$) different total SCO and social communication sub scores from clusters 1,2,5&6. Additionally, cluster 6 was significantly different from clusters 2&5 in total SCQ and social communication sub scores. Clusters 4 and 2 were different in restricted repetitive behaviour sub score.



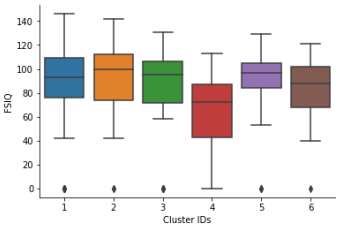
(a) Clusters and Sample Density



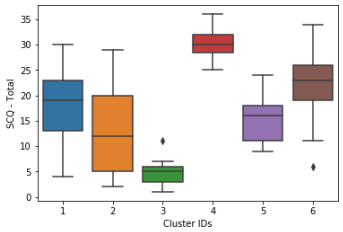
(b) Scatter Score Lines



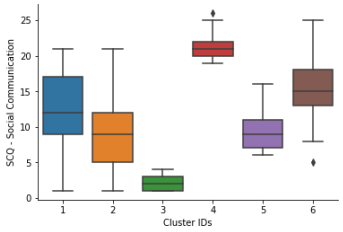
(c) Clinical Diagnosis



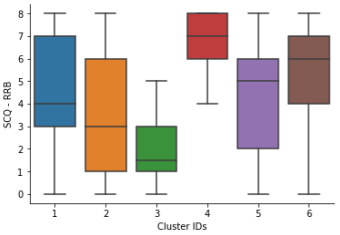
(d) FSIQ Distributions



(e) SCQ Totoal Score



(f) SCQ Social Communication Sub-score



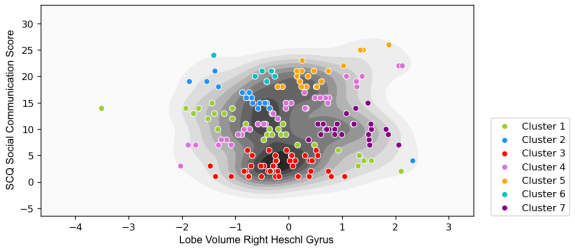
(g) SCQ RRB. Sub-score

Cluster ID	Population	Cluster Adj. R^2	Coefficient
1	37	0.867	-4.6723
2	41	0.927	4.251
3	24	0.019	-0.557
4	15	0.34	-4.3664
5	21	0.831	3.2765
6	33	0.825	3.7344

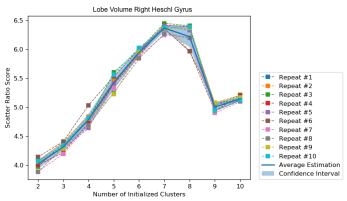
(h) Cluster Statistics in Lobe Volume Right Lingual Gyrus

Figure 5.17: Cluster Characteristics in Lobe Volume Right Lingual Gyrus

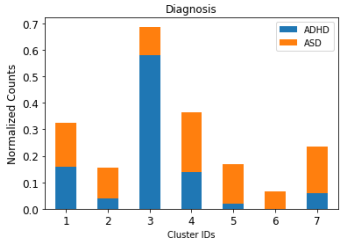
Figure 5.18 shows a summary of the identified clusters in the volume of the right heschl gyrus. There are 5 clusters (# 1,2,4,5&6) with Adj. R^2 of greater than 0.3. In total SCQ and social communication sub scores, cluster 1 and 2 were significantly different ($p < 0.001$) from each other; clusters 5 and 6 were different from clusters 1,2&4. Cluster 1 and 5 were significantly different in restrictive repetitive behaviour sub scores.



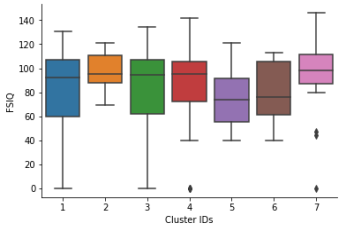
(a) Clusters and Sample Density



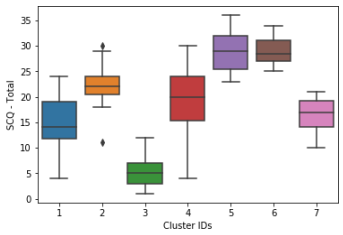
(b) Scatter Score Lines



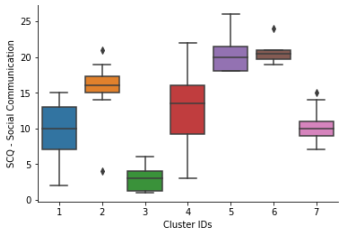
(c) Clinical Diagnosis



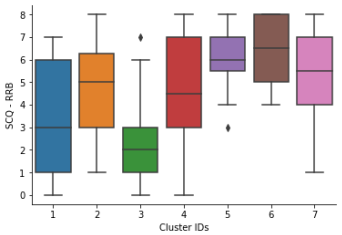
(d) FSIQ Distributions



(e) SCQ Totoal Score



(f) SCQ Social Communication Sub-score



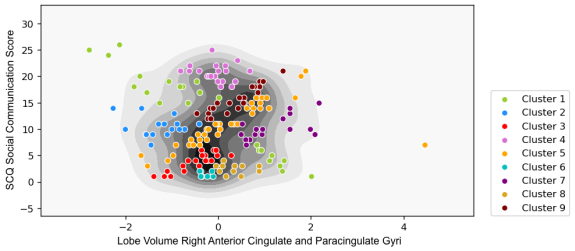
(g) SCQ RRB. Sub-score

Cluster ID	Population	Cluster Adj. R^2	Coefficient
1	28	0.849	-2.6972
2	16	0.936	-3.9302
3	42	-0.007	0.401
4	34	0.952	4.34
5	19	0.616	3.8742
6	8	0.559	-3.3761
7	24	-0.036	-0.402

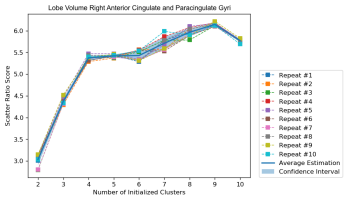
(h) Cluster Statistics in Lobe Volume Right Heschl Gyrus

Figure 5.18: Cluster Characteristics in Lobe Volume Right Heschl Gyrus

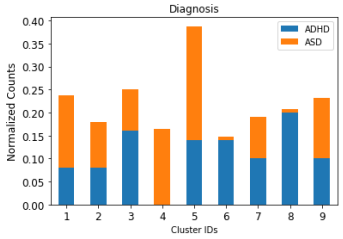
Figure 5.19 shows a summary of the identified clusters in the volume of the right anterior cingulate and paracingulate gyri. There are 5 clusters (# 1,3,5,7&9) with Adj. R^2 of greater than 0.3. In total SCQ and social communication sub scores, cluster 7 and 9 were significantly different ($p < 0.001$) from each other; cluster 3 was different from clusters 1,5,7&9.



