5-2023

Interrogating autism from a multidimensional perspective: an integrative framework.

Mohamed T. Ali
University of Louisville

INTERROGATING AUTISM FROM A MULTIDIMENSIONAL PERSPECTIVE: AN INTEGRATIVE FRAMEWORK

By
Mohamed T. Ali
MSc. in Systems & Biomedical Engineering, 2017

A Dissertation
Submitted to the Graduate School of the
University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in Interdisciplinary Studies: Specialization in Translational Bioengineering

Interdisciplinary Studies
University of Louisville
Louisville, Kentucky

May 2023

INTERROGATING AUTISM FROM A MULTIDIMENSIONAL PERSPECTIVE:
AN INTEGRATIVE FRAMEWORK

By

Mohamed T. Ali
MSc. in Systems & Biomedical Engineering, 2017

Dissertation approved on

April 3, 2023

by the following Dissertation Committee:

Dissertation Director
Ayman S. El-Baz, Ph.D.

Gregory Neal Barnes, M.D., Ph.D

Tommy Roussel, Ph.D.

Tamer M. A. Mohamed, Ph.D.

Hermann Frieboes, Ph.D.

DEDICATION

To the one who has always been there for me, through thick and thin, always offering her unconditional love and support. Without you, I would not be where I am today, pursuing my dreams and ambitions. You have been my rock, my confidant, and my constant source of inspiration. To my Mom.

ACKNOWLEDGMENTS

My sincere thanks and appreciation are extended to my supervisor, Dr. Ayman El-Baz, for his constant supervision, patience, and valuable advice throughout my graduate program. My sincere thanks go to him for giving me the opportunity to work on this interesting topic as a member of his great research team. I would like also to thank my helpful lab members, especially my post-doctors: Dr. Ahmed Shalaby, Dr. Ali Helmy, Dr. Yaser Elnakieb, Dr. Ahmed Elnakib and my helpful colleague and friend Mohamed Elsharkawy. I am thankful to be in a such lab, with an atmosphere that enabled my work and made it visible.

I would like to acknowledge the other members of my Ph.D. committee for spending time and effort in reading and reviewing my work: Dr. Tommy Roussel, Dr. Hermann Frieboes, and Dr. Tamer M. A. Mohamed, with special thanks to Dr. Gregory Barnes for his frequent help.

ABSTRACT

INTERROGATING AUTISM FROM A MULTIDIMENSIONAL PERSPECTIVE: AN INTEGRATIVE FRAMEWORK

Mohamed T. Ali

April 3, 2023

Autism Spectrum Disorder (ASD) is a condition characterized by social and behavioral impairments, affecting approximately 1 in every 44 children in the United States. Common symptoms include difficulties in communication, interpersonal interactions, and behavior. While symptoms can manifest as early as infancy, obtaining an accurate diagnosis may require multiple visits to a pediatric specialist due to the subjective nature of the assessment, which may yield varying scores from different specialists. Despite growing evidence of the role of differences in brain development and/or environmental and/or genetic factors in autism development, the exact pathology of this disorder has yet to be fully elucidated by scientists. At present, the diagnosis of ASD typically involves a set of gold-standard diagnostic evaluations, such as the Autism Diagnostic Observation Schedule (ADOS), the Autism Diagnostic Interview-Revised (ADI-R), and the more cost-effective Social Responsive Scale (SRS). Administering these diagnostic tests, which involve assessing communication and behavioral patterns, along with obtaining a clinical history, requires the expertise of a team of qualified clinicians. This process is time-consuming, effortful, and involves a degree of subjectivity due to the reliance on clinical judgment. Aside from conventional observational assessments, recent developments in neuroimaging and machine learning offer a fast and objective alternative for diagnosing ASD using brain imaging. This comprehensive work explores the use of different imaging modal-

ities, namely structural MRI (sMRI) and resting-state functional MRI (rs-fMRI), to investigate their potential for autism diagnosis. The proposed study aims to offer a new approach and perspective in comprehending ASD as a multidimensional problem, within a behavioral space that is defined by one of the available ASD diagnostic tools. This dissertation introduces a thorough investigation of the utilization of feature engineering tools to extract distinctive insights from various brain imaging modalities, including the application of novel feature representations. Additionally, the use of a machine learning framework to aid in the precise classification of individuals with autism is also explored in detail. This extensive research, which draws upon large publicly available datasets, sheds light on the influence of various decisions made throughout the pipeline on diagnostic accuracy. Furthermore, it identifies brain regions that may be impacted and contribute to an autism diagnosis. The attainment of high global state-of-the-art cross-validated, and hold-out set accuracy validates the advantages of feature representation and engineering in extracting valuable information, as well as the potential benefits of employing neuroimaging for autism diagnosis. Furthermore, a suggested diagnostic report has been put forth to assist physicians in mapping diagnoses to underlying neuroimaging markers. This approach could enable an earlier, automated, and more objective personalized diagnosis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) characterizes Autism Spectrum Disorders (ASD) as neurodevelopmental disorders that share common impairments in social communication and interactions, as well as restricted and repetitive behavioral patterns [1]. Moreover, ASD is identified as a heterogeneous neurodevelopmental disorder with a strong genetic basis and varied clinical presentations [2]. The severity of symptoms, language and cognitive abilities, and symptom patterns distinguish the three subtypes of ASD, including ASD, Asperger's disorder, and pervasive developmental disorder–not otherwise specified [3].

According to the Centers for Disease Control and Prevention (CDC), a recent statistic reported in 2021 showed that 1 in 44 children in the United States were diagnosed with autism in 2018. This represents a significant increase of 127% from the numbers reported in 2016, 155% from the numbers in 2010, and 241% since 2000. Figure 1 depicts the rising trend of autism diagnosis in the US [4]. The financial burden of autism, including direct medical and non-medical expenses, amounted to 265 billion dollars in 2015, and it is projected to increase to 461 billion by 2025 [4]. Identifying co-occurring conditions at the earliest possible stage could lead to better services and earlier interventions, which could benefit children with ASD. Additionally, ASD imposes significant demands on the medical, social, and political aspects of any nation [5].

To explore various types of abnormalities associated with ASD, several magnetic resonance imaging (MRI) based modalities have been employed, including (i) structural MRI (sMRI) [6] to investigate anatomical features, (ii) functional MRI (fMRI) [7] to examine brain activity either during rest or task performance, and (iii) diffusion tensor imaging (DTI) to study brain connectivity [8, 9].

Autistic symptoms generally emerge within the first two years of life, with early signs detectable in infants as young as 12 months old. Despite this, the average age of ASD diagnosis is approximately five years [10, 11]. The current standard for ASD diagnosis relies on behavior observation tests conducted by clinicians, such as the Autism Diagnostic Observation Schedule (ADOS) [12], Autism Diagnostic Interview-Revised (ADI-R) report [13], or Social Responsiveness Scale (SRS). However, these methods can be time-consuming, require specialized training, and can be subjective in their assessment [10].

1 Current Autism Diagnosis Techniques

The current gold standard for diagnosing ASD involves using a combination of the Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R) [14]. These methods involve interviewing the individual and scoring

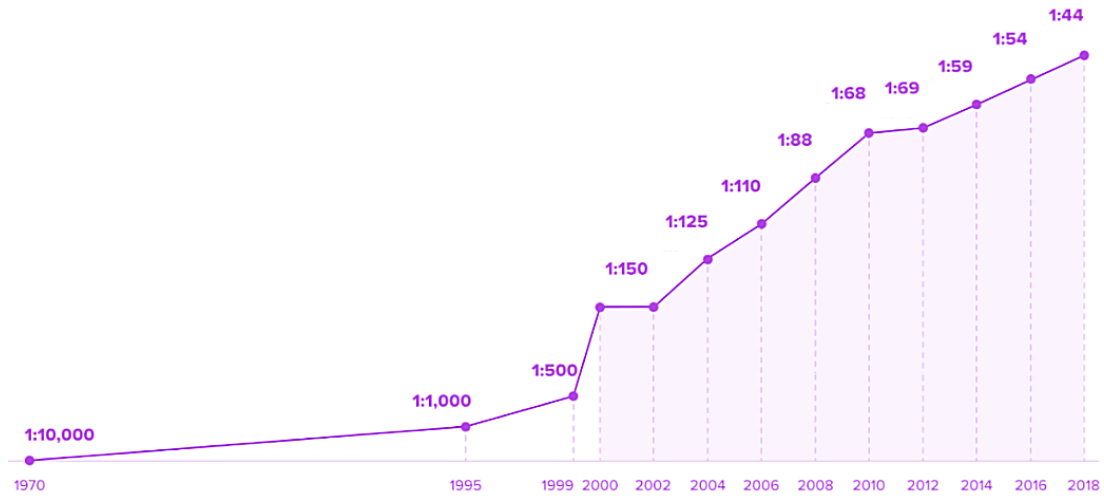


Figure 1. The graph displays a steady increase in the prevalence of autism among children in the United States with respect to each surveillance year.

them based on various behavioral traits related to social reciprocity, communication, and repetitive behavior [14, 15]. While considered the gold standard, the ADOS and ADI-R require extensive training and are time-consuming, and can be subject to bias based on the clinician's experience, patient behavior, and parent knowledge [1, 16–18]. Moreover, given the wide range of symptoms associated with ASD, it can be challenging to identify individuals who may have the disorder [17]. An alternative diagnostic tool is the Social Responsiveness Scale (SRS), which is an objective, efficient, and cost-effective method increasingly being used as a clinical screening without an instrument [19, 20]. In the following subsections, we will briefly discuss each diagnostic module.

ADOS

The Autism Diagnostic Observation Scale (ADOS) is a well-established diagnostic instrument utilized to assess social, communicative, and imaginative deficiencies in individuals with ASD [1]. It is a widely recognized tool that is primarily focused on evaluating social and communication impairments, which are the most prominent indicators of ASD. To cater to different age groups, the ADOS is available in five different modules, each designed for specific age groups and tasks. To be diagnosed with ASD, individuals suspected of having the disorder must undergo standardized assessments of social interaction, play, and imaginative use of materials using the appropriate module. However, the main limitation of this diagnostic tool is that the individual administering and scoring the test must undergo special training to ensure reliability and validity of the results.

ADI-R

The Autism Genetic Research Exchange has defined the Autism Diagnostic Interview-Revised (ADI-R) as a parent interview, which includes ninety-three items that assess the child's behavior across three domains: social interactions, communication and language, and repetitive behaviors. The ADI-R also includes items related to treatment planning, such as self-injury and overactivity. The assessment is mainly dependent on parents' responses and scores are assigned for each of the three domains. A diagnosis of autism is made when a child meets or exceeds the specified cutoffs in all three domains of communication, social interaction, and patterns of behavior, and the disorder has manifested by 36 months of age.

SRS

The Social Receptivity Scale (SRS) is a diagnostic tool used to measure the severity of social deficits and associated symptoms in individuals with ASD based on responses from parents or teachers. The scale was specifically designed to be completed after a quick training session and takes approximately 15-20 minutes to complete [19]. The SRS employs a quantitative approach to measure the full range of autistic symptoms that occur in nature using a 65-item rating scale. The initial rating obtained from the SRS serves as an indication of the severity of the individual's autistic symptoms.

Limitations of Current Techniques

Although no medical test exists for diagnosing autism, evaluators use a combination of observations, feedback from parents, and tests designed to assess the child's skills and behaviors across various developmental domains. However, these methods are subjective, time-consuming, and challenging, and their accuracy is limited to around 80%-85% [21]. Furthermore, clinicians may not always agree with the results due to grading biases [22].

Clinicians and researchers have been using various metrics to measure the severity of autism, reflecting their ongoing efforts to find valid measures for capturing the construct of ASD severity. Currently, the gold standard for diagnosing autism spectrum disorders is behavioral observational testing by a clinician, which is subjective, time-consuming, and cannot detect autism before the child reaches at least two years old [10, 11].

The severity of autism is highly correlated with age, cognitive abilities, and language abilities [23]. As such, many measures reflect the developmental characteristics of children, rather than inherent autistic traits, casting doubt on their validity. Therefore, there is a motivation to develop a quantitative, objective alternative based on neuroimaging that can provide a less subjective evaluation, potentially leading to faster and more reliable diagnoses for clinicians [17, 18].

2 Objectives

The main objectives of this dissertation are:

- to present a promising machine learning (ML) framework for autism diagnosis, which has been tested on publicly available datasets such as ABIDE-I and ABIDE-II, and can be extended to other datasets.
- to develop a unified perspective for solving autism diagnosis using different imaging modalities, namely, structural magnetic resonance imaging (sMRI) and functional magnetic resonance imaging (fMRI).
- to introduce a different perspective to observe Autism Spectrum Disorder (ASD), which considers the disorder's heterogeneity.
- to introduce Neuroatlases that deconstruct ASD into behavioral components.
- to introduce a technique to map alterations in the brain cortex to corresponding behavioral disorders, which is then mapped to the severity of ASD.
- to develop an ML framework that diagnoses ASD in two stages, mimicking the clinical settings of diagnosing ASD.

In the next subsections, we will provide details about the main imaging modalities and tools utilized in this dissertation.

3 Brain Imaging in Autism Diagnosis

MRI is considered to be the most powerful noninvasive clinical diagnostic tool, as it provides the best comparison of soft tissue among all image modalities, making it the most useful modality for imaging the brain [24]. It offers three main modalities: sMRI, DTI, and fMRI. sMRI has been used to study the morphological brain changes in ASD, including the shape and volume of different brain regions. DTI provides information on anatomical connections and has revealed disorganized micro-structural WM integrity in individuals with ASD. On the other hand, fMRI is used to detect dynamic physiological information from active brain regions by measuring the change in Blood Oxygenation Level-Dependent (BOLD) signal in various brain states, such as resting state or task-evoked states, which can reveal functional architecture abnormalities in the ASD population [25, 26].

Structural MRI (sMRI)

sMRI is widely used in non-research medical settings, as it provides exquisite detail of the brain, spinal cord, and vascular anatomy in all three planes [24]. The magnetic resonance imaging (MRI) technique measures the relaxation time of tissue, which can be divided into two categories, T1 and T2. The longitudinal relaxation time (T1) determines how quickly excited protons return to equilibrium, while the transverse relaxation time (T2) is the time constant that determines how quickly excited protons reach equilibrium or fall out of phase with each other. T1-weighted and T2-weighted MRI scans can be obtained by selecting the time to echo (TE) and the repetition time (TR) accordingly. T1-weighted images use a short TE time and TR time,

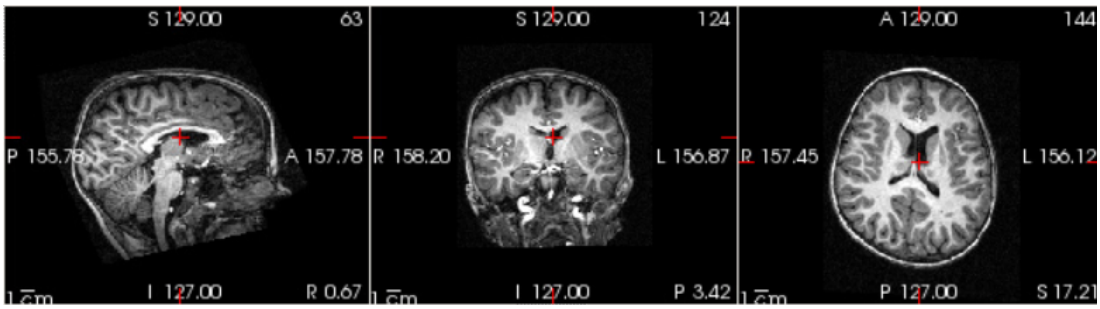


Figure 2. This image depicts a structural magnetic resonance imaging (sMRI) scan of the head, exhibiting intricate anatomical features of the brain as seen in three planes: sagittal, coronal, and axial.

while T2-weighted images are obtained using longer TE and TR. Each type of scan weighs a different characteristic of the tissue (T1 / T2). sMRI is a specific type of MRI that reveals anatomical information of each tissue type, distinguishing gray matter (GM), white matter (WM), cerebrospinal fluid (CSF), and other structures. The output features usually include shape and morphological and geometric features, such as cortical thickness, surface area, volume, and curvature measures, which are relevant for studying changes in brain structure associated with autism [24]. Figure 2 illustrates sMRI and its capabilities.

Functional MRI (fMRI)

Using functional imaging, it is possible to identify areas of the brain and the underlying processes that enable a particular cognitive or behavioral function to be performed. It is possible to make inferences about brain function and location based on the type of signal being analyzed. There is still a long way to go before fMRI is widely used in clinical practice.

fMRI is a technique that enables the study of brain activation patterns in response to certain stimuli, as well as the identification of brain hemodynamic changes that correspond to changes in mental activity [27]. By using functional imaging, researchers can identify brain regions and underlying processes that enable a particular cognitive or behavioral function to be performed. This method allows for inferences to be made about brain function and location based on the type of signal being analyzed. However, despite its potential, fMRI still has a long way to go before it is widely used in clinical practice.

Functional MRI utilizes the Blood Oxygenation Level-Dependent (BOLD) signal, which is based on the fact that the harder a specific area of the brain works while performing a given task, the greater its metabolic demand [28]. As metabolic function increases, vessels in a particular brain region may dilate, resulting in a shift in the gradient between highly oxygenated and highly deoxygenated hemoglobin in the capillaries, leading to a detectable change in the MRI signal. In this way, we can indirectly measure brain activity based on the physiological changes that occur within the brain. This method allows researchers to match various neuronal activities with

BOLD signals and invasively learn how the brain is functioning [29].

Diffusion Tensor Imaging (DTI)

DTI is a non-invasive imaging technique that enables the examination of the structure of white matter tracts in vivo by measuring the diffusion of water molecules within the brain. This provides a macroscopic view of the white matter structure within the voxel being imaged. DTI requires the use of magnetic gradients in at least six non-collinear directions, in addition to a base volume, for proper DTI construction. Using this approach, a 3x3 diffusion tensor matrix can be calculated at each voxel, with each element representing diffusion across one direction, such as xx or xy [30].

$$D = \begin{bmatrix} D_{x,x} & D_{x,y} & D_{x,z} \\ D_{y,x} & D_{y,y} & D_{y,z} \\ D_{z,x} & D_{z,y} & D_{z,z} \end{bmatrix}$$

To find the principal diffusion directions and strength, the 3 eigenvalues λ_1, λ_2 and λ_3 and their corresponding eigenvectors v_1, v_2 and v_3 are calculated, where the eigenvector corresponding to the largest eigenvalue is the principle diffusion direction (i.e, diffusion in across the fiber), while the other two eigenvectors correspond to the radial diffusion directions (i.e, diffusion perpendicular to the fiber) [31]. A special case is an isotropic medium, where the diffusion ellipsoid takes the shape of a sphere because $\lambda_1 = \lambda_2 = \lambda_3$. While in the case of an anisotropic medium, the diffusion is represented as an ellipsoid as shown pointing in the v_1 direction with λ_1. Figure 3 illustrates the resulting representation. In addition, another representation is DTI tractography, where major tracts or bundles from a region of interest (ROI) can be tracked, and 3D color-coded maps can be visualized, with each color representing a different direction (x / y / z) of fiber crossing, see Figure 4.

DTI is a non-invasive technique that uses the diffusion of water molecules within white matter (WM) tracts to visualize WM structures in vivo. This is accomplished by applying magnetic gradients in at least 6 non-colinear directions, in addition to a base volume, to construct a 3x3 diffusion tensor matrix at each voxel, with each element representing diffusion across one direction. The most commonly used DTI metrics are Mean Diffusivity (MD) and Fractional Anisotropy (FA). MD is related to cellular density, while FA captures the directional changes of diffusion and represents the degree of alignment of WM tracts and cellular structure, ranging from 0 (random or isotropic) to 1 (unidirectional or anisotropic). Other parameters include Axial Diffusivity (AD) and Radial Diffusivity (RD), which respectively measure the diffusion in a direction parallel and perpendicular to WM tracts, and are related to axon and myelin integrity [32–34].

WM tracts are composed of axons that transmit signals between different brain regions. Dysmaturation of myelination in individuals with autism spectrum disorder (ASD) has been associated with alterations in synaptogenesis and changes in axonal fiber density, caliber, and homogeneity. These changes lead to impairments in the microstructural organization and integrity of WM [26, 35, 36].

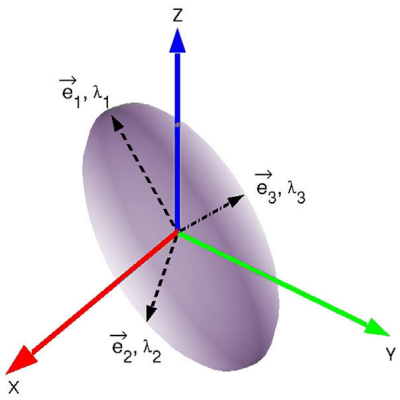


Figure 3. The ellipsoid model is utilized to visually represent the diffusion tensor in this graphic, which is determined by three eigenvectors that specify the orientation of the ellipsoid in three dimensions and three eigenvalues that define the primary axes values of the ellipsoid.

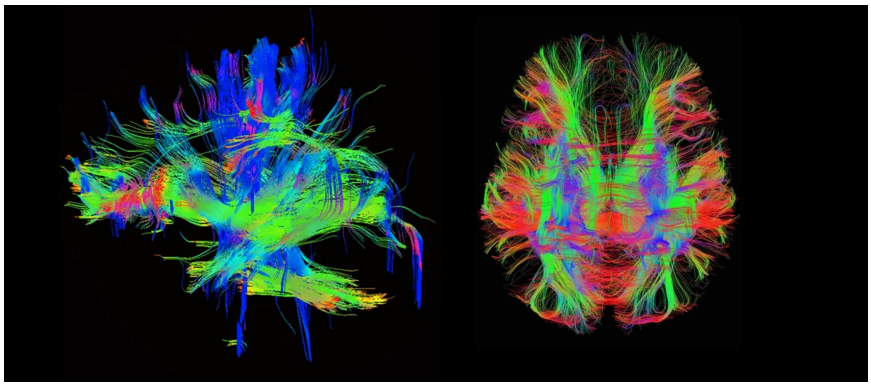


Figure 4. This image showcases tractography of the white matter tracts using diffusion tensor imaging (DTI).

4 Computer Aided Diagnosis (CAD) using Machine Learning (ML)

Obtaining an early and accurate diagnosis is of utmost importance in mitigating the worsening of a condition and enhancing prognostic outcomes [37–39]. Computer-aided algorithms have been developed to aid clinicians in achieving early and precise diagnoses. A computer-aided diagnosis (CAD) generally involves three primary stages: feature extraction, feature reduction, and classification. Feature extraction is contingent on the type of medical data employed, with distinct feature extraction approaches and data representations for each data modality. After appropriate data representation and feature extraction, feature reduction and machine learning remain areas that could benefit from further enhancements and algorithms. The following section will introduce the fundamental concepts and types of both feature reduction and machine learning stages.

Dimensionality Reduction: importance and techniques

Effective reduction of high-dimensional feature space is a critical aspect of medical computer-aided diagnosis (CAD) systems that rely on machine learning algorithms. The "curse of dimensionality," as described by Bellman [40], refers to the exponential increase in complexity resulting from the addition of new dimensions to the feature space, which is common in MRI imaging research and medical data where the feature space is extremely large, creating potential for over-fitting. To address this issue, several techniques for feature reduction have been developed, including principal component analysis (PCA), linear discriminant analysis (LDA), and autoencoders [41–43]. However, these methods do not always retain the semantic meaning of the original feature space, making it challenging to identify clinical findings underlying classification results. As a result, feature selection methods that preserve the original feature space's semantics have gained attention for more practical use by physicians to better understand pathological abnormalities in autistic brains. Therefore, efficient feature selection techniques, which typically incorporate filtering techniques, are introduced in the following sections [41–43].

$$F = \frac{between\ group\ variance}{within\ group\ variance}$$

Kernel F-score is a feature selection technique that involves using one-way ANOVA F-test to compute a scoring function for each feature in the input X. The computed scores are then used to select a subset of the features that are deemed most significant for classification purposes. Specifically, the first K features with the highest scores in X are retained. The significance of a large score is attributed to the lack of equality in the means of the K groups, thereby providing 'distinctive' information. The F-score is calculated as the ratio of "between group variance" to "within group variance". Medical datasets with high dimensional feature spaces have been shown to benefit from using F-score values as a feature selection method [44].

The **Recursive feature elimination (RFE)** is a technique used for feature reduction that ranks features based on their importance and removes the least important ones in an iterative manner. The RFE method is designed to improve generalization performance by removing less important features that have the least effect on training errors. This method employs a classifier as a kernel, which assigns weights to features or determines their importance, and recursively eliminates a small number of features at each step. The performance of the RFE depends on the estimator used for training at each step. Many kernels are used in literature, including the commonly used supported vector machine (SVM).

The RFE-CV is another variant of the RFE method that uses cross-validation to determine the performance of the estimator at each step. This variant helps to avoid over-fitting and leads to better generalization performance, albeit with more training time due to the need for k-fold cross-validation computations. Several studies in the literature have shown the effectiveness of the RFE method for feature selection in medical datasets, with SVM-based RFE being one of the most commonly used variants [45, 46].

Machine learning (ML) techniques

ML involves the use of artificial learning mechanisms to replicate human thinking and problem-solving processes. As an essential component of AI, it employs a large volume of data points to enable machines to predict, categorize, and identify patterns. The three primary types of ML are supervised, unsupervised, and reinforcement learning. Supervised learning relies on labeled input-output pairs to classify data, where each input corresponds to a ground truth output. A case in point is the supervised classification of diseases, where labeled subsets are used to train the model and labeled testing subsets for evaluation. Unsupervised learning, in contrast, categorizes data without the aid of labeled input-output pairs. Instead, it employs the patterns inherent in the input data to efficiently categorize them. Reinforcement learning employs positive/negative reward signals instead of labeled input-output pairs to teach machines. This type of ML is the most similar to human learning as it involves learning by reinforcement.

Supervised learning, which is the central focus of this study, includes classification as an example. A broad range of classifiers exists, such as SVM, RF, LR, ANN, and others. In addition to traditional machine learning methods, deep learning has gained considerable attention and demonstrated significant potential in many medical applications. Among the most widely used deep learning architectures are CNNs, which incorporate convolutional and fully connected layers to perform both feature extraction and classification. CNNs have demonstrated superiority over traditional algorithms in image recognition tasks. However, due to their complexity, CNNs are computationally intensive and often challenging to interpret

In this dissertation, a range of machine learning algorithms were employed. Below, we provide an overview of some of the classifiers used in this work:

- K-Nearest Neighbor (KNN):

  The K-nearest neighbors (KNN) classifier is a machine learning algorithm that assigns data points to one of K groups based on their proximity to the nearest data points, as measured by a specified distance metric. The choice of the parameter 'K' and the distance metric are critical and need to be determined before training. KNN is highly sensitive to the quality of the input data, and it is essential to preprocess and normalize the data to achieve optimal performance

- Naïve Bayes:
  Naive Bayes is a probabilistic classification algorithm that utilizes Bayes theorem to determine the posterior likelihood and classify unstructured data based on it. This approach assumes that the predictors are independent, which is not always true in practice. Despite this simplification, Naive Bayes has shown high accuracy and is particularly useful for text classification and spam filtering.