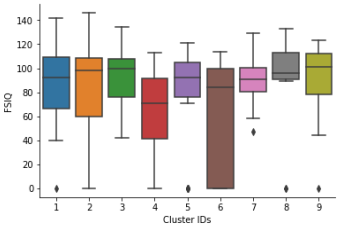
(a) Clusters and Sample Density



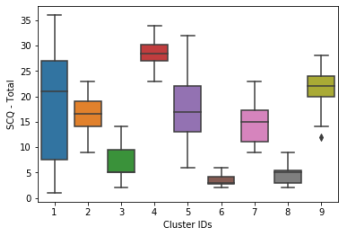
(b) Scatter Score Lines



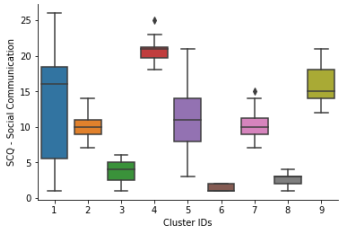
(c) Clinical Diagnosis



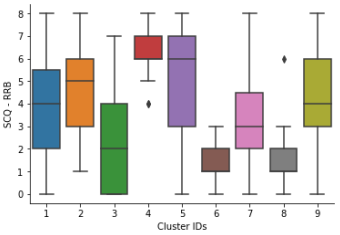
(d) FSIQ Distributions



(e) SCQ Totoal Score



(f) SCQ Social Communication Sub-score



(g) SCQ RRB. Sub-score

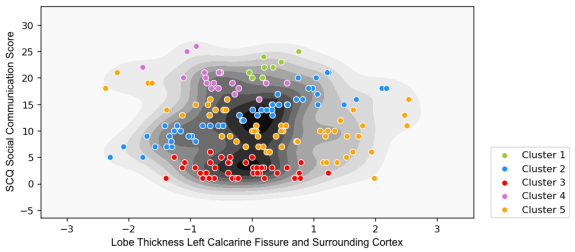
Cluster ID	Population	Cluster Adj. R^2	Coefficient
1	23	0.91	-5.4067
2	16	-0.045	-0.6205
3	19	0.736	3.296
4	20	-0.03	-0.7155
5	37	0.39	2.4437
6	8	-0.093	-1.0344
7	16	0.306	2.665
8	11	-0.094	-0.367
9	21	0.833	4.6371

(h) Cluster Statistics in Lobe Volume Right Anterior Cingulate and Paracingulate Gyri

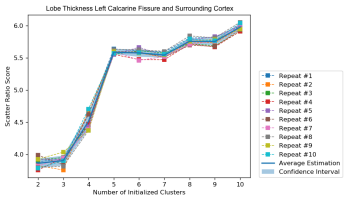
Figure 5.19: Cluster Characteristics in Lobe Volume Right Anterior Cingulate and Paracingulate Gyri

Figure 5.20 shows a summary of the identified clusters from the cortical thickness of the left calcarine fissure and surrounding cortex. There are 3 clusters (# 1,2&5) with Adj. R^2 of greater than 0.3. In total SCQ and social communication sub scores,

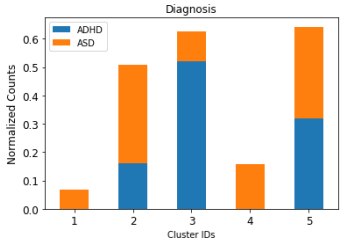
cluster 1 was significantly different ($p < 0.01$) from clusters 2 and 5. Additionally, cluster 2 was different from cluster 5 in social communication sub score.



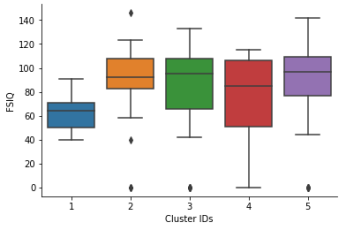
(a) Clusters and Sample Density



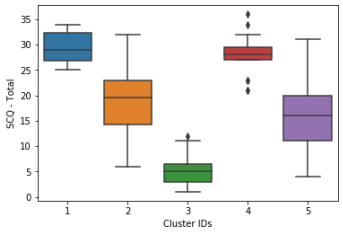
(b) Scatter Score Lines



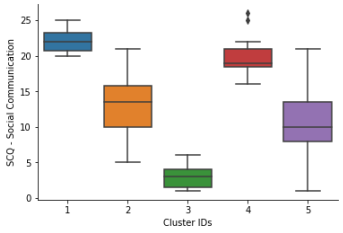
(c) Clinical Diagnosis



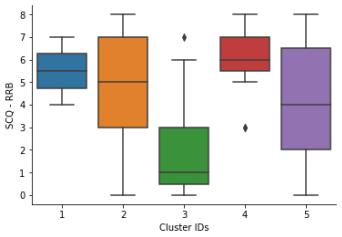
(d) FSIQ Distributions



(e) SCQ Totoal Score



(f) SCQ Social Communication Sub-score



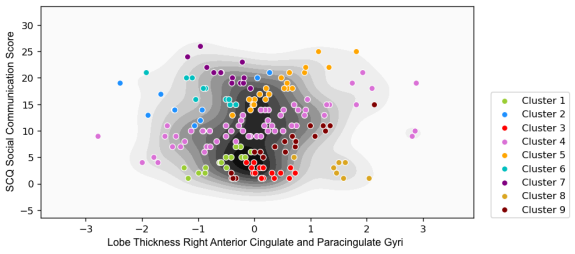
(g) SCQ RRB. Sub-score

Cluster ID	Population	Cluster Adj. R^2	Coefficient
1	8	0.535	5.4007
2	50	0.833	3.3906
3	39	0.005	-0.3994
4	19	0.274	-2.6793
5	55	0.314	-2.0443

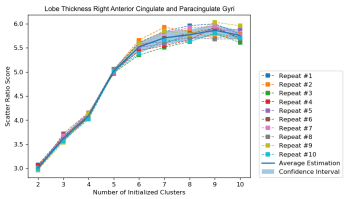
(h) Cluster Statistics in Lobe Thickness Left Calcarine Fissure and surrounding Cortex

Figure 5.20: Cluster Characteristics in Lobe Thickness Left Calcarine Fissure and surrounding Cortex

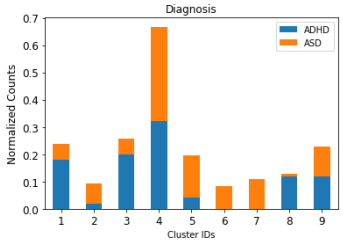
Figure 5.21 shows a summary of the identified clusters from the cortical thickness of the right anterior cingulate and paracingulate gyri. There are 7 clusters (# 1,4,5,6,7,8&9) with Adj. R^2 of greater than 0.3. In total SCQ and social communication sub scores, clusters 5, 6 and 7 were significantly different ($p < 0.001$) from clusters 1,4,8&9; cluster 4 was also different from clusters 1,8&9. Additionally, cluster 1 was different from clusters 4,5&7 in restrictive repetitive behaviour sub score; cluster 7 was also different from clusters 8 and 9.



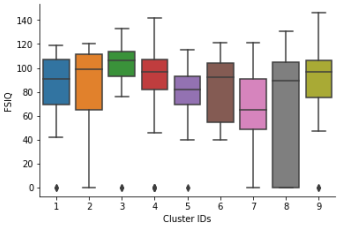
(a) Clusters and Sample Density



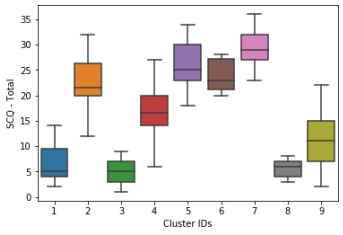
(b) Scatter Score Lines



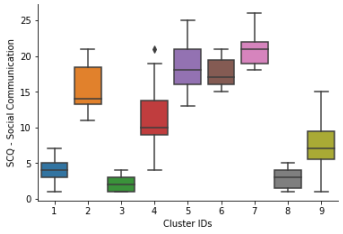
(c) Clinical Diagnosis



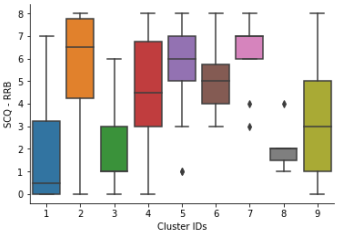
(d) FSIQ Distributions



(e) SCQ Totoal Score



(f) SCQ Social Communication Sub-score



(g) SCQ RRB. Sub-score

Cluster ID	Population	Cluster Adj. R^2	Coefficient
1	16	0.682	4.0647
2	10	0.024	1.5583
3	17	0.07	-1.4121
4	58	0.512	2.0962
5	21	0.872	5.7916
6	10	0.743	-4.0517
7	13	0.547	-5.0731
8	7	0.426	-2.8595
9	19	0.925	5.3327

(h) Cluster Statistics in Lobe Thickness Right Anterior Cingulate and Paracingulate Gyri

Figure 5.21: Cluster Characteristics in Lobe Thickness Right Anterior Cingulate and Paracingulate Gyri

Chapter 6

Discussion

6.1 Chapter Overview

We start by discussing the proposed B-RC algorithm in section 6.2. The findings on simulated and real word data are discussed in sections 6.3 and 6.4 respectively.

6.2 The B-RC Algorithm

In this thesis, we proposed a new regression clustering algorithm, the B-RC algorithm. The algorithm is novel in two ways: first, we build on the concept of bagging to explore the existence of linear correlations using a limited portion of the data set and see if there are other samples in the data set that can be explained by the same correlation pattern. Second, we propose a novel approach to characterize similarity between two points based on their relative distance to a regression line. This enables us to allow for the possibility of multiple regression lines in the data and allows us to build affinity matrices that can be used for clustering. This approach is advantageous when examining brain-behaviour relationships because as we discussed in 2 this relationship can be consisted of multiple subtypes with distinct brain-behaviour co relates where a single regression line may not be able to explain the entire data set.

We built on the concept of bagging to see subtypes and multiple relations compared to traditional regression methods. Further, we used robust regression models (eg.

RANSAC) which have good performance even in presence of outliers which was not the case for previous RC algorithms [133, 10]. Additionally, our method provides insight into the data by introducing a novel way of constructing an affinity matrix based on regression residuals and we introduced a within-to-between cluster similarity measure to estimate the number of potential clusters. Although in this thesis we demonstrated the proposed approach using the RANSAC regressor and spectral clustering models, the algorithm can easily be extended to employ other regression algorithms and affinity based clustering methods. These choices should be made based on the specific application.

6.3 Algorithm Performance on Simulated Data

We evaluated the proposed algorithm using a simulated data set. Our results showed high agreement between true cluster labels and those discovered by the proposed algorithm under low noise levels (ARI: 0.5-1). The algorithm was able to detect the correct clusters under different cluster balancing conditions (eg. Figures 5.4a and 5.4b). To further understand the behaviour of the proposed method, we examined its sensitivity to bag size, noise, and cluster balance.

6.3.1 Sensitivity to Bag Size

The bag size enables the discovery of multiple associations. Moreover, it sets the resolution of the algorithm; meaning, that bag size determines the portion size of the data set that search for existence of a correlate in a single iteration. In other words, a small bag size allows for discovering small clusters in the data set.

We examined the performance of the algorithm using bag sizes equal to 5%, 33% and 63.2% of the sample size. We expected that a small bag size would increase the chances of identifying the clusters in the data set. This expectation was based on the following: First, the number of possible combinations (different bags) increase with bag size until the bag size is approximately half the size of the entire set (using the binomial coefficients). Second, we can only run the algorithm for a limited number of iterations that is less that the total possible combinations of data points. Hence, we believed small bag size would perform better. Additionally, we hypothesized that

for large bag sizes, the pipeline may fail to identify clusters that are relatively small compared to the size of other clusters and random noise points. This expectation was based on the fact that larger bags would be guaranteed to include random noise points (or points from other clusters) that would make it difficult for the RANSAC regressor to pick up the smaller clusters in any of the iterations. Based on these hypotheses we expected the bag size of 5% to perform better than 33% and 63.2%.

Our results were generally consistent with the above hypotheses. However, as noise increased, we saw a plateau in ARI scores. In this case, since all random points shared the same label, a small number of agreements between the predicted and true labels existed, resulting in a constant and low ARI score which is above the absolute random ARI score of 0.

Our results also suggested that although a small bag size provides the best performance in most cases, larger bag sizes can be advantageous in situations with high noise levels by better preserving the larger clusters.

6.3.2 Sensitivity to Noise

As expected, our results showed that the performance of the proposed algorithm degrades with increasing noise. As more random points in a set increase, the more random regressions (regression lines that would suggest actual cluster data points are similar to other noisy points) would contribute to the similarity matrix, leading to noisier clusters. Despite the general performance degradation, there were instances that the algorithm performed well at the highest noise level of 50% (eg. Figure 5.8c and 5.8d). In these instances, the algorithm was initialized with bag sizes of 33% and 63.2% which could have helped to identify the clusters. Robustness to noise is an important factor when dealing with unknown data since the noise level in the data may not always be a known factor.

6.3.3 Detection of Variable Cluster Sizes

We examined the proposed method's ability to detect clusters of varying sizes. We expected for the algorithm to perform better in experiments with balanced clusters compared to unbalanced sets since all clusters should be represented evenly in the

affinity matrix (cluster are equally likely to be discovered due to having the same number of samples). This hypothesis was only partially supported by our results. Decreased performance for the balance cases seemed related to inaccuracies in estimating the number of clusters (eg. Figure 5.5).

6.3.4 Determining the Number of Clusters

We examined the effectiveness of the scatter score (ϕ_n) proposed in Equation 3.6 for automatically determining the number of clusters. As discussed in chapter 3, we expected to see a local maximum on the scatter line for the correct number of clusters. While this was in fact the case for many cases (eg. Figure 5.3a), the scatter ratio was not effective in identifying the correct number of cluster in some cases (eg. Figure 5.5).

Previous attempts to estimate the number of clusters, most notably the eigenvalue gap statistic [121, 126], have shown inconsistent performance under different conditions [28]. Our results suggests that the proposed method for estimating the number of clusters can be effective in some instances but it is far from perfect. Unfortunately, there is no perfect way of estimating the number of clusters and we have had to settle for less than satisfactory results. We recommend that our method is used as a guideline to visual inspection and other techniques based on the nature of the application. For example, we saw in chapter 5.2.4 that in a lot of instances (eg. Figure 5.5d) the detected number of clusters is within ± 1 of the actual number of clusters; this can be advantageous when combined by visual inspection.

6.3.5 Clustering Result Illustrations

As seen in Figure 5.8, the proposed algorithm is able to recover the original clusters under various conditions including high noise levels in some instances. As seen in Figures 5.9a and 5.9b, the algorithm fails to detect the clusters properly in cases with high noise levels. Figures 5.9c and 5.9d demonstrate that sometimes the algorithm can also fail to detect the correct clusters in cases with 0% noise which can be due to incorrect initialization of the number of clusters.

6.4 Discovering Brain-Behaviour Associations

The analytical approach in ASD and ADHD studies has traditionally focused on finding reduced or increased gray matter in different brain regions that would generalize to the population of these psychiatric conditions [94]. However, our study suggests the existence of multiple brain-behaviour correlates that cross the diagnosis boundaries. In this section, we will go over each of the significant brain regions. Our study suggests that in some groups the social communication deficits could be inversely related with a specific cortical regions while in some groups this relation can be direct.

6.4.1 Structural Brain Features

The reported regions are broadly involved in social function, social cognition, language, perception, speech, attention and emotion processing. Many of these regions have been previously reported in ASD (eg. middle frontal gyrus [58, 44, 54, 108], inferior frontal gyrus [20, 33, 105, 59, 112, 111, 33, 47, 90, 89], anterior cingulate cortex [58, 22], insula [65, 37], inferior occipital gyrus [101, 124, 94], heschl gyrus [85, 49, 93], calcarine fissure and and surrounding cortex [13, 97, 49]), in ADHD (eg. cingulate [110, 105, 33, 20], inferior frontal cortex [105], anterior cingulate cortex[88], inferior occipital gyrus [94]) and some have also been frequently reported in social function studies (eg. precuneus [99], anterior cingulate cortex [119, 65, 27, 18, 99, 41], insula [119, 65, 27, 18]) In conclusion, our results suggest the existence of multiple brain-behaviour correlates in these regions and supports the notion that these correlates may not be diagnostic specific.