- Logistic Regression (LR):
  LR is a popular classification algorithm that has been widely used for many decades. Unlike traditional regression models, it uses a logistic function to

model the relationship between input variables and the probability of belonging to a specific class. LR is highly regarded in the medical field due to its interpretability and ability to identify the contribution of individual predictors to the classification outcome.

- Random Forest (RF):
  Random forest is a popular ensemble learning method based on decision trees that is often used for classification and regression tasks. It works by constructing multiple decision trees on bootstrap samples of the data and then combining their predictions through majority voting or averaging. Each tree in the forest is built using a random subset of features, which helps to reduce variance and prevent overfitting. Random forest has been shown to be a powerful classifier in various applications, including medical diagnosis.

- Support Vector Machines (SVM):
  The SVM is a popular classifier that is often used in high-dimensional spaces, especially when the number of dimensions is greater than the number of samples. Essentially, an SVM algorithm attempts to locate a hyperplane in the feature space of N dimensions, which separates the data points in a clear manner. The performance of the SVM depends significantly on the choice of the SVM kernel, which is a crucial function of the SVM. The kernel function maps the original data set into a higher-dimensional space, where the algorithm constructs a hyperplane to divide the data set into classes based on its spatial distribution

- Passive Aggressive classifier:
  The algorithm is named based on its two distinct mechanisms: 1) Passive: the model remains unchanged if the prediction is accurate, meaning that the example data is insufficient to alter the model. 2) Aggressive: the model is modified if the prediction is inaccurate, allowing for adjustments to improve the model.

- XGBoost:
  The framework of extreme gradient boosting (XGBoost) is aimed at regularizing gradient boosting and is an open-source library. By using gradient boosting, which is an ensemble learning approach that combines the outputs from several decision trees, XGBoost calculates a prediction score.

- Artificial Neural Networks (ANN):
  ANNs are modeled after the human brain and are well-suited for processing large and complex datasets. The basic units of ANNs are nodes, which function similarly to neurons, and they are connected to multiple inputs to generate the target output through multiple layers. Each node has the ability to approximate any continuous function on its direct input. Initially, the weights for the nodes are randomly assigned and then iteratively adjusted to optimize the output for all data points. ANNs have demonstrated great success in a variety of applications, such as computer vision, natural language processing, and speech recognition.

Performance Measures

Various metrics are utilized to evaluate the performance of ML components for autism diagnosis. In the subsequent sections, we will provide a concise summary of these metrics.

The performance of ML components is commonly evaluated using various metrics to assess the accuracy of autism diagnosis. In order to define these metrics, we use the following notations: TP represents true positive, TN represents true negative, FN represents false negative, and FP represents false positive. The performance metrics are described below:

- Specificity: $\frac{TN}{FP+TN}$
- Sensitivity (recall): $\frac{TP}{TP+FN}$
- Accuracy: $\frac{TP+TN}{TP+TN+FP+FN}$
- F1-score: $\frac{2*Precision*Recall}{Precision+Recall} = \frac{2*TP}{2*TP+FP+FN}$
- Balanced Accuracy Score: $\frac{1}{2}(\frac{TP}{Pos} + \frac{TN}{Neg})$
- Precision: $\frac{TP}{TP+FP}$
- The area under the curve (AUC) of the receiver operating characteristics (ROC) is a commonly used metric to evaluate the performance of binary classification models. The AUC is a value between 0 and 1, where a higher value indicates better performance. An AUC of 1 indicates perfect classification, while an AUC of 0.5 indicates that the model is not performing better than random chance.

5 Dissertation Organization

In this thesis, six sections are comprised, and the following statements provide an overview of the contents in each section.:

Chapter II: The present chapter provides an overview of recent publications that have utilized statistical analysis and machine learning techniques to predict or identify Autism Spectrum Disorder (ASD) based on MRI data. The chapter critically reviews recent advancements towards understanding ASD using both structural MRI (sMRI) and functional MRI (fMRI) imaging modalities. The literature review presented in this chapter emphasizes the unique contribution of this dissertation to the existing body of knowledge.

Chapter III: In this chapter, a solution is presented for the understanding and diagnosis of ASD. The proposed solution consists of a framework that includes preprocessing steps, feature extraction, feature selection, and ML training and optimization. The main objective of this framework is to identify the brain regions that have the strongest correlation with ASD and to utilize those regions for diagnosis. To evaluate the proposed framework, it was applied to the ABIDE I dataset, which is a publicly available dataset comprising 664 subjects after quality control.

Chapter IV: In this chapter, a second and optimized solution for ASD diagnosis is introduced, taking into account the heterogeneity of the disorder. Similar to the first solution, this chapter proposes an ML framework that includes a preprocessing step, feature extraction, feature selection, and training and optimization steps. However, it adds a two-step classification process that considers the behavioral components of the disorder. The first step divides the disorder into its behavioral components, and the second step aggregates the classifications of each component to produce a final diagnosis for each subject. The proposed framework is applied to ABIDE II, which consists of 1038 subjects after quality control.

CHAPTER II
AUTISM COMPUTER AIDED DIAGNOSIS: A SURVEY

Magnetic-Resonance Imaging (MRI) has made in-vivo brain studies possible. The first MRI study of autism was published in the late 1980s by Gaffney et al. [47] and Courchesne et al. [48], and since then, hundreds of studies have been conducted. Structural MRI (sMRI) examination is widely used to investigate brain morphology due to its high contrast sensitivity and spatial resolution, as well as its non-invasive nature, which is especially important for children and adolescents [49]. Different MRI modalities have been utilized to investigate the effect of ASD on the brain, including sMRI, fMRI, and/or DTI. sMRI has been used by studies that focus on the geometry of the cerebral cortex and brain morphology [50–52]. Two major types of structural imaging studies exist depending on the features used: geometric features, such as surface area, circumference, curvature, and thickness, which are 2D-surface features related to the brain cortex [53], and volumetric features, which refer to the volume of subcortical structures such as the hippocampus, putamen, thalamus, etc. [54]. On the other hand, functional MRI (fMRI) has been used by studies investigating alterations in brain activation between ASD and typically developed (TD) groups [7]. Functional imaging studies can be divided into two broad types: task-based fMRI and resting-state fMRI. Task-based fMRI studies involve the study of functional activities and cognitive behaviors of the brain based on the induced stimulus by tasks [55]. On the other hand, DTI, which is the most recent submodality used in ASD studies, is concerned with the analysis of the structural connectivity of the brain white matter (WM) [56]. It characterizes the diffusion of water molecules in biological tissues and examines normative white matter (WM) development, as well as neurodevelopmental and neurodegenerative disorders [57, 58]. While early theories about autism were structurally based, sMRI remains widely used in current studies to investigate brain morphology due to its high contrast sensitivity and spatial resolution, as well as its safety for children and adolescents [49]. This is particularly relevant since the earliest theories about autism proposed that individuals with autism might have a larger brain volume than their typically developing peers, known as the big brain theory [59–61]. Another theory rooted in neuropathology, the minicolumnar pathology in autism, also has implications for large-scale anatomy [62]. Therefore, in this study, we focused on sMRI as the primary imaging modality for investigating the effects of ASD on brain morphology.

Following the pioneering work of Gaffney et al. [47] and Courchesne et al. [48], subsequent studies on autism and the brain have honed in on specific structures such as the cerebellum [63–65], amygdala [66–68], hippocampus [66, 69, 70], and corpus callosum [31, 71, 72] [73]. Structural MRI (sMRI) has enabled researchers to examine structural changes in the brains of individuals with ASD, utilizing analytical methods

such as voxel-based morphometry (VBM) and surface-based morphometry (SBM) [50]. VBM-based studies focus on tissue density and volume, while SBM studies concentrate on the intrinsic topology of the cerebral cortex, which cannot be measured directly with VBM [50, 51].

The proposed study aims to investigate cortical morphology differences between individuals with ASD and typically developed individuals, and therefore is classified as an SBM study. Prior research suggests that multiple aspects of cerebral morphology differ between ASD brains and TD brains, including cortical thickness, surface area, and folding pattern. These features are associated with dendritic arborization, the number of minicolumns, and intrinsic and extrinsic connectivity. Therefore, examining the relationships between these features can provide insight into the multifactorial etiologies of ASD. Previous SBM studies have utilized cortical morphology to identify statistical differences between TD and ASD brains, and some have used machine learning algorithms to develop predictive models of ASD. [2, 17, 52, 74–85].

The Autism Brain Imaging Data Exchange (ABIDE) has provided researchers with access to large datasets, which has resulted in an increase in publications that combine machine learning (ML) with various neuroimaging biomarkers. The objective of these studies is to reduce subjectivity and establish a more objective data-driven approach to identify, classify, and predict the prognosis of children with ASD [11].

The purpose of this chapter is to provide a comprehensive review of recent publications that use machine learning (ML) methods to predict or identify autism spectrum disorder (ASD) based on MRI, with a particular focus on functional imaging (fMRI) studies published in the last five years. Electronic databases, including PubMed and Google Scholar, were manually searched until July 2022 using specific search terms. The eligibility criteria included original research articles, published or accepted for publication, and available online in English. The studies compared a group of ASD individuals with a group of typically developed controls, and age- or sex-based studies were included. Studies that compared ASD with other neuro-developmental, cognitive, or psychiatric disorders such as Attention-Deficit Hyperactive Disorder (ADHD) were excluded. The review focused on sMRI and fMRI studies that used ML algorithms with neuroimaging data as a biomarker for differentiating ASD individuals from typically developed controls. Other imaging modalities such as structural MRI, MR spectroscopy, or positron emission tomography were excluded from the data extraction process. Narrative, systematic reviews, meta-analyses, and case reports were excluded from data extraction but used as reference searches.

1 Structural MRI (sMRI)

In 1943, Kanner's seminal publication on ASD described the condition as characterized by cognitive and behavioral difficulties without any consistent cerebral dysmorphology, which has intrigued the scientific community ever since [86]. Understanding abnormal brain structures or dysfunctions in individuals with ASD is crucial for the development of effective diagnostic and treatment strategies. However, the behavioral approach used to diagnose ASD makes it difficult to identify neurological biomarkers.

Despite numerous studies, the etiology and pathogenesis of ASD remain unclear, and it is recognized as a complex multifactorial disorder [87–90]. Therefore, there are currently no specific drugs available to treat the core symptoms of ASD.

Statistical analysis

Levitt et al. [52] used a 3D mapping of cortical sulcal patterns to analyze brain structure differences between 21 ASD and 20 TD individuals aged between 7-13 years old. The authors found significant differences in the position of the superior frontal sulci, sylvian fissure, superior temporal sulcus, and inferior frontal sulcus in the autistic group. They concluded that these findings suggest delayed maturation in brain regions involved in memory, emotion processing, language, and eye gaze. Nordahl et al. [74] utilized surface-based morphometry (SBM) to analyze cortical shapes in ASD individuals aged between 7.5 to 18 years old. The authors found shape abnormalities in different regions for low-function ASD, high-function ASD, and Asperger's syndrome groups, which indicates altered early brain development trajectory in autism. Nunes et al. [75] conducted a longitudinal study on TD and ASD populations within the age range of 6-30 years old using ABIDE I and ABIDE II data sets. The authors reported age-related changes in cortical thickness (CT) primarily in frontal and temp-parietal areas, which differed between ASD and TD groups. The linear slope of CT (curvature) was identified as the most reliable feature for localizing atypical brain development in ASD. Khundrakpam et al. [2] analyzed the ABIDE I dataset to determine age-specific differences in cortical thickness and their relationship with ASD symptoms severity. The authors found increased cortical thickness in ASD, primarily left-lateralized, from six years old onwards, with differences diminishing during adulthood. These findings suggest a dynamic nature of morphological abnormalities in ASD.

Recent research suggests that there are significant differences in the developmental trajectory of the cortex in individuals with ASD compared to those without ASD. Levitt et al. [52] used 3-dimensional (3D) mapping to study cortical sulcal patterns in 21 ASD and 20 TD individuals aged between 7 to 13 years old. They found significant differences in the anterior and superior shifting of the superior frontal sulci bilaterally, anterior shifting of the right sylvian fissure, the superior temporal sulcus, and the left inferior frontal sulcus in the autistic group relative to the normal group. The authors suggested that these findings indicate delayed maturation in brain regions involved in diverse functions, consistent with delayed myelination patterns seen on MRI in ASD. Nordahl et al. [74] utilized shape-based modeling to examine cortical abnormalities in a range of ASD individuals aged between 7.5 and 18 years old. They divided the ASD group into three subgroups based on function and found that all subgroups showed shape abnormalities in specific regions of the cortex. The authors concluded that these findings are consistent with the evidence of altered developmental trajectories in multiple brain regions and identified several regions that may have abnormal patterns of connectivity in individuals with ASD.

According to recent research, the developmental trajectory of the cortex in individuals with ASD differs significantly from that of typically developing (TD) individ-

uals, with altered brain development localized in specific regions. Nunes et al. [75] conducted a longitudinal study utilizing the ABIDE I and ABIDE II data sets and found that there were no overall differences in cortical thickness (CT) between the two groups across the entire age range of 6-30 years old. However, they did report differences in age-related changes in CT located mainly in frontal and temporal-parietal areas, with the curvature being the most reliable feature for localizing atypically developed brain areas in ASD. Similarly, Khundrakpam et al. [2] studied the inconsistencies in cortical abnormalities in ASD using the ABIDE I dataset and included 560 subjects out of 1100 available subjects. They reported increased cortical thickness in ASD, primarily left lateralized, from six years onwards, with differences decreasing during adulthood. These findings highlight the dynamic nature of morphological abnormalities in ASD and emphasize the importance of studying brain development across ages to understand the altered regional trajectories compared to TD individuals.