Chapter 7

Conclusions

7.1 Chapter Overview

In this chapter we will first review the contributions of this study in section 7.2. Then, we highlight the limitations of our work and potential future directions in section 7.3.

7.2 Contributions

The contributions of this study can be categorized into technical and exploratory themes as followed:

- **Bagged Regression Clustering (B-RC)**: We introduced a novel regression clustering pipeline that is able to perform clusterwise linear regression. We demonstrated that it can identify clusters when they intersect (intersecting regression lines). Additionally, B-RC was able to identify the clusters when the data was contaminated with noise (as high as 50% noise in some instances). We also introduced a novel way of estimating the number of clusters based on within-to-between similarity scores of an affinity matrix.
- **Application of Regression Clustering in Studies of Neurodevelopmental Disorders**: We demonstrated for the first time, the use case of clusterwise regression analysis in brain-behaviour studies. Our analysis on social functions

in ASD and ADHD suggests the existence of multiple distinct brain-behaviour correlates where some of the correlates expanded across the traditional diagnostic categories and some were diagnostic specific. Our analysis identified 12 potential regions to be related to social difficulties in ASD and ADHD, and reported the type of relationship (eg. increasing or decreasing) seen in each cluster. Our results should be interpreted with caution given the limitations of the proposed algorithm under high-noise cases.

7.3 Directions for Future Work

In this section, we will review the limitations of our work and elaborate on our thoughts regarding the future direction of this research.

- We only used a single phenotypic measure to quantify behavioural characteristics. Future work could consider other methods for quantifying symptom severity scores.
- The sample size used in this study was limited, with a variability across the population of different diagnostic categories. Future replication studies are needed with larger sample sizes.
- One limitation of the proposed algorithm is considering one response/feature at a time. There is value in analyzing multiple features at the same time (consideration of multiple features in regression). For example, in the case of neurodevelopmental disorders, there could be a network of brain features that could consist of brain-behaviour correlates. Future studies are needed to explore brain-behaviour correlates in networks of regions.
- Although the proposed algorithm could find simulated clusters in data set with outlier points, the performance was inconsistent. Future studies are needed to improve this aspect.
- One of the requirements of the proposed B-RC algorithm is the user defined bag size. Our analysis showed that different bag sizes can have their own advantages and disadvantages. Adopting methods such as the "simulated annealing" can be beneficial and future studies are required to improve this aspect of the algorithm.

Bibliography

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

[2] Daniel Alcalá-López, Jonathan Smallwood, Elizabeth Jefferies, Frank Van Overwalle, Kai Vogeley, Rogier B. Mars, Bruce I. Turetsky, Angela R. Laird, Peter T. Fox, Simon B. Eickhoff, and Danilo Bzdok. Computing the Social Brain Connectome Across Systems and States. *Cerebral Cortex*, 28(June):1–26, 2017.

[3] Andr Altmann, Laura Toloşi, Oliver Sander, and Thomas Lengauer. Permutation importance: A corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347, 2010.

[4] David G. Amaral. The promise and the pitfalls of autism research: An introductory note for new autism researchers. *Brain Research*, 1380:3–9, 2011.

[5] Stephanie H. Ameis, Jason P. Lerch, Margot J. Taylor, Wayne Lee, Joseph D. Viviano, Jon Pipitone, Arash Nazeri, Paul E. Croarkin, Aristotle N. Voineskos, Meng Chuan Lai, Jennifer Crosbie, Jessica Brian, Noam Soreni, Russell Schachar, Peter Szatmari, Paul D. Arnold, and Evdokia Anagnostou. A diffusion tensor imaging studyin children with ADHD, autism spectrum disorder, OCD, and matched controls: Distinct and non-distinct white matter disruption and dimensional brain-behavior relationships. *American Journal of Psychiatry*, 2016.

[6] American Psychiatric Association. *DSM-5*. American Psychiatric Association, 2013.

[7] Evdokia Anagnostou and Margot J. Taylor. Review of neuroimaging in autism spectrum disorders: What have we learned and where we go from here, 2011.

[8] Gideon E. Anholt, Danielle C. Cath, Patricia Van Oppen, Merijn Eikelenboom, Johannes H. Smit, Harold Van Megen, and Anton J.L.M. Van Balkom. Autism and adhd symptoms in patients with ocd: Are they associated with specific oc symptom dimensions or oc symptom severity. *Journal of Autism and Developmental Disorders*, 2010.

[9] Yuta Aoki, Yuliya N. Yoncheva, Bosi Chen, Tanmay Nath, Dillon Sharp, Mariana Lazar, Pablo Velasco, Michael P. Milham, and Adriana Di Martino. Association of White Matter Structure With Autism Spectrum Disorder and Attention-Deficit/Hyperactivity Disorder. *JAMA Psychiatry*, 2017.

[10] Adil M. Bagirov, Julien Ugon, and Hijran Mirzayeva. Nonsmooth nonconvex optimization approach to clusterwise linear regression problems. *European Journal of Operational Research*, 229(1):132–142, 2013.

[11] Adil M. Bagirov, Julien Ugon, and Hijran G. Mirzayeva. Nonsmooth Optimization Algorithm for Solving Clusterwise Linear Regression Problems. *Journal of Optimization Theory and Applications*, 164(3):755–780, 2015.

[12] Jon Baio, Lisa Wiggins, Deborah L. Christensen, Matthew J Maenner, Julie Daniels, Zachary Warren, Margaret Kurzius-Spencer, Walter Zahorodny, Cordelia Robinson, Rosenberg, Tiffany White, Maureen S. Durkin, Pamela Imm, Loizos Nikolaou, Marshalyn Yeargin-Allsopp, Li-Ching Lee, Rebecca Harrington, Maya Lopez, Robert T. Fitzgerald, Amy Hewitt, Sydney Pettygrove, John N. Constantino, Alison Vehorn, Josephine Shenouda, Jennifer Hall-Lande, Kim Van, Naarden, Braun, and Nicole F. Dowling. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR. Surveillance Summaries*, 2018.

[13] Elise B Barbeau, John D Lewis, Julien Doyon, Habib Benali, Thomas A Zeffiro, and Laurent Mottron. A greater involvement of posterior brain areas in interhemispheric transfer in autism: fMRI, DWI and behavioral evidences. *NeuroImage: Clinical*, 8:267–280, 2015.

[14] Danielle A. Baribeau, Krissy A.R. R Doyle-Thomas, Annie Dupuis, Alana Iaboni, Jennifer Crosbie, Holly McGinn, Paul D. Arnold, Jessica Brian, Azadeh

Kushki, Rob Nicolson, Russell J. Schachar, Noam Soreni, Peter Szatmari, and Evdokia Anagnostou. Examining and Comparing Social Perception Abilities Across Childhood-Onset Neurodevelopmental Disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 54(6):479–486, 2015.

[15] Danielle S Bassett and Olaf Sporns. Network neuroscience. *Nature Neuroscience*, 20(3):353–364, 2017.

[16] Yoav; Benjamini and Yosef; Hochberg. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*, 1995.

[17] BIC - The McConnell Brain Imaging Centre. CIVET 2-1-0.

[18] Sarah Jayne Blakemore. The social brain in adolescence. *Nature Reviews Neuroscience*, 9(4):267–277, 2008.

[19] N. Boddaert, N. Chabane, H. Gervais, C. D. Good, M. Bourgeois, M. H. Plumet, C. Barthélémy, M. C. Mouren, E. Artiges, Y. Samson, F. Brunelle, R. S.J. Frackowiak, and M. Zilbovicius. Superior temporal sulcus anatomical abnormalities in childhood autism: A voxel-based morphometry MRI study. *NeuroImage*, 23(1):364–369, 2004.

[20] Bjrn Bonath, Jana Tegelbeckers, Marko Wilke, Hans Henning Flechtner, and Kerstin Krauel. Regional Gray Matter Volume Differences Between Adolescents With ADHD and Typically Developing Controls: Further Evidence for Anterior Cingulate Involvement. *Journal of Attention Disorders*, 22(7):627–638, 2018.

[21] C Bonferroni. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, 8:3–62, 1936.

[22] Leonardo Bonilha, Fernando Cendes, Chris Rorden, Mark Eckert, Paulo Dalgalarrondo, Li Min Li, and Carlos E. Steiner. Gray and white matter imbalance - Typical structural abnormality underlying classic autism? *Brain and Development*, 30(6):396–401, 2008.

[23] E. Bora and C. Pantelis. Meta-analysis of social cognition in attention-deficit/hyperactivity disorder (ADHD): Comparison with healthy controls and autistic spectrum disorder, 2016.

[24] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

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

[26] Sarah Brieber, Susanne Neufang, Nicole Bruning, Inge Kamp-Becker, Helmut Remschmidt, Beate Herpertz-Dahlmann, Gereon R. Fink, and Kerstin Konrad. Structural brain abnormalities in adolescents with autism spectrum disorder and patients with attention deficit/hyperactivity disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 48(12):1251–1258, 2007.

[27] Leslie Brothers. The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci*, 1990.

[28] Pierrick Bruneau, Olivier Parisot, and Benot Otjacques. A heuristic for the automatic parametrization of the spectral clustering algorithm. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 1313–1318. IEEE, 2014.

[29] Christina O. Carlisi, Luke J. Norman, Steve S. Lukito, Joaquim Radua, David Mataix-Cols, and Katya Rubia. Comparative Multimodal Meta-analysis of Structural and Functional Brain Abnormalities in Autism Spectrum Disorder and Obsessive-Compulsive Disorder. *Biological Psychiatry*, 82(2):83–102, 2017.

[30] R Carper. Cerebral Lobes in Autism: Early Hyperplasia and Abnormal Age Effects. *NeuroImage*, 16(4):1038–1051, 2002.

[31] T A Collaboration, A M Price-Whelan, B M Sipőcz, H M Günther, P L Lim, S M Crawford, S Conseil, D L Shupe, M W Craig, and N Dencheva. The Astropy Project: Building an inclusive, open-science project and status of the v2. 0 core package. *arXiv*, 2018.

[32] E. Courchesne, C. M. Karns, H. R. Davis, R. Ziccardi, R. A. Carper, Z. D. Tigue, H. J. Chisum, P. Moses, K. Pierce, C. Lord, A. J. Lincoln, S. Pizzo, L. Schreibman, R. H. Haas, N. A. Akshoomoff, and R. Y. Courchesne. Unusual brain growth patterns in early life in patients with autistic disorder: An MRI study. *Neurology*, 57(2):245–254, 2001.

[33] Ana Cubillo, Rozmin Halari, Anna Smith, Eric Taylor, and Katya Rubia. 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(2):194–215, 2012.

[34] Wayne S. DeSarbo and William L. Cron. A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, 1988.

[35] Roberto Di Mari, Roberto Rocci, and Stefano Antonio Gattone. Clusterwise linear regression modeling with soft scale constraints. *International Journal of Approximate Reasoning*, 91:160–178, 2017.

[36] Chase C. Dougherty, David W. Evans, Scott M. Myers, Gregory J. Moore, and Andrew M. Michael. A Comparison of Structural Brain Imaging Findings in Autism Spectrum Disorder and Attention-Deficit Hyperactivity Disorder. *Neuropsychology Review*, 26(1):25–43, 2016.

[37] Krissy A. R. Doyle-Thomas, Azadeh Kushki, Emma G. Duerden, Margot J. Taylor, Jason P. Lerch, Latha V. Soorya, A. Ting Wang, Jin Fan, and Evdokia Anagnostou. The Effect of Diagnosis, Age, and Symptom Severity on Cortical Surface Area in the Cingulate Cortex and Insula in Autism Spectrum Disorders. *Journal of Child Neurology*, 28(6):732–739, 2013.

[38] Emma G. Duerden, Kathleen M. Mak-Fan, Margot J. Taylor, and S. Wendy Roberts. Regional differences in grey and white matter in children and adults with autism spectrum disorders: An activation likelihood estimate (ALE) meta-analysis. *Autism Research*, 5(1):49–66, 2012.

[39] Christine Ecker, Susan Y Bookheimer, and Declan G M Murphy. Neuroimaging in autism spectrum disorder: brain structure and function across the lifespan. *The Lancet Neurology*, 14(11):1121–1134, 2015.

[40] J Ellegood, E Anagnostou, B A Babineau, J N Crawley, L Lin, M Genestine, E DiCicco-Bloom, J K Y Lai, J A Foster, O Peñagarikano, D H Geschwind, L K Pacey, D R Hampson, C L Laliberté, A A Mills, E Tam, L R Osborne, M Kouser, F Espinosa-Becerra, Z Xuan, C M Powell, A Raznahan, D M Robins, N Nakai, J Nakatani, T Takumi, M C van Eede, T M Kerr, C Muller, R D Blakely, J Veenstra-VanderWeele, R M Henkelman, and J P Lerch. Clustering autism: using neuroanatomical differences in 26 mouse models to gain insight into the heterogeneity. *Molecular Psychiatry*, 20(1):118–125, 2 2014.

[41] Amit Etkin, Tobias Egner, and Raffael Kalisch. Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Sciences*, 15(2):85–93, 2011.

[42] Martin A. Fischler and Robert C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[43] Nicholas E.V. Foster, Krissy A.R. Doyle-Thomas, Ana Tryfon, Tia Ouimet, Evdokia Anagnostou, Alan C. Evans, Lonnie Zwaigenbaum, Jason P. Lerch, John D. Lewis, and Krista L. Hyde. Structural Gray Matter Differences during Childhood Development in Autism Spectrum Disorder: A Multimetric Approach. *Pediatric Neurology*, 53(4):350–359, 2015.

[44] C M Freitag, E Luders, H E Hulst, K L Narr, P M Thompson, A W Toga, C Krick, and C Konrad. Total brain volume and corpus callosum size in medication-naive adolescents and young adults with autism spectrum disorder. *Biological Psychiatry*, 66(4):316–319, 2009.

[45] Tanya E Froehlich, Bruce P Lanphear, Jeffery N Epstein, William J Barbaresi, Slavica K Katusic, and Robert S Kahn. Prevalence, recognition, and treatment of attention-deficit/hyperactivity disorder in a national sample of US children. *Archives of pediatrics & adolescent medicine*, 161(9):857–864, 2007.

[46] Belinda A. Gargaro, Nicole J. Rinehart, John L. Bradshaw, Bruce J. Tonge, and Dianne M. Sheppard. Autism and ADHD: How far have we come in the comorbidity debate?, 2011.

[47] Hilde M. Geurts, K. Richard Ridderinkhof, and H. Steven Scholte. The relationship between grey-matter and ASD and ADHD traits in typical adults. *Journal of Autism and Developmental Disorders*, 2013.

[48] Ian R. Gizer, Courtney Ficks, and Irwin D. Waldman. Candidate gene studies of ADHD: A meta-analytic review, 2009.

[49] Shulamite A Green, Jeffrey D Rudie, Natalie L Colich, Jeffrey J Wood, David Shirinyan, Leanna Hernandez, Nim Tottenham, Mirella Dapretto, and Susan Y Bookheimer. Overreactive brain responses to sensory stimuli in youth with autism spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52(11):1158–1172, 2013.

[50] Nouchine Hadjikhani, Robert M. Joseph, Josh Snyder, and Helen Tager-Flusberg. Anatomical differences in the mirror neuron system and social cognition network in autism. *Cerebral Cortex*, 16(9):1276–1282, 2006.

[51] Madeline B. Harms, Alex Martin, and Gregory L. Wallace. Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, 20(3):290–322, 2010.