ML studies

Moradi et al. [91] aimed to predict the severity of ASD symptoms solely based on cortical thickness, utilizing support vector regression (SVR) and Elastic Net penalized linear regression. The authors used data from 156 individuals with ASD, between 8 to 40 years old, compiled from four sites in the Autism Brain Imaging Data Exchange 1 (ABIDE-1). They reported an average correlation of 0.51 and an average mean absolute error of 1.36. Dekhil et al. [92] utilized features from both sMRI and fMRI modalities, including morphological features from sMRI such as surface area, volume, thickness, curvature, and folding index, and Person correlation coefficients between time courses of different brain regions from fMRI. They used 185 subjects obtained from National Database for Autism Research (NDAR) and reported 75% classification accuracy using fMRI data only, 79% using sMRI data only, and 81% when fusing both features together. Yassin et al. [85] performed multi-class classification among TD, ASD, and schizophrenia subjects, and binary classification between each pair of classes. They used CT, subcortical structures' volumes, and surface area as features for classification, and achieved accuracy of 69% for multi-class classification, 75% for ASD vs. schizophrenia classification, 75.8% for ASD vs. TD classification, and 70.6% for schizophrenia vs. TD classification. Ali et. al. [93] proposed a feature selection algorithm to select the most relevant morphological features to ASD, utilizing ABIDE I dataset and achieving accuracy of 82% using neural networks and 72% using support vector machines. Gao et. al. [94] proposed a combination of convolutional neural network (CNN) and individual structural covariance network to classify ASD, utilizing subjects from ABIDE I dataset and achieving accuracy of 71.8

CHAPTER III
DIAGNOSING AUTISM SPECTRUM DISORDER AS A HOMOGENEOUS DISORDER: STRUCTURAL MRI APPROACH

The aim of this chapter is to present a comprehensive ML model to detect imaging markers for autism, and then utilize these imaging markers to train a set of linear and non-linear classifiers to distinguish between ASD and TD. The main motivations behind using solely morphological features which are extracted from the brain cor-tex, while neglecting the subcortical structures, are (i) segmentation of subcortical structures is more challenging and prone to error more than the cortex segmentation, and (ii) most of the significant findings in the literature are achieved by utilizing the SBM methods. The proposed model defines a global neuro-atlas annotating all the brain regions associated with ASD among all subjects of the data set, as well as, a local neuro-atlas for each site independently. Neuro-atlases are built via employing sophisticated a machine learning algorithm utilizing different classifiers, recursive feature elimination with cross-validation (RFECV) using four different classifiers (SVM with linear kernel (LSVM), RF, LR with l_1-norm (Lg1), LR with l_2-norm (Lg2)). Eventually, a training step via three linear and five non-linear classifiers. Furthermore, the proposed model utilizes features extracted only from the brain cortex such as curvature, CT, surface area, and volume. In this chapter, we try to avoid the limitations existed in the aforementioned literature such as working on a small sample size [79, 85], or neglecting the heterogeneity of the ASD [7, 84]. Furthermore, as it was previously mentioned that age is a significant confounding variable on ASD, we adjusted our data to account for the effect of age on the extracted brain features. Consequently, in this chapter, we are utilizing the ABIDE I data set [95], which comprises 1112 subjects collected from different sites/hospitals in the US. In this chapter, we are answering the question of whether a site-based classification, i.e., local model, would improve the classification accuracy over a one ML model for the whole set. Moreover, we study if there would be any common features between the selected features from each site, and the selected feature from the global model. In order to answer the aforementioned questions, two implementations were carried out for the proposed model: (i) on each site of the ABIDE I dataset separately to find local the local neuro-atlas of each site, and (ii) on the whole dataset to find the global global neuro-atlas for the whole autism spectrum represented by the available subjects. The main contribution of the proposed work can be summarized as follows: (i) building a comprehensive ML pipeline to find morphological features and brain regions that are correlated with autism, (ii) finding the anomalous neuro-circuits caused by autism (e.g. neuro-atlases) , and (iii) investigating a global ML model that can be used to diagnose ASD subjects with different demographics and scanning parameters.

1 Materials and Methods

A comprehensive ML pipeline is proposed in this chapter to select morphological features and brain regions that relates to ASD. The ML pipeline starts with downloading the sMRI volumes of ASD and TD subjects provided by ABIDE I dataset [96], then the preprocessing of the sMRI volumes is performed by Freesurfer V.6.0 [97–100]. Preprocessing consists of three stages which are: (i) intensity normalization, (ii) skull stripping, and (iii) brain segmentation. Each of the aforementioned stages comprises a set of substages which are going to be briefly discussed in the following sections. After preprocessing, features are extracted in the form of two numerical representation for each morphological feature for each brain region. A data matrix, and a target vector are created and passed to feature selection algorithm to select the candidate imaging markers. Reduced data matrix based on the candidate imaging markers, and the target vector are then passed to the ML algorithms to select the best ML model that can be used for classifying ASD and TD subjects. The whole pipeline is automated with Python 3.7 [101]. We utilized *pandas* as the data manipulation package [102], *numpy* and *scipy* for numerical analysis and matrices operations [103], *scipy* for performing statistical tests [104], *nibabel* for reading and writing Freesurfer files [105]

Figure 5 demonstrates the general block diagram of the proposed model for each of the global model, and the local model. For the global model, the proposed block diagram is applied only one time over the whole dataset. On the other hand, for the local model, the proposed block diagram is applied on each independent site. Results of both the global model, and each site of the local models are analyzed and compared to each other. Each site's results using the local model answers the research question about findings of local imaging markers. Similarly, the global model answers the research question about finding global imaging markers for the all the subjects included in the dataset. In the following sections, each of the main blocks in both models is discussed in detail.

Dataset

ABIDE I is a famous publicly available dataset. Using ABIDE I achieves two-fold advantages: (i) It facilitates replicating the results, since it is publicly available. (ii) It comprises a large sample size, which adds more significance to the findings. ABIDE I contains sMRI and resting-state fMRI data acquired on individuals with ASD and TD individuals from 17 independent sites. ABIDE I includes 1112 subjects divided into 530 subjects with ASD, and 573 subjects with TD. The original studies included in ABIDE received approval from each participating site's Institutional Review Board (IRB). All sites diagnosed autism using the Autism Diagnostic Interview-Revised (ADI-R), or Autism Diagnostic Observation Schedule (ADOS). Moreover, each site provided basic phenotypic data on each subject, including age, sex, and intelligence Quotient (IQ). For more details about ABIDE, refer to [95].

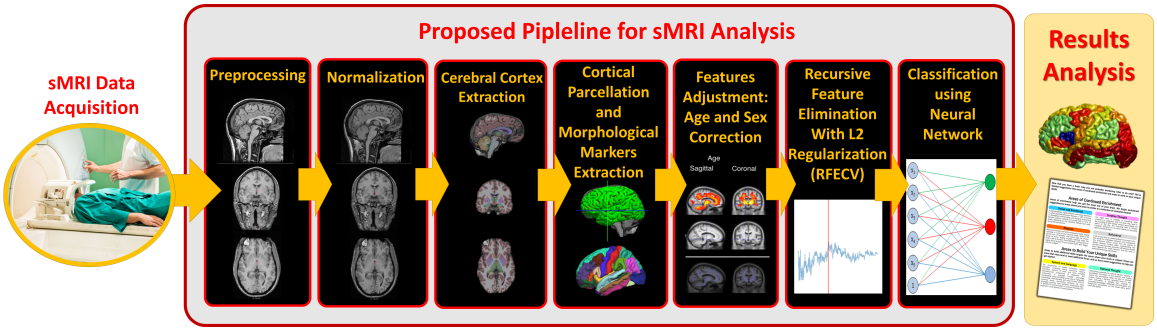


Figure 5. Overview of the proposed system starting from acquiring MRI volumes up to the diagnosis.

Pre-processing

Preprocessing is a crucial requirement to eliminate the between-subjects variability that may stem from data acquisition, different scanners, artifacts, or partial volume effects. Moreover, brain MRI scans usually contain non-brain tissues as it is shown in Figure 5. FreeSurfer performs multiple steps on each sMRI volume to extract the morphological features. Those steps are intensity normalization, brain extraction and skull stripping, brain segmentation and area labeling, tessellation of the gray-white matter boundary, surface inflation and spherical atlas registration, and eventually cortical surface parcellation to the Desikan-Killiany (DK) atlas. It is worth noting that the main assumption behind the preprocessing is that as long as FreeSurfer succeeds in extracting the morphological features of the cerebral cortex and parcelate them to DK, then confounding variables relevant to the MRI scanner wont be major concern. This assumption is based on the fact that FreeSurfer outputs the morphological features in their physical unit e.g. mm, mm^2, mm^3.

Intensity normalization

Variations in both intensity and contrast across sMRI images, resulting in the corruption of the sMRI images, are typically due to magnetic susceptibility artifacts and RF-field inhomogeneities. This corruption is undesirable for any segmentation procedure which utilizes intensity information in order to classify voxel data into different tissue types [106].

To correct the aforementioned corruption, the following procedure of 11 steps is repeated with iterating oversteps from (viii)-(x). The procedures are: (i) Construct a set of histograms from overlapping slices parallel to the x-y Cartesian plane in the magnetic co-ordinate system, (ii) Smooth the resulting histograms using a fairly broad Gaussian window, (iii) Use a peak-finding algorithm to determine the mean white matter intensity, (iv) Discard the outliers from the array of the detected mean white matter intensities, (v) fit a set of cubic splines to the resulting coefficients of the valid slices, (vi) Use the splines to interpolate the coefficients for each point along the z axis, (vii) Adjust each intensity value by the coefficient at its z coordinate, (viii) Find all points in the volume that are at the center of a $5 \times 5 \times 5$ neighborhood of intensity

values that all lie within 10% of the white matter peak, (ix) Build a Voronoi diagram and set all voxels unassigned in step (viii) to the correction value of the nearest control point, (x) Perform a few iterations of "soap-bubble" smoothing, and (xi) Scale the intensity at each voxel in the volume by the computed correction field [107]. The result is shown in Figure 5, the visual transformation of the brightness level between the preprocessing step and the normalization step is to reduce the variance of brightness for the same tissue inter-subjects due to different data acquisition methods. For more mathematical and implementation details, the reader is referred to [107, 108]

Brain Extraction or Skull Stripping

Brain extraction of skull stripping is the process of automatically strip the skull (or any non-brain tissue) from the intensity normalized image. In order to remove the skull and any non-brain tissue, a tessellated ellipsoidal template is deformed into the shape of the inner surface in the skull. Two kind of forces drive the deformation process: (i) An MRI-based force, and (ii) A curvature reducing force.

The MRI-based force is designed to drive the template outward from the brain. It is calculated based on nonlocal information obtained by sampling the MRI data along the surface normal to each vertex of the template tessellation. The curvature reducing force enforces a smoothness constraint on the deformed template, which can be seen as encoding *a priori* knowledge about the smoothness of the inner surface of the skull [107]. The result of this step is illustrated in Figure 5.

Brain Segmentation & Area labeling

The segmentation process is a two-step procedure: (i) A preliminary classification is performed based solely on the intensity information, and (ii) This volume is examined and the regions that contain more than one tissue type are marked for further processing [107]. After segmentation, a 3D surface reconstruction and brain parcellation to an anatomical atlas is performed on the segmented volume. The 3D surface reconstruction is performed via 2 steps: (i) tessellation of the gray-white matter boundary as described in [109, 110], and (ii) surface inflation and spherical atlas registration as described in [108, 111]. Brain parcellation to an anatomical atlas, which is Desikan-Killiany (DK) atlas, is described in [112]. DK atlas parcellates the brain into 68 cortical labels, 34 for each hemisphere. The results of the segmentation and the DK atlas parcellation are shown in Figure 5. For more detailed information on each of the aforementioned preprocessing steps, the reader is referred to the following publication [96].

Feature extraction

There are two outputs of FreeSurfer which are (i) a set of volumes for each subject describing each step of the pipeline (normalization, skull stripping and segmentation) as shown in Figure 5, and (ii) surfaces parcellated to DK atlas and containing the morphological features values at each point on a predefined mesh grid created on the brain, as shown in Figure 20.

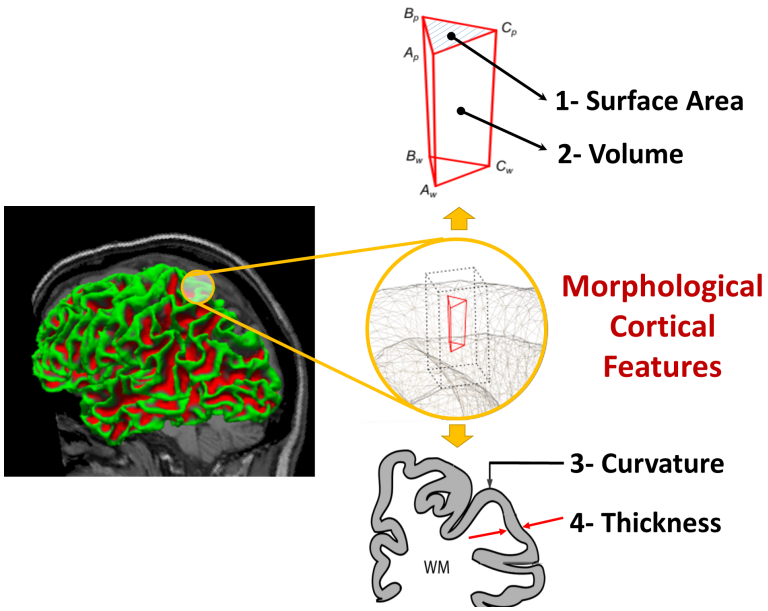


Figure 6. Morphological features extracted from brain surfaces by freesurfer.

in this chapter, we utilized the following morphological features to represent the brain of each subject: (i) surface area (S_a), (ii) volume (V), (iii) thickness (Th), and (iv) curvature (c) (see Figure 20). It is worth noting that Th is calculated as the closest distance from the gray/white matter boundary to the gray/CSF boundary at each vertex on the tessellated surface [113], while c is measured as the average of the reciprocal of the principal radii [114]. For each of those features, we calculated the median value (MV), inter-quartile range (IQR), and MV ± IQR within each brain region parcellated to the DK atlas. There are two reasons behind choosing MV − IQR and MV + IQR to represent each morphological feature of each brain region: (i) the distribution of morphological features' values within each brain region is not necessary Gaussian as it is shown in Figure 7, and (ii) to include lower and upper bound that each morphological feature can possess within a specific brain region while excluding the outliers. DK atlas parcellates the brain into 68 brain regions; 34 brain regions on the left hemisphere, and 34 brain regions on the right hemisphere. Therefore, each subject is represented by a vector of 68 brain regions ×4 features ×2 = 544 elements of a feature vector.

ABIDE I comprises 17 different sites with total number of 1112 subjects after performing quality control, and removing all subjects with bad brain segmentation, at which the data were collected. Thus, the data are heterogeneous, and it is invalid to assume blocking for all confounding variables while working on the whole dataset. Therefore, we proposed two exclusion criteria: (i) exclusion criterion for subjects, and (ii) exclusion criterion for sites. The exclusion criterion for subjects is simply done by removing the subjects with missing feature values. The exclusion criterion for sites depends on how balanced each site is. In other words, after applying subjects' excluding criterion, if we find a site where the ratio ASD:TD, or its reciprocal TD:ASD, exceeds 0.6, we discard that site.There is a trade-off between removing all

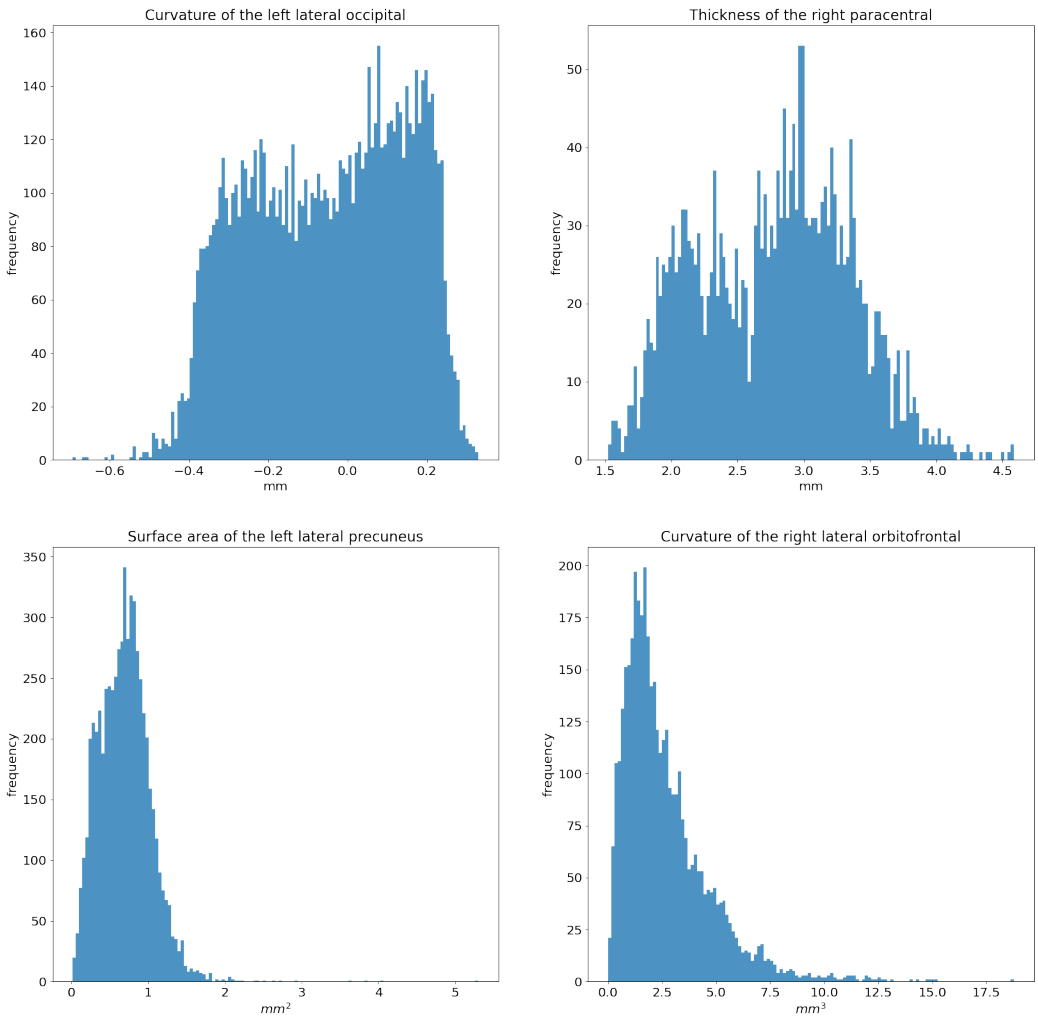


Figure 7. Distribution of the S_a values within different brain regions.

the subjects of the unbalanced sites, or raise the unbalanced threshold. Empirically, we found that 0.6 would be a reasonable ratio, given that we utilize balanced accuracy score to evaluate the system performance, to assume balance and include as many sites as possible in the study The rationale behind the site's exclusion criterion is to avoid including too many subjects of one class that have been collected with certain criterion without having their corresponding subjects from the other class that possess the same collection criterion, i.e., trying to avoid introducing more heterogeneity due to the subject's exclusion criterion.

Table 1 shows the summary statistics of the data set after applying both exclusion

Site	total	ASD		TD	
		n	age (min, max)	n	age (min, max)
Caltech	37	19	(17.5, 55.4)	18	(17, 56.2)
CMU	27	14	(19, 39)	13	(20, 40)
Leuven	63	29	(12.1, 32)	34	(12.2, 29)
MaxMun	51	23	(7, 58)	28	(7, 46)
OHSU	26	12	(8, 15.2)	14	(8.2, 11.9)
Olin	34	19	(11, 24)	15	(10, 23)
Pitt	56	29	(9.33, 35.2)	27	(9.4, 33.2)
Stanford	38	19	(7.5, 12.9)	19	(7.7, 12.4)
Trinity	47	22	(12, 23)	25	(12, 25.6)
UCLA	95	53	(8.36, 17.94)	42	(9.2, 17.7)
UM	134	61	(8.5, 18.6)	73	(8.2, 28.8)
Yale	56	28	(7, 17.7)	28	(7.6, 17.8)
Total	664	328		336	

Table 1. ABIDE data phenotypical information summary after sites' preprocessing

criteria, subject's exclusion criteria and site exclusion criteria. Total number of five sites have been discarded which are: KKI, SDSU, NYU, SBL, and USM, representing a total of 305 subjects. Over the whole data set, there is no statistically significant difference between ASD and TD group ($chi = -0.0271, p = 0.869$). Furthermore, there is no statistically significant difference between the age of each group ($t = -0.5438, p = 0.5867$). However, there is a statistically significant difference between the gender within each group; for TD group, chi-square test was conducted over the gender distribution ($\chi^2 = 84.188, p < 0.001$), and for ASD group, chi-square test was conducted over the gender distribution ($\chi^2 = 94.010, p < 0.001$). At the end of this step, a data matrix is created as follows:

$$D = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,544} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,544} \\ \vdots & \vdots & \ddots & \vdots \\ f_{664,1} & f_{664,2} & \cdots & f_{664,544} \end{bmatrix}, \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{664} \end{bmatrix}$$

where D is the data matrix with size 664 subjects ×544 features; each row represents the feature vector of a specific subject. $f_{i,j}$ denoted the feature value j of subject i, and y_i denoted the diagnosis of subject i. It is worth mentioning that D is the data matrix for the global model.

For the local model, we created 12 data matrices (D_L) such that $D_L = \{d_t : d_t \in^{M,544} \ \& \ 1 \leq t \leq 12\}$ each corresponding to one of the sites; d_t denotes the data matrix corresponding to site t. Each d_t has the size of $M \times 544$ such that M denotes the number of subjects within site t; sequentially, y_t denotes the diagnosis vector corresponding to site t, and y_L denotes the set containing all the y_t for all sites.

$$d_t = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,544} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,544} \\ \vdots & \vdots & \ddots & \vdots \\ f_{M,1} & f_{M,2} & \cdots & f_{M,544} \end{bmatrix}, \qquad y_t = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_M \end{bmatrix}, \qquad D_L = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_{12} \end{bmatrix}, \qquad y_L = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{12} \end{bmatrix}$$

Feature Adjustment & Normalization

Since it has been mentioned in the literature that there is an effect of age on ASD brain morphology [115], morphological features have been adjusted, for the effect of both age and sex, in the proposed work. Adjusted metrics of regional V and S_a were calculated using cortical growth curves from Coupé et al. [116]. Denote by $V_s(a)$ the mean volume of cortical grey matter in individuals of sex s and age a. Then each regional volume V_r is replaced by its age-relative, adjusted metric $V'_r = V_r/V_s(a)$. Similarly, each regional surface area S_r is converted to an adjusted metric $S'_r = S_r/V_s(a)^{2/3}$ [116].

The feature vector corresponding to every subject contains the MV − IQR and MV + IQR of each morphological feature for every region. We consider MV − IQR and MV + IQR to be the lower bound, and the upper bound of every morphological feature for every brain region respectively. Morphological features don't share the same units of measurement; for instance, surface area is measured in mm^2, while V is measured in mm^3. Consequently, we anticipate having different ranges of values, which might adversely affect the performance of the classifiers [117].

in this chapter, we utilized minimum-maximum normalization between 0 to 1 as it is one of the most common normalization methods used for biomedical data [118]. Consequently, each column in the data matrix D is normalized between 0 to 1 using the equation 4.

$$\tilde{f}_{i,j} = \frac{f_{i,j} - \min_i\{f_{i,j}\}}{\max_i\{f_{i,j}\} - \min_i\{f_{i,j}\}} \tag{1}$$

where $\tilde{f}_{i,j}$ and $f_{i,j}$ denote the normalized feature value j and the original feature value j corresponding to the subject i, and $\min(f_j)$ and $\max(f_j)$ correspond to minimum and maximum values of the feature vector j respectively. The output normalized matrix is denoted by D_n for the global model, and d_{tn} for the local model where $1 \leq t \leq 12$.

Building Neuro-Atlas

To implement a Computer-Aided Diagnosis System (CAD) for accurate diagnosis of autism, we have to use a neuro-atlas tailored to the specific developmental patterns of the brain in autism. Unfortunately, there is no general purpose brain atlas in the literature that we can use in our CAD system; thus, developing an atlas for autistic subjects that shows the areas and imaging markers that are associated with autism is the main motivation behind this work. To achieve this goal, we used the modern tools of machine learning (e.g., Recursive Feature Elimination via Cross Validation

(RFECV)) to select the most significant features and their corresponding areas that are correlated with autism spectrum disorder. Since, RFECV is one of supervised feature selection algorithm, we have to split the data into $k - folds$, $k = 10$ in our case (as shown on Figure 8, and for each fold (i) train a predetermined classifier using the training set, (ii) evaluate the performance of the trained classifier on the validation set, (iii) save the classifier's score on the validation set, (iv) find the least significant feature according to the trained classifier, (v) remove the least significant feature from the model, (vi) repeat the whole processes until you end up with only one feature. Again, repeat the whole process for each fold, calculate the average performance of the $k - fold$ cross-validation (CV) when: using all features to train the classifier; using all features but one, and so forth, to the point of classification on a single feature. Find the number of the features at which the classifier has the maximum performance score, assuming it is N_f features. N_f is the optimum number of features to be selected. Perform the whole algorithm again over all the subjects to find the most N_f significant features. For further details regarding the algorithm and its implementation, the reader is suggested to read Guyon et al. [119] and Pedregosa et al. [120], respectively.

To build a neuro-atlas for autism, ABIDE I dataset and RFECV are utilized to select those significant brain regions along with their morphological features. For both the local model and the global model, RFECV is run with four different classifier architectures, which are RF,least absolute shrinkage and selection operator (LASSO), RIDGE regression (RIDGE), and SVM with linear kernel, resulting in four different models. Those four models represent two major categories of features' sets: (i) A feature set that forms a feature space, where the subjects are non-linearly separable as much as possible, and (ii) A feature set that forms a feature space, where the subjects are linearly separable as much as possible. The first category corresponds to the features' set selected by RFECV+RF, and the second category corresponds to the features' sets selected by RFECV+LASSO, RFECV+RIDGE, RFECV+SVM. Each of the RFECV models is performed with 10-fold CV; such that we iterate over all the 544 features, removing one feature at a time, perform 10-fold CV on the current sample, and calculate the average balanced accuracy score. The balanced accuracy score was introduced in 2010 to solve the optimistic estimate occurs when a biased classifier is tested on an imbalanced dataset [121]. The balanced accuracy score is defined by equation 6:

$$score = \frac{1}{2}\left(\frac{TP}{Pos} + \frac{TN}{Neg}\right) \tag{2}$$

Where $score$ denotes the balanced accuracy score, TP denotes the true positive classified by the model, Pos denotes the total number of positive cases in the sample, TN denotes the true negatives classified by the model, and Neg is the total number of negative cases in the sample.

For the site-based model, RFECV is performed on each site separately. The selected set of features is then extracted for each site to have a new data matrix with number of columns less than or equal to the original number of columns. For the global model, RFECV is performed only one time on the normalized data matrix

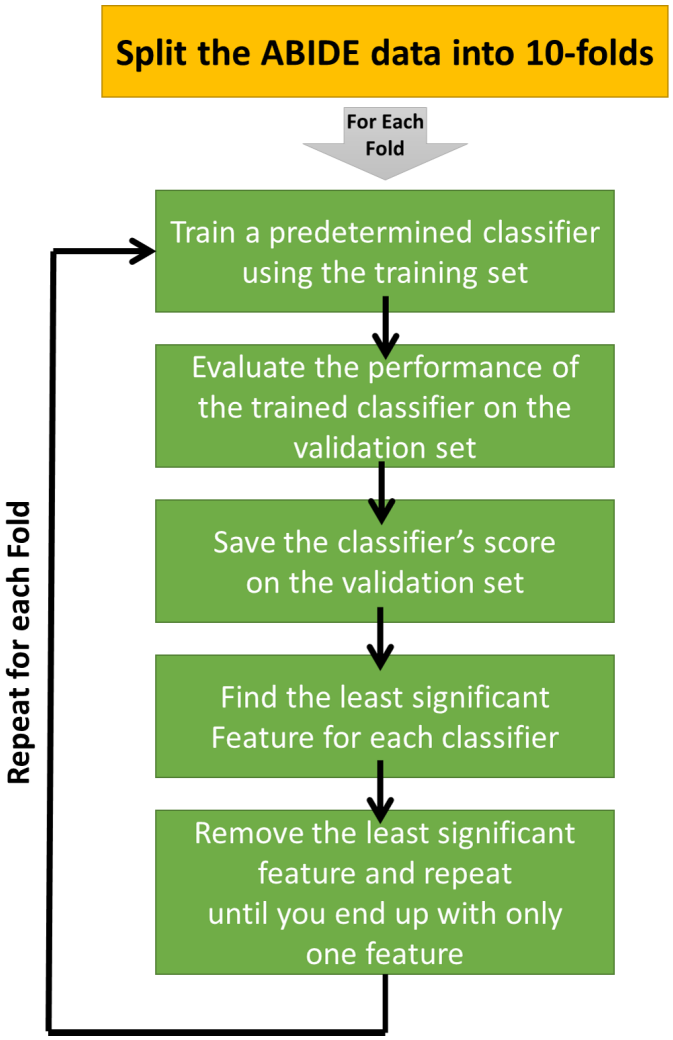


Figure 8. The flowchart of The RFECV algorithm.

D_n, and the selected set of features is calculated and then extracted from D_n. It is worth noting that at this point we assume that the selected features from each site are the imaging markers candidate for ASD subjects collected from that site i.e., local imaging markers; while the selected features from D_n, in the global model, are the global markers candidates that define the ASD subjects in the whole dataset. In case of the global model, the input to the RFECV step is D_n and the output is S_n where the size of S_n is $664 \times M$ such that $M <= 554$. In case of the site-based model, the input to the RFECV step is 12 d_n (normalized data matrices of each site), and the output is s_{kn} where the size of s_{kn} is $N \times M$ such that N is the number of subjects within site k, and $M <= 544$.

Eventually, a global neuro-atlas is created using the whole data set, and a local neuro-atlas is created for each site. We claim that the globla neuro-atlas, as well as, the local neuro-atlases can be used as a guide for future analysis of ASD or ABIDE I dataset.

ML classifiers

Having the imaging markers candidates, they should be placed under test to see how good they are at separating the two classes. A set of eight different ML classifiers representing both linear and non-linear hypotheses is selected to test the local, and the global imaging markers candidates. The utilized eight ML classifiers are split into two main divisions: (i) Linear classifiers, (ii) Non-linear classifiers. The linear classifiers set comprises LR, LSVM, and passive aggressive. The non-linear classifiers set comprises RF, SVM with radial basis function (SVM-RBF), eXtra Gradient Boost trees (xgboost), Gaussian Naive Bayes (GNB), and neural network (NN) shallow and deep. To optimize the hyper-parameters of each classifier, Data matrix is reduced accordingly to the results of RFECV algorithm. The data is split into 5-folds. For each of the predefined classifiers, the hyperparameters and their ranges, where the search will be conducted, are defined in 7, then a nested for-loop for each classifier and for each hyperparameter value of that classifier, a 5-fold CV is performed, and the results of each fold is saved. Eventually, the hyperparameters that corresponds to the maximum CV average score is saved as the optimum parameters. The results of the highest performed classifiers are saved with their hyperparameter values. The performance metric for each classifier is set to the balanced accuracy score.

Personalized diagnosis

We propose a personalized map per subject to show the affected brain regions and to gage the probability of a diagnostic difference when comparing autistic individuals to control. We define a personalized map as a set of scores associated with a set of features denoting the importance of a particular feature in diagnosing a subject as TD or ASD. In a previous work [7], we created a personalized map for ABIDE I dataset using both fMRI and sMRI features. However, in the proposed work, we are introducing a personalized map for ABIDE I dataset using only sMRI morphological features. The motivation behind creating the personalized map for only the sMRI morphological features is the high performance of the proposed pipeline.

A personalized map is created for each local model. The personalized maps are easily created with classifiers that either associate weights to the input features such as linear classifiers, or place the input features in a tree schematic that denotes the importance of each feature based on the level of the feature. However, since NN is the used classifier for the local models, it is difficult to determine which input feature contributes the most to the classification decision and which input feature contributes the least.

Local interpretable model-agnostic explanations (LIME) [122] is a novel explanation technique that explains the prediction of any classifier. The main idea behind LIME is that it builds linear models around the predictions of an opaque model to explain it. LIME is used to explain the classification decisions made by four different local models We have used LIME to explain the decisions of four local models on two random subjects within each of the four sites. Afterward, we visualize the scores representing the contribution of each feature in the classification decision as shown

in Figure 10.

2 Experimental Results

In this section, the results for each site are demonstrated in terms of (i) autism imaging local markers (AILM) for each site, and (ii) the balanced accuracy score of the trained ML when using the AILM. The balanced accuracy scores of the trained ML models are meant to validate the AILM selected via one of the proposed feature selection models (RFECV+lg1, RFECV+lg2, RFECV+svm, RFECV+rf). In order to measure the significance of our findings, we compare the accuracy of the proposed pipeline in classifying ASD to the results of Katuwal et. al [123] study. Katuwal et. al. utilized ABIDE I dataset to provide a site-based diagnosis using RF, SVM, and Gradient Boosting Machine (GBM) classifiers. In addition to the results of the local model, we applied the proposed on the whole dataset to find, what we have called, autism imaging global markers (AIGM). Similar to local model, AIGM are put to test to study how discrimintive they are at classifying ASD. However, for the global model we do not only compare the results with another study, but also, we compare the results with the proposed model without performing feature selection i.e. using 544 features. Therefore, for the global model the two comparisons performed are (i) compare the performance, utilizing balanced accuracy score, of the proposed model to the performance of the same model while excluding the feature selection step, and (ii) comparing to the results of Sabuncu et. al. [124]. Sabuncu et. al. performed univariate and multivariate methods to classify ASD subjects within ABIDE I dataset. The authors performed 5 fold cross validation and used SVM, Neighborhood approximation forest (NAF), which is a variant of random forest, and relevance voxel machine (RVM) classifiers. Sabuncu et. al. has used different set of features as well, however, the highest classification accuracy was achieved using the cortical thickness values sampled onto the fsaverage5 template, provided by Freesurfer, and smoothed on the surface with an approximate Gaussian kernel of a full-width-half-max (FWHM) of 5 mm.

Autism Local Neuro-Atlases: RFECV with the four classifiers is applied on each independent site, and the set of features corresponding to the maximum balanced accuracy score is selected. To compress the data visualization into four subplots similar to the case of the global model, we visualize only the optimum number of features selected by each RFECV model along with the highest achieved balanced accuracy score. Figure 9 demonstrates the number of selected features corresponds to the maximum balanced accuracy score achieved when utilizing each RFECV model on every site of the dataset.

ML local models: almost perfect cross-validation results are achieved when using the local-atlases selected by RFECV+lg2, as it showed the highest performance in case of the global model, as well as in case of sites-based model, the highest number of sites with balanced accuracy score above 0.65 is achieved with RFECV+lg2 algorithm. Results are compared with Katuwal et. al. [123] in table 2. The proposed pipeline has achieved better accuracy as shown in Table 2. Although Katuwal et.

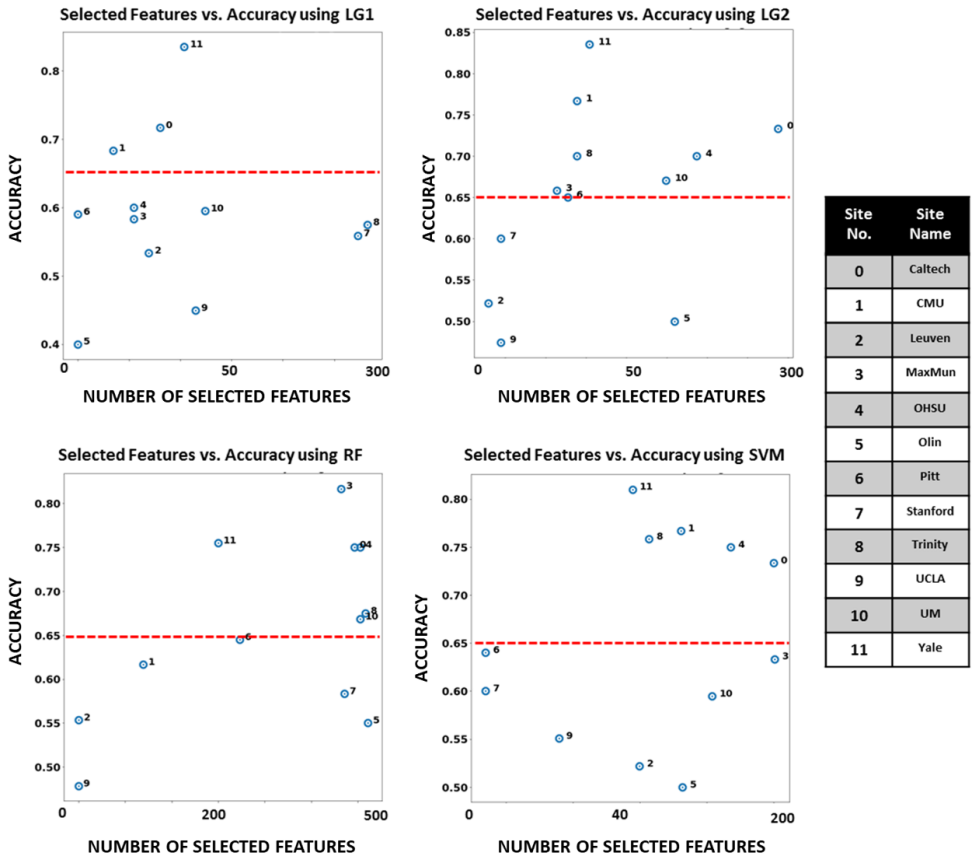


Figure 9. Number of selected features vs. the maximum balanced accuracy score achieved using these features when applying RFECV using the four core classifiers, using the local model.

al. selected, for some sites, a smaller number of features, the proposed pipeline has outperformed their method in terms of accuracy.

Personalized diagnosis: Figure 10 visualizes the results of applying LIME on eight different subjects from four different sites. brain regions with gray color represents brain regions that are not selected as candidates of the local neuro-atlas for that site. Brain regions with deep blue color contribute heavily toward classifying the subject to be ASD and it gets milder as the color moves from blue towards green. On the other hand, brain regions with deep red color contribute heavily toward classifying the subject to be TD and it gets milder as the color moves from the red color to the green color. All the subjects shown in Figure 10 are correctly classified by their corresponding local models. Figure 10 proposes that the local models classify subjects based on the majority voting of the classifications per brain region. However, we cannot assume that using each brain region solely for classification and aggregating their decisions would yield the same results. We assume that each of the local neur-atlas creates a multidimensional space at which the highest possible accuracy of classification of ASD is achieved, and it is hard to achieve the same accuracy by utilizing each feature independently from the others.

Table 2. Comparison between the proposed pipeline and previous results from literature.

Site	Katuwal et. al [123] results		Proposed pipeline	
	Number of selected features	Accuracy (%)	# of selected features	Accuracy (%)
Caltech	5	97	217	100
CMU	1	94	18	100
Leuven	-	-	20	91.5 ± 5
MaxMun	-	-	151	97.5 ± 1
OHSU	12	77	3	100
Olin	1	86	60	100
Pitt	-	-	16	100
Stanford	-	-	7	100
Trinity	-	-	18	100
UCLA	2	64	55	82.2 ± 5
UM	3	72	59	97.2 ± 1
Yale	2	75	21	100

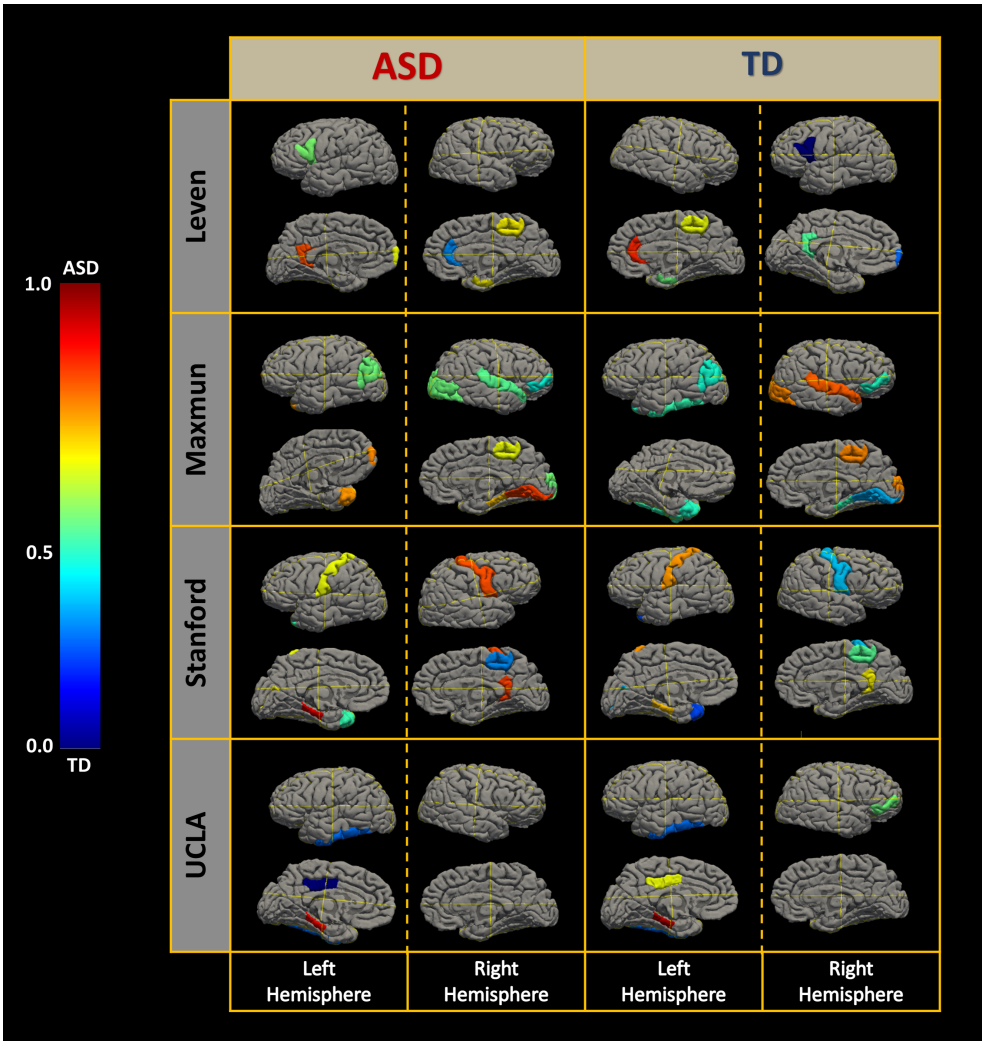


Figure 10. Personalized Diagnosis.

Autism Global Neuro-Atlas: The high accuracy obtained using the local

model has encourage us to try to find a global model. To achieve this goal, we performed the same feature selection criteria on the whole data set to determine the autism global neuro-atlas.Figure 11 demonstrates the optimum number of the selected features using each of the four RFECV models. The number of features selected on each trial is represented on the horizontal axis. On the other hand, the average performance of the $5 - folds$, when RFECV utilized that specific number of features, is represented on the vertical axis. The optimum number of features corresponding to the maximum average balanced accuracy score is denoted by a vertical red line. The features set containing that optimum number of features is then utilized to train the ML models of the following step in the pipeline. The results demonstrated in figure 11 are as follow. RFECV+lg1 selected only one feature, RFECV+lg2 selected 207 features, RFECV+rf selected 2 features, and RFECV+lSVM selected 11 features. It is worth noting that all the selected features are already included in the RFECV+lg2 set. The selected features' set, representing the global neuro-atlas, is found in supplementary materials 1.

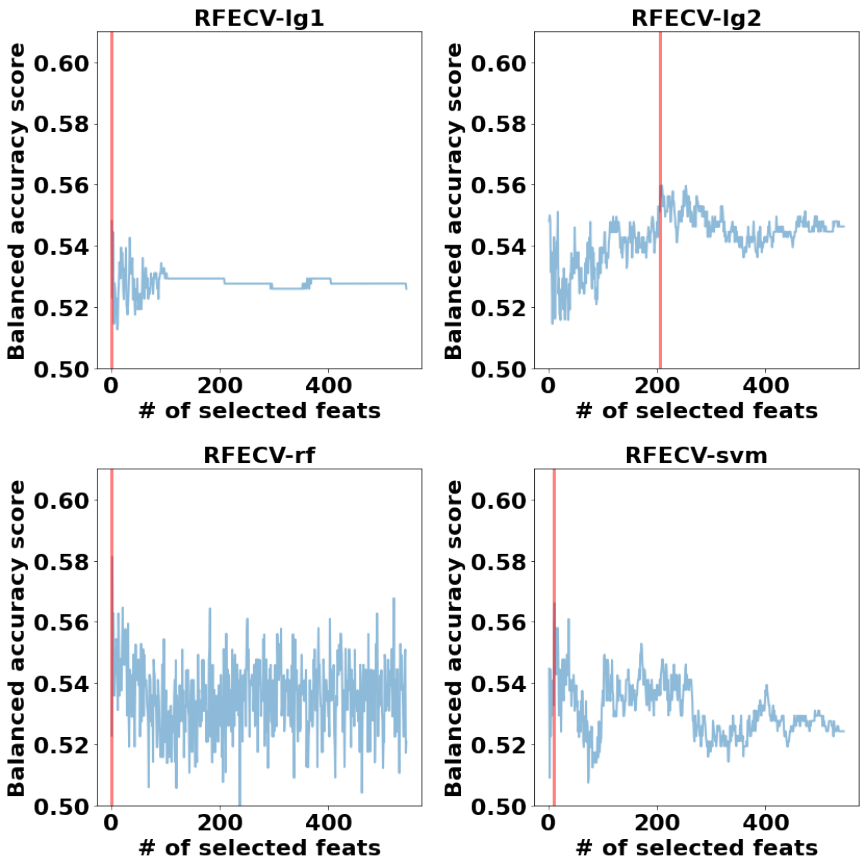


Figure 11. Number of selected features vs. the balanced accuracy score when applying RFECV with different classifiers. The red vertical line labels the number of features corresponding to the maximum balanced accuracy score.

ML global model: the balanced accuracy scored is calculated for each of the four selected features' set utilizing all of the eight classifiers. Figure 12 demonstrates

the mean and std. deviation over the 5-fold cross-validation when using the optimum hyper-parameters of each classifier. The highest balanced accuracy score, marked as red dot on the figure, is achieved when using the features (global neuro-atlas) selected with lg2 classifier and utilizing NN (0.716 ± 0.024) The optimum hyper-parameters used for NN are as follows: loss function is Cross-Entropy, solver for weight optimization is stochastic gradient descent, learning rate is set to be adaptive, l2 penalty (regularization term) is 0.0001, activation function is hyperbolic tangent function (Tanh).

Table 3. The classification accuracy score of Sabuncu et. al. study, and the proposed model with and without RFECV, along with the classifier used to achieve the score for each model and the number of features included in each model.

	Sabuncu et. al. results	Proposed pipeline without RFECV	Proposed pipeline
Accuracy	0.59 ± 0.02	0.597 ± 0.04	0.716 ± 0.02
Classifier	RVM	XGB	NN
# of features	20484	544	207

Similarly, for the null hypothesis, the same hyper-parameters optimization has been carried out for the eight classifiers, except that all the 544 were utilized without performing any feature selection. Figure 13 demonstrates the balanced accuracy score calculated for each of the optimized eight classifiers. The highest achieved score, marked as a red dot on the figure, is achieved with XGB classifier (0.597 ± 0.04). A summary of the global model results is shown in Table 3.

3 Chapter Discussion

In this section, we will discuss the global and local neuro-atlases. We will cover the common findings between the proposed study and previous literature. in this chapter, we focus on two aims: (i) Creating a discriminative model using a set of morphological features extracted from sMRI volumes of the brain of ASD and TD subjects, and (ii) defining global and local neuro-atlases that can be used to define ASD; consequently, those atlases can be used to understand the underling neurophysiology of the disorder. We define a discriminative model as a ML classifier which is trained using a set of features that best discriminate ASD and TD in the current dataset. in this chapter, we introduced two different models: (i) the global model which aims to classify ASD and TD in a heterogeneous dataset, (ii) the site-based model/local model which aims to classify ASD and TD in a less heterogeneous dataset based on demography.

Analysis of the relation between global and local models

RFECV algorithm is performed on the data matrix using four different classifiers: Lg1, Lg2, SVM, and RF. Each model selects the set of features that maximizes the classification's balanced accuracy score via 10-fold CV process. Those sets of features are considered as local neuro-atlases for ASD within each site. The algorithm is performed only one time in case of the global model, while it is performed 12 times on each site in case of the local model. The main motivation behind estimating a

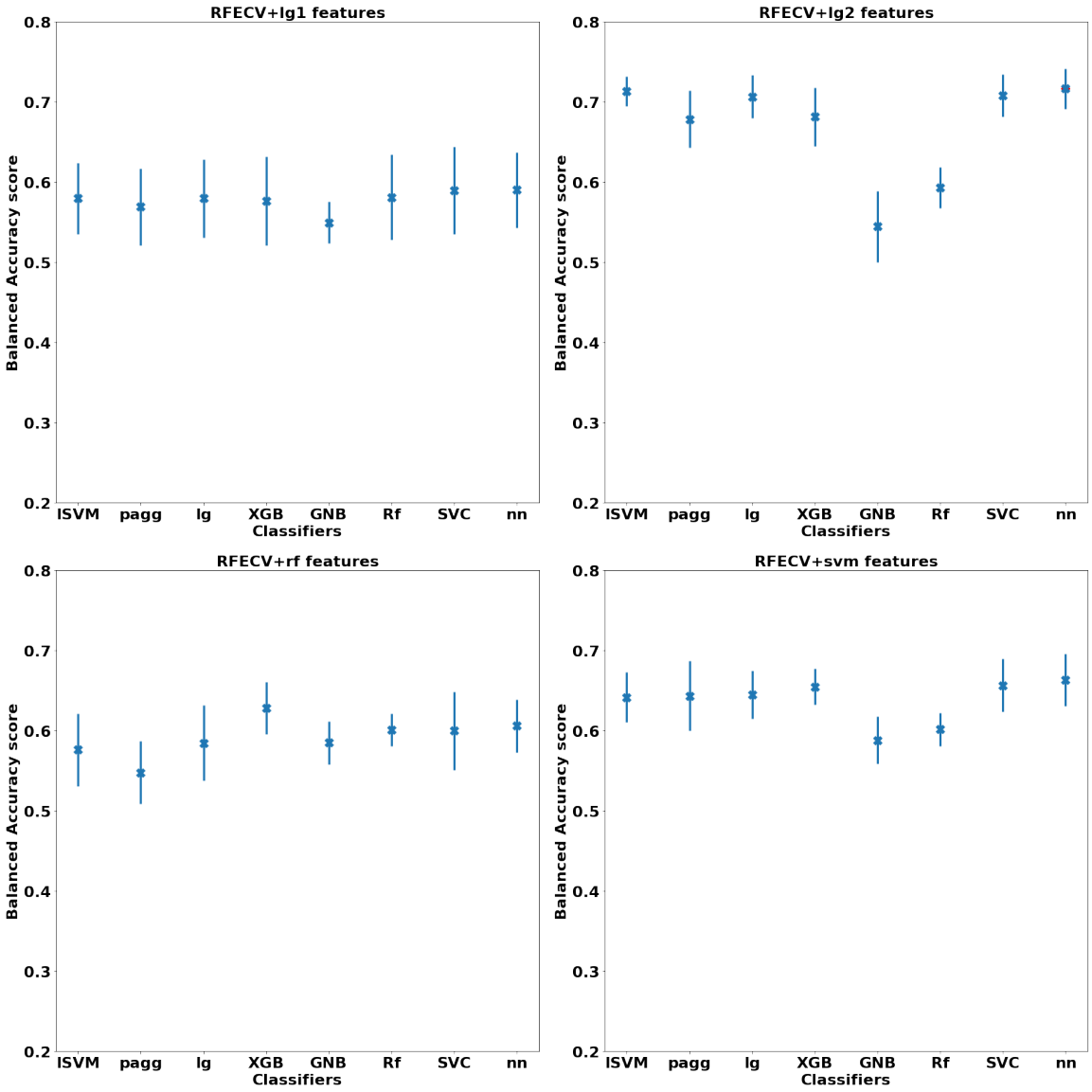


Figure 12. The highest testing balanced accuracy score ± standard deviation achieved by each of the optimized classifiers with applying RFECV with the core classifiers. The red dot labels the classifiers with the highest mean testing accuracy over the 5-folds cross-validation.

global model is the promising results of the local model. The discriminative set of features is determined at the ML step where a set of linear and non-linear classifiers are trained on the selected features set. The set of features corresponding to the maximum balanced accuracy score via 5-fold CV is said to be the discriminative feature set (neuro-atlas) for either the global model, or for a specific site. In the case of global model, RFECV+lg2 with 207 morphological features corresponding to a balanced accuracy score of (0.715 ± 0.024) is set to be the model that has selected the discriminative set of features. The 207 features can be found in supplementary materials 1. Table 4 shows the RFECV model corresponding to the maximum clas-

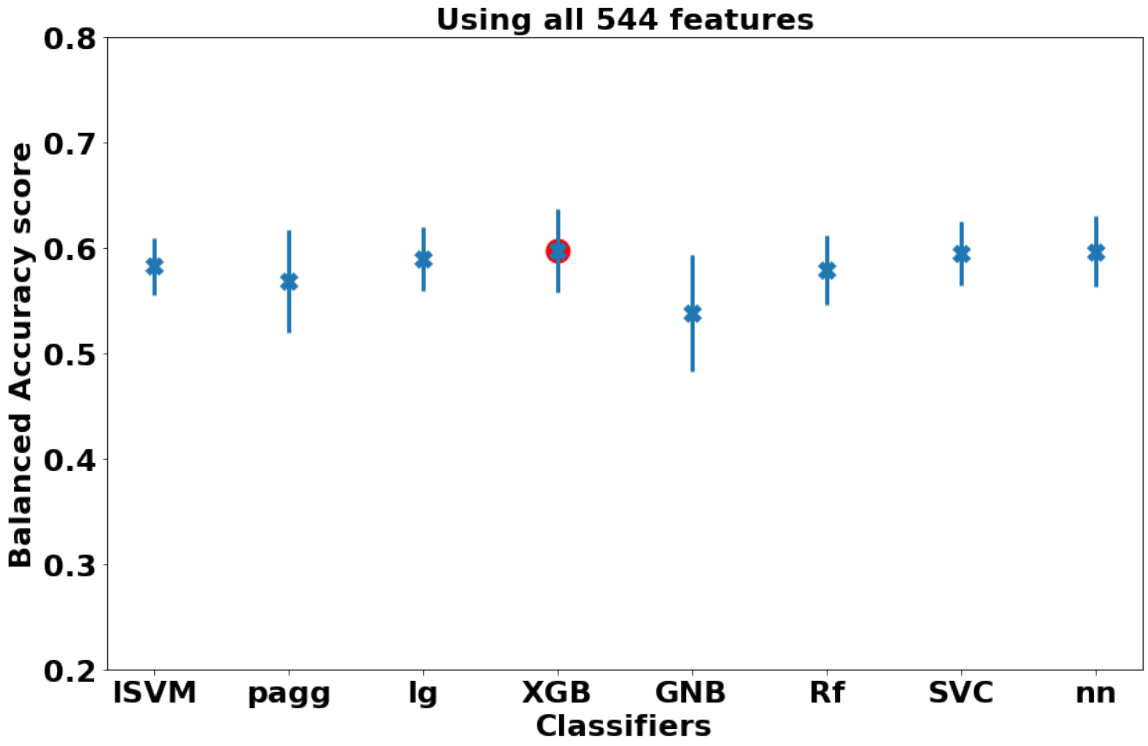


Figure 13. The highest testing balanced accuracy score, plus or minus one standard deviation, achieved by each of the optimized classifiers without applying any feature selection algorithms. The red point labels the classifiers achieving the highest performance.

Table 4. Summary statistics of the selected features using the local model

Site	# of features	# of mutual features with the global model
Caltech	217	74
CMU	18	5
Leuven	6	4
MaxMun	14	7
OHSU	79	32
Olin	60	26
Pitt	16	8
Stanford	7	1
Trinity	18	5
UCLA	7	2
UM	54	23
Yale	21	9

sification balanced accuracy score of a 5-fold CV process, along with the size of the selected features' set and the number of common features with the global model. It is anticipated to have common features between the global model and local models since the global model can be thought of as something "similar" to the aggregation of the local models. However, this raises the question about whether any of the mutual features are repeated within multiple sites as well as the global model.

Across all the 544 features, the highest number of selections to be given for a feature is five times i.e., a feature is selected by the global model and four different local models. We selected the features that are nominated by the global model as well

Table 5. The most frequent morphological features and brain regions to be selected by RFECV+lg2 discriminative model

Morphological feature	Hemisphere	Brain region
Curvature	Left	Middle Temporal Gyrus
Volume	Left	Middle Temporal Gyrus
Volume	Left	Transverse Temporal Gyrus
Surface area	Right	Transverse Temporal Gyrus
Curvature	Left	Frontal Pole
Curvature	Right	Rostral Anterior Cingulate
Curvature	Right	Transverse Temporal Gyrus
Thickness	Left	Middle Temporal Gyrus
Thickness	Left	Rostral Middle Frontal Gyrus
Thickness	Left	Superior Temporal Gyrus
Volume	Right	Lateral Occipital Sulcus
Volume	Right	Posterior Cingulate Cortex
Surface area	Left	Superior Frontal Gyrus
Surface area	Right	Banks of Superior Temporal Sulcus
Surface area	Right	Pars Orbitalis
Surface area	Right	Pars Orbitalis

as at least three different local models in order to study them to be candidate imaging markers. Table 5 demonstrates the morphological features, the hemisphere, and the brain regions with the highest frequency of selection among sites in a decreasing order.

A total number of 16 features have met the aforementioned criteria. Among those 16 features, the distribution of the morphology is as follows: (i) surface with 5 occurrences, (ii) curvature with 4 occurrences, (iii) volume with 4 occurrences, and (iv) thickness with 3 occurrences. There are eight brain regions from the left hemisphere, and eight brain regions from the right hemisphere. The most frequent brain regions among the 16 most common are: Middle Temporal Gyrus, and Transverse Temporal Gyrus. The unique selected brain regions are shown on brain template image in Figure 14.

Detected ASD Neurocircuits

The development of a CAD system for the early diagnosis of autism must include central features which correspond to the effect of the increasing neuropil within a brain region. Through these experimental approaches, structural MRI parameters related to the expanding neuropil are relevant to defining ASD neurocircuits. The effect of the local diagnosis identifies ASD related brain regions which fit into the Research Domain Criteria (RDoC) neural circuits and are similar circuits predictive of ASD diagnosis at 24 months. Neurocircuits detected in the proposed work is shown in Table 6

RDoC-defined language and social neural circuits are overly represented in the morphological data. Middle Temporal Gyrus has been found to be associated with autism volume of left Middle Temporal Gyrus [125,126], functional connectivity of left and right Middle Temporal Gyrus [127, 128]. Transverse Temporal Gyrus has been found to possess significant difference between ASD and TD in terms of magnetic mismatch field evoked by voice stimuli in 3- to 5- year-old subjects [129]. Superior Temporal Gyrus is found to be possess a greater gray matter volume in ASD subjects

Table 6. The Brain Regions Linked to RDoC Neural Circuits, Behavioral/Cognitive Domains of the ADOS, and ASD Structural Diagnosis

Component	RDoC Neurocircuit	ADOS Domain	Anatomical Correspondence
Restricted Interest	Reward Learning/Habit	RRB	Frontal Pole
Attention	Ventral/Dorsal Attention	Total	Rostral Middle Frontal Gyrus
			Lateral Occipital Sulcus
Language	Receptive/Expressive	SA	Middle Temporal Gyrus
			Transverse Temporal Gyrus
			Pars Orbitalis
			Superior Temporal Gyrus
			Superior Temporal Sulcus
Social	Affiliation & Attachment	SA	Frontal Pole
Social	Self Aware	SA	Superior Frontal Gyrus
			Posterior cingulate gyrus
Social	Understanding the Mental States of Others	SA	Rostral ACC
			Superior Temporal Sulcus
Executive Function	Working Memory	SA	Superior Frontal Gyrus
			Rostral Middle Frontal Gyrus
	Performance Monitoring	SA	Rostral ACC

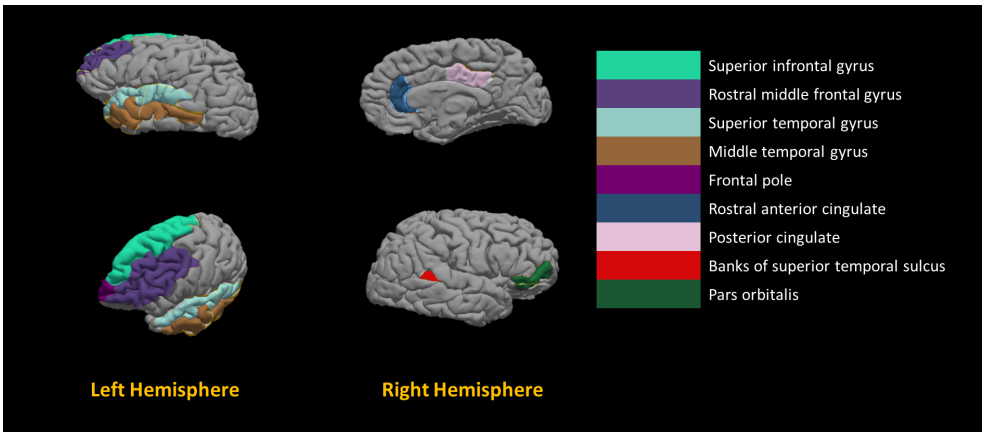


Figure 14. Visualization of the brain regions that are most commonly chosen by the RFECV+LG2 method.

who show drive to assess or construct rule-based systems [130], a diminished functional connectivity between cerebellum in ASD subjects of ABIDE I dataset [131], other functional connectivity differences, occurred in superior temporal gyrus, between ASD and TD subjects in 3 age cohorts $< 12, 12-19, > 20$ years old [132], more different morphological alterations have been reported for superior temporal gyrus in different studies [125, 133–135]. Banks of Superior Temporal Sulcus is found to demonstrate less developmental trajectory of the surface area [136], and an increased thickness of the right Banks of Superior Temporal Sulcus [137].

Reward learning, attentional, social, and executive function RDoC defined neural circuits are also impacted and classify individuals as ASD or TD. Frontal pole has been found to have a decreased thickness in the left hemisphere for the ASD subjects in the age of $(9.2 \pm 2.1 years)$ [138]. Rostral Anterior Cingulate has found to demonstrate abnormally increased activation on specific visual tasks in fMRI and reduced fraction anisotropy in the white matter [139], cortical thickness alteration [83]. Rostral Middle

Frontal Gyrus is found to be the most important feature for classification in a dataset of children older than 6 years old [140], the volume of rostral middle frontal was found to be the only statistically significant difference between ASD and TD group with the age of $33 \pm 9.1 year - old$ [141]. Lateral Occipital Sulcus is found to have a negative slope of relationship between local gyrification index and age greater in ASD than TD; moreover, a negative correlation between the local gyrification index and cortical thickness if found to be less in magnitude in ASD than TD [142], A study focusing on finding brain regions that can be targeted by noninvasive brain stimulation (NIBS), for ASD treatment, has labeled Lateral Occipital Cortex as a potential target for NIBS [143].

Similarly, the social function associated circuits involving the Posterior Cingulate Cortex are found to possess irregularly distributed neurons, and there is an increased density of neuron in the underlying white matter in the same region [144]; furthermore, in an fMRI study, ASD group showed weaker connectivity between the posterior cingulate cortex and superior frontal gyrus, stronger connectivity between posterior cingulate cortex and both the right temporal lobe and right Para hippocampal gyrus, and poorer social functioning in the ASD group was correlated with weaker connectivity between the posterior cingulate cortex and the superior frontal gyrus [145]. Banks of Superior Temporal Sulcus is found to demonstrate less developmental trajectory of the surface area [136], and an increased thickness of the right Banks of Superior Temporal Sulcus [137].

The proposed pipeline is anticipated to achieve better results than those in the literature because of the way that the morphological feature values are aggregated is less prone to outliers, RFECV implementation with more than one classifier to cover as many assumptions on the relationship between the features and the target as possible while selecting the features, performing hyper-parameter optimization using grid search on eight classifiers to achieve the optimum results given the selected set of features. All the codes utilized in the proposed pipeline implementation are available upon request.

CHAPTER IV
DIAGNOSING AUTISM SPECTURM DISORDER AS A HETEROGENEOUS DISORDER: STRUCTURAL MRI APPROACH

This chapter provides a novel approach to understand and diagnose ASD. In this chapter we consider ASD as a multidimensional disorder such that each dimension represents a behavioral domain. We measure the severity of every subject across each behavioral domain, and by aggregating the severity scores, we assign a global diagnosis for each subject. This approach allows us to understand ASD by dissembling it into different behavioral domains which would allow physicians to design personalized treatments to each subject based on the severity of his/her behavioral disorder. Moreover, we utilize interpretable model, LIME, in order to map the decision made by each classifier to the corticial features which contributed the most to that decision which would help in building behavioral neuroatlases to deeply understand how ASD alters the brain cortex.

1 Background

In this chapter, we are classifying ASD by identifying subjects in a multi-dimensional behavioral space. A multi-dimensional behavioral space is defined by an array of SRS behavioral score for each subject. Thus, we identify location of each subject on the spectrum via identifying his severity within each behavioral module. Furthermore, we identify the cortical regions which are most correlated with the severity of each behavioral category. **To the best of our knowledge, we are the first team to classify ASD subjects based on their predict behavioral severity scores, and annotate the most correlated brain region with each behavioral category**. However, the closest study, we could, find to our work is the study conducted by Moradi et al. [91] who predicted symptoms severity score of individuals with ASD based solely on cortical thickness. In the following section we will cover Moradi et al. study as well as the most recent studies that classify ASD using machine learning and structure MRI (sMRI).

In this chapter, we propose a novel personalized comprehensive ML model to classify ASD via two classification phase. First phase, we classify the severity of each behavioral category as they are defined in the SRS module (Communication, Mannerism, Cognition, Motivation, Awareness, Total). The severity levels are defined as follow: (i) TD, (ii) mild, (iii) moderate, and (iv) sever, such that mild, moderate, and sever are all considered to be ASD. Second phase, we combine the predicted severity scores for each behavioral category to get a final classification for each subject.

First phase, we search for the neuroimaging markers that define every severity level (TD, mild, moderate, and sever) of each behavioral category. Those neuroimaging markers are used for training a machine learning model to classify subjects based on their severity level (TD, mild, moderate, or sever) of each behavioral category. The first phase is repeated 51 times while shuffling the training-validation set. The cortical regions with statistically significant classification accuracy are selected for each behavioral module to build the corresponding behavioral neuroatlas. At the end of the second phase, we propose a CAD report, illustrated on 2 randomly selected case studies from the testing set, which maps the results of phase one with the most contributing cortical features, and