[52] Z. Hawi, T. D.R. Cummins, J. Tong, B. Johnson, R. Lau, W. Samarrai, and M. A. Bellgrove. The molecular genetic architecture of attention deficit hyperactivity disorder, 2015.

[53] Heather Cody Hazlett, Michele Poe, Guido Gerig, Rachel Gimpel Smith, James Provenzale, Allison Ross, John Gilmore, and Joseph Piven. Magnetic resonance imaging and head circumference study of brain size in autism: birth through age 2 years. *Archives of general psychiatry*, 62(12):1366–1376, 2005.

[54] Heather Cody Hazlett, Michele D. Poe, Guido Gerig, Rachel Gimpel Smith, and Joseph Piven. Cortical gray and white brain tissue volume in adolescents and adults with autism. *Biological Psychiatry*, 59(1):1–6, 2006.

[55] C Hennig. Models and methods for clusterwise linear regression. In *Classification in the Information Age*, pages 179–187. Springer, 1999.

[56] Christian Hennig. *Datenanalyse mit Modellen für Cluster linearer Regression*. diplom. de, 2000.

[57] Christian Hennig. Fixed point clusters for linear regression: computation and comparison. *Journal of classification*, 19(2):249–276, 2002.

[58] Leanna M Hernandez, Jeffrey D Rudie, Shulamite A Green, Susan Bookheimer, and Mirella Dapretto. Neural Signatures of Autism Spectrum Disorders: Insights into Brain Network Dynamics. *Neuropsychopharmacology*, 40(1):171–189, 2015.

[59] Elseline Hoekzema, Susana Carmona, J. Antoni Ramos-Quiroga, Vanesa Richarte Fernández, Marisol Picado, Rosa Bosch, Juan Carlos Soliva, Mariana Rovira, Yolanda Vives, Antonio Bulbena, Adolf Tobeña, Miguel Casas, and Oscar Vilarroya. Laminar Thickness Alterations in the Fronto-Parietal Cortical Mantle of Patients with Attention-Deficit/Hyperactivity Disorder. *PLoS ONE*, 2012.

[60] Seok-jun Hong, L Valk, Adriana Di Martino, Michael P Milham, and Boris C Bernhardt. Multidimensional Neuroanatomical Subtyping of Autism Spectrum Disorder. *Cerebral Cortex*, 28(September):1–11, 2017.

[61] Lawrence Hubert. Comparing Partitions. *Journal of Classification*, 233:193–218, 1985.

[62] John D Hunter. Matplotlib: A 2D graphics environment. *Computing in science & engineering*, 9(3):90–95, 2007.

[63] Krista L. Hyde, Fabienne Samson, Alan C. Evans, and Laurent Mottron. Neuroanatomical differences in brain areas implicated in perceptual and other core features of autism revealed by cortical thickness analysis and voxel-based morphometry. *Human Brain Mapping*, 31(4):556–566, 2010.

[64] Eric Jones, Travis Oliphant, and Pearu Peterson. SciPy: Open source scientific tools for Python. 2001. *URL http://www. scipy. org*, 2016.

[65] Daniel P. Kennedy and Ralph Adolphs. The social brain in psychiatric and neurological disorders, 2012.

[66] Budhachandra S. Khundrakpam, John D. Lewis, Penelope Kostopoulos, Felix Carbonell, and Alan C. Evans. Cortical Thickness Abnormalities in Autism Spectrum Disorders Through Late Childhood, Adolescence, and Adulthood: A Large-Scale MRI Study. *Cerebral Cortex*, 27(3):1721–1731, 2017.

[67] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian E Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica B Hamrick, Jason Grout, and Sylvain Corlay. Jupyter Notebooks-a publishing format for reproducible computational workflows. In *ELPUB*, pages 87–90, 2016.

[68] William H. Kruskal and W. Allen Wallis. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 1952.

[69] Tarald O. Kvalseth. Cautionary Note about R2. *The American Statistician*, 39(4):279–285, 1985.

[70] Matthew Kyan, Paisarn Muneesawang, Kambiz Jarrah, and Ling Guan. *Unsupervised Learning: A Dynamic Approach*. John Wiley & Sons, 2014.

[71] Joseph Lee Rodgers and W. Alan Nice Wander. Thirteen ways to look at the correlation coefficient. *American Statistician*, 42(1):59–66, 1988.

[72] Alison E. Lenet. Shifting Focus: From Group Patterns to Individual Neurobiological Differences in Attention-Deficit/Hyperactivity Disorder. *Biological Psychiatry*, 82(9):e67–e69, 2017.

[73] Zhao Li, Su hua Chang, Liu yan Zhang, Lei Gao, and Jing Wang. Molecular genetic studies of ADHD and its candidate genes: A review, 2014.

[74] Paul Lichtenstein, Eva Carlström, Maria Råstam, Christopher Gillberg, and Henrik Anckarsäter. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. *American Journal of Psychiatry*, 2010.

[75] L. Lim, K. Chantiluke, A. I. Cubillo, A. B. Smith, A. Simmons, M. A. Mehta, and K. Rubia. Disorder-specific grey matter deficits in attention deficit hyperactivity disorder relative to autism spectrum disorder. *Psychological Medicine*, 45(5):965–976, 2015.

[76] Anath C. Lionel, Jennifer Crosbie, Nicole Barbosa, Tara Goodale, Bhooma Thiruvahindrapuram, Jessica Rickaby, Matthew Gazzellone, Andrew R. Carson, Jennifer L. Howe, Zhuozhi Wang, John Wei, Alexandre F.R. Stewart, Robert Roberts, Ruth McPherson, Andreas Fiebig, Andre Franke, Stefan Schreiber, Lonnie Zwaigenbaum, Bridget A. Fernandez, Wendy Roberts, Paul D. Arnold, Peter Szatmari, Christian R. Marshall, Russell Schachar, and Stephen W. Scherere. Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Science Translational Medicine*, 2011.

[77] Anath C. Lionel, Kristiina Tammimies, Andrea K. Vaags, Jill A. Rosenfeld, Joo Wook Ahn, Daniele Merico, Abdul Noor, Cassandra K. Runke, Vamsee K. Pillalamarri, Melissa T. Carter, Matthew J. Gazzellone, Bhooma Thiruvahindrapuram, Christina Fagerberg, Lone W. Laulund, Giovanna Pellecchia, Sylvia Lamoureux, Charu Deshpande, Jill Clayton-Smith, Ann C. White, Susan Leather, John Trounce, H. Melanie Bedford, Eli Hatchwell, Peggy S. Eis, Ryan K.C. Yuen, Susan Walker, Mohammed Uddin, Michael T. Geraghty, Sarah M. Nikkel, Eva M. Tomiak, Bridget A. Fernandez, Noam Soreni, Jennifer Crosbie, Paul D. Arnold, Russell J. Schachar, Wendy Roberts, Andrew D. Paterson, Joyce So, Peter Szatmari, Christina Chrysler, Marc Woodbury-Smith, R. Brian Lowry, Lonnie Zwaigenbaum, Divya Mandyam, John Wei, Jeffrey R. MacDonald, Jennifer L. Howe, Thomas Nalpathamkalam, Zhuozhi Wang, Daniel Tolson, David S. Cobb, Timothy M. Wilks, Mark J. Sorensen,

Patricia I. Bader, Yu An, Bai Lin Wu, Sebastiano Antonino Musumeci, Corrado Romano, Diana Postorivo, Anna M. Nardone, Matteo Della Monica, Gioacchino Scarano, Leonardo Zoccante, Francesca Novara, Orsetta Zuffardi, Roberto Ciccone, Vincenzo Antona, Massimo Carella, Leopoldo Zelante, Pietro Cavalli, Carlo Poggiani, Ugo Cavallari, Bob Argiropoulos, Judy Chernos, Charlotte Brasch-Andersen, Marsha Speevak, Marco Fichera, Caroline Mackie Ogilvie, Yiping Shen, Jennelle C. Hodge, Michael E. Talkowski, Dimitri J. Stavropoulos, Christian R. Marshall, and Stephen W. Scherer. Disruption of the ASTN2/TRIM32 locus at 9q33.1 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. *Human Molecular Genetics*, 2014.

[78] Haibing Lu and Shengsheng Huang. Clustering Panel Data. *Data Mining for Marketing*, 1, 2011.

[79] Luca Magri and Andrea Fusiello. Multiple Models Fitting as a Set Coverage Problem. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[80] Luca Magri and Andrea Fusiello. Multiple structure recovery with maximum coverage. *Machine Vision and Applications*, 29(1):159–173, 2018.

[81] H. B. Mann and D. R. Whitney. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 1947.

[82] Grinne M. McAlonan, Vinci Cheung, Charlton Cheung, John Suckling, Grace Y. Lam, K. S. Tai, L. Yip, Declan G.M. Murphy, and Siew E. Chua. Mapping the brain in autism. A voxel-based MRI study of volumetric differences and intercorrelations in autism. *Brain*, 128(2):268–276, 2005.

[83] Wes McKinney. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.

[84] Meghan Miller, Russell B. Hanford, Catherine Fassbender, Marshall Duke, and Julie B. Schweitzer. Affect recognition in adults with ADHD. *Journal of Attention Disorders*, 2011.

[85] Elaheh Moradi, Budhachandra Khundrakpam, John D. Lewis, Alan C. Evans, and Jussi Tohka. Predicting symptom severity in autism spectrum disorder based on cortical thickness measures in agglomerative data. *NeuroImage*, 144(September 2016):128–141, 2017.

[86] Benjamin M Neale, Sarah E Medland, Stephan Ripke, Philip Asherson, Barbara Franke, Klaus-Peter Lesch, Stephen V Faraone, Thuy Trang Nguyen, Helmut Schäfer, and Peter Holmans. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(9):884–897, 2010.

[87] Eric E. Nelson, Ellen Leibenluft, Erin B. McClure, and Daniel S. Pine. The social re-orientation of adolescence: A neuroscience perspective on the process and its relation to psychopathology. *Psychological Medicine*, 35(2):163–174, 2005.

[88] Luke J. Norman, Christina Carlisi, Steve Lukito, Heledd Hart, David Mataix-Cols, Joaquim Radua, and Katya Rubia. Structural and functional brain abnormalities in attention-deficit/hyperactivity disorder and obsessive-compulsive disorder: A comparative meta-analysis. *JAMA Psychiatry*, 73(8):815–825, 2016.

[89] Laurence O'Dwyer, Colby Tanner, Eelco V. Van Dongen, Corina U. Greven, Janita Bralten, Marcel P. Zwiers, Barbara Franke, Dirk Heslenfeld, Jaap Oosterlaan, Pieter J. Hoekstra, Catharina A. Hartman, Wouter Groen, Nanda Rommelse, and Jan K. Buitelaar. Decreased Left Caudate Volume Is Associated with Increased Severity of Autistic-Like Symptoms in a Cohort of ADHD Patients and Their Unaffected Siblings. *PLoS ONE*, 11(11):1–20, 2016.

[90] Laurence O'Dwyer, Colby Tanner, Eelco V. Van Dongen, Corina U. Greven, Janita Bralten, Marcel P. Zwiers, Barbara Franke, Jaap Oosterlaan, Dirk Heslenfeld, Pieter Hoekstra, Catharina A. Hartman, Nanda Rommelse, and Jan K. Buitelaar. Brain volumetric correlates of autism spectrum disorder symptoms in attention deficit/hyperactivity disorder. *PLoS ONE*, 9(6):1–13, 2014.

[91] M Ofner, A Coles, M Decou, M T Do, A Bienek, J Snider, and A Ugnat. Autism spectrum disorder among children and youth in Canada 2018: a report of the National Autism Spectrum Disorder Surveillance System. *Public Health Agency of Canada. https://www. canada. ca/content/dam/phac-aspc/documents/services/publications/diseases-conditions/autism-spectrum-*

disorder-children-youth-canada-2018/autism-spectrum-disorder-children-youth-canada-2018. pdf. Accessed, 10, 2018.

[92] Travis E Oliphant. *A guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.

[93] K Oconnor. Auditory processing in autism spectrum disorder: a review. *Neuroscience & Biobehavioral Reviews*, 36(2):836–854, 2012.

[94] Min Tae M. Park, Armin Raznahan, Philip Shaw, Nitin Gogtay, Jason P. Lerch, and M. Mallar Chakravarty. Neuroanatomical phenotypes in mental illness: Identifying convergent and divergent cortical phenotypes across autism, ADHD and schizophrenia. *Journal of Psychiatry and Neuroscience*, 43(3):201–212, 2018.

[95] Fabian Pedregosa, Gal Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, and Vincent Dubourg. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[96] Fernando Pérez and Brian E Granger. IPython: a system for interactive scientific computing. *Computing in Science & Engineering*, 9(3), 2007.

[97] Ruth C M Philip, Maria R Dauvermann, Heather C Whalley, Katie Baynham, Stephen M Lawrie, and Andrew C Stanfield. A systematic review and meta-analysis of the fMRI investigation of autism spectrum disorders. *Neuroscience & Biobehavioral Reviews*, 36(2):901–942, 2012.

[98] Pratibha Reebye. AttentionDeficit Hyperactivity Disorder: A Handbook For Diagnosis And Treatment, Third Edition. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 17(1):31–33, 2 2008.

[99] Crystal Reeck, Daniel R. Ames, and Kevin N. Ochsner. The Social Regulation of Emotion: An Integrative, Cross-Disciplinary Model. *Trends in Cognitive Sciences*, 20(1):47–63, 2016.

[100] Rockville. Social Processes: Workshop Proceedings. Technical report, The National Institute of Mental Health, 2012.

[101] Donald C. Rojas, Eric Peterson, Erin Winterrowd, Martin L. Reite, Sally J. Rogers, and Jason R. Tregellas. Regional gray matter volumetric changes in

autism associated with social and repetitive behavior symptoms. *BMC Psychiatry*, 6:1–13, 2006.

[102] Nanda N.J. Rommelse, Barbara Franke, Hilde M. Geurts, Catharina A. Hartman, and Jan K. Buitelaar. Shared heritability of attention-deficit/hyperactivity disorder and autism spectrum disorder. *European Child and Adolescent Psychiatry*, 19(3):281–295, 2010.

[103] Angelica Ronald, Henrik Larsson, Henrik Anckarsäter, and Paul Lichtenstein. Symptoms of Autism and ADHD: A Swedish Twin Study Examining Their Overlap. *Journal of Abnormal Psychology*, 123(2):440–451, 2014.

[104] Angelica Ronald, Emily Simonoff, Jonna Kuntsi, Philip Asherson, and Robert Plomin. Evidence for overlapping genetic influences on autistic and ADHD behaviours in a community twin sample. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 2008.

[105] Katya Rubia, Analucia Alegria, and Helen Brinson. Imaging the ADHD brain: Disorder-specificity, medication effects and clinical translation. *Expert Review of Neurotherapeutics*, 14(5):519–538, 2014.

[106] Michael Rutter, Anthony Bailey, Catherine Lord, Carlo Cianchetti, and Giuseppina Sannio Fancello. Social Communication Questionnaire. *Western Psychological Services*, 2003.

[107] Laura Ruzzano, Denny Borsboom, and Hilde M. Geurts. Repetitive Behaviors in Autism and ObsessiveCompulsive Disorder: New Perspectives from a Network Analysis. *Journal of Autism and Developmental Disorders*, 2014.

[108] Wataru Sato, Takanori Kochiyama, Shota Uono, Sayaka Yoshimura, Yasutaka Kubota, Reiko Sawada, Morimitsu Sakihama, and Motomi Toichi. Reduced Gray Matter Volume in the Social Brain Network in Adults with Autism Spectrum Disorder. *Frontiers in Human Neuroscience*, 11:1–12, 2017.

[109] Margaret Semrud-Clikeman, Jodene Goldenring Fine, Jesse Bledsoe, and David C. Zhu. Regional Volumetric Differences Based on Structural MRI in Children With Two Subtypes of ADHD and Controls. *Journal of Attention Disorders*, 21(12):1040–1049, 2017.

[110] Philip Shaw, Jason Lerch, Deanna Greenstein, Wendy Sharp, Liv Clasen, Alan Evans, Jay Giedd, F. Xavier Castellanos, and Judith Rapoport. Longitudinal

mapping of cortical thickness and clinical outcome in children and adolescents with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, 63(5):540–549, 2006.

[111] Philip Shaw, Meaghan Malek, Bethany Watson, Deanna Greenstein, Pietro De Rossi, and Wendy Sharp. Trajectories of cerebral cortical development in childhood and adolescence and adult attention-deficit/hyperactivity disorder. *Biological Psychiatry*, 2013.

[112] Philip Shaw, Meaghan Malek, Bethany Watson, Wendy Sharp, Alan Evans, and Deanna Greenstein. Development of cortical surface area and gyrification in attention-deficit/hyperactivity disorder. *Biological Psychiatry*, 2012.

[113] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[114] Jordan W. Smoller, Kenneth Kendler, Nicholas Craddock, Phil Hyoun Lee, Benjamin M. Neale, John N. Nurnberger, Stephan Ripke, Susan Santangelo, Patrick S. Sullivan, Benjamin N. Neale, Shaun Purcell, Richard Anney, Jan Buitelaar, Ayman Fanous, Stephen F. Faraone, Witte Hoogendijk, Klaus Peter Lesch, Douglas L. Levinson, Roy P. Perlis, Marcella Rietschel, Brien Riley, Edmund Sonuga-Barke, Russell Schachar, Thomas S. Schulze, Anita Thapar, Kenneth K. Kendler, Jordan S. Smoller, Michael Neale, Roy Perlis, Patrick Bender, Sven Cichon, Mark D. Daly, John Kelsoe, Thomas Lehner, Douglas Levinson, Mick O'Donovan, Pablo Gejman, Jonathan Sebat, Pamela Sklar, Mark Daly, Bernie Devlin, Patrick Sullivan, and Michael O'Donovan. Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *The Lancet*, 2013.

[115] Elena Sokolova, Anoek M. Oerlemans, Nanda N. Rommelse, Perry Groot, Catharina A. Hartman, Jeffrey C. Glennon, Tom Claassen, Tom Heskes, and Jan K. Buitelaar. A Causal and Mediation Analysis of the Comorbidity Between Attention Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD). *Journal of Autism and Developmental Disorders*, pages 1–10, 2017.

[116] H SpÄath. Cluster dissection and analysis, 1985.

[117] Helmuth Späth. A fast algorithm for clusterwise linear regression. *Computing*, 29(2):175–181, 1982.

[118] H Spiith. Algorithm 39 Clusterwise Linear Regression. *Computing*, 22:367–373, 1979.

[119] Damian A. Stanley and Ralph Adolphs. Toward a neural basis for social behavior. *Neuron*, 80(3):816–826, 2013.

[120] R. Thomas, S. Sanders, J. Doust, E. Beller, and P. Glasziou. Prevalence of Attention-Deficit/Hyperactivity Disorder: A Systematic Review and Meta-analysis. *Pediatrics*, 135(4):e994–e1001, 2015.

[121] Robert Tibshirani, Trevor Hastie, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistics. *Journal of the Royal Statistical society*, 63(2):411–423, 2001.

[122] Andia H. Turner, Kiefer S. Greenspan, and Theo G.M. van Erp. Pallidum and lateral ventricle volume enlargement in autism spectrum disorder. *Psychiatry Research - Neuroimaging*, 252:40–45, 2016.

[123] Jolanda M.J. Van Der Meer, Anoek M. Oerlemans, Daphne J. Van Steijn, Martijn G.A. Lappenschaar, Leo M.J. De Sonneville, Jan K. Buitelaar, and Nanda N.J. Rommelse. Are autism spectrum disorder and attention-deficit/hyperactivity disorder different manifestations of one overarching disorder? Cognitive and symptom evidence from a clinical and population-based sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, 2012.

[124] Daan Van Rooij, Evdokia Anagnostou, Celso Arango, Guillaume Auzias, Marlene Behrmann, Geraldo F. Busatto, Sara Calderoni, Eileen Daly, Christine Deruelle, Adriana Di Martino, Ilan Dinstein, Fabio Luis Souza Duran, Sarah Durston, Christine Ecker, Damien Fair, Jennifer Fedor, Jackie Fitzgerald, Christine M. Freitag, Louise Gallagher, Ilaria Gori, Shlomi Haar, Liesbeth Hoekstra, Neda Jahanshad, Maria Jalbrzikowski, Joost Janssen, Jason Lerch, Beatriz Luna, Mauricio Moller Martinho, Jane McGrath, Filippo Muratori, Clodagh M. Murphy, Declan G.M. Murphy, Kirsten O'Hearn, Bob Oranje, Mara Parellada, Alessandra Retico, Pedro Rosa, Katya Rubia, Devon Shook, Margot Taylor, Paul M. Thompson, Michela Tosetti, Gregory L. Wallace, Fengfeng Zhou, and

Jan K. Buitelaar. Cortical and subcortical brain morphometry differences between patients with autism spectrum disorder and healthy individuals across the lifespan: Results from the ENIGMA ASD working group. *American Journal of Psychiatry*, 175(4):359–369, 2018.

[125] Jeremy Veenstra-VanderWeele and Randy D Blakely. Networking in Autism: Leveraging Genetic, Biomarker and Model System Findings in the Search for New Treatments. *Neuropsychopharmacology*, 37(1):196–212, 2012.

[126] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[127] Ulrike von Luxburg. Clustering Stability: An Overview. *Foundations and Trends® in Machine Learning*, 2(3):235–274, 2010.

[128] Jian Bao Wang, Li Jun Zheng, Qing Jiu Cao, Yu Feng Wang, Li Sun, Yu Feng Zang, and Hang Zhang. Inconsistency in abnormal brain activity across cohorts of adhd-200 in children with attention deficit hyperactivity disorder. *Frontiers in Neuroscience*, 11(JUN):1–10, 2017.

[129] Michael Waskom, O Botvinnik, P Hobson, J Warmenhoven, J B Cole, Y Halchenko, J Vanderplas, S Hoyer, S Villalba, and E Quintero. Seaborn: statistical data visualization. *URL: https://seaborn. pydata. org/(visited on 2017-05-15)*, 2014.

[130] Peter M. Wehmeier, Alexander Schacht, and Russell A. Barkley. Social and Emotional Impairment in Children and Adolescents with ADHD and the Impact on Quality of Life. *Journal of Adolescent Health*, 46(3):209–217, 2010.

[131] Choong-Wan Woo, Luke J Chang, Martin A Lindquist, and Tor D Wager. Building better biomarkers: brain models in translational neuroimaging. *Nature Neuroscience*, 20(3):365–377, 2017.

[132] Fiona Zandt, Margot Prior, and Michael Kyrios. Repetitive behaviour in children with high functioning autism and obsessive compulsive disorder. *Journal of Autism and Developmental Disorders*, 2007.

[133] B. Zhang. Regression clustering. *Third IEEE International Conference on Data Mining*, pages 451–458, 2003.

[134] Alex P Zijdenbos, Reza Forghani, and Alan C Evans. Automatic" pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. *IEEE transactions on medical imaging*, 21(10):1280–1291, 2002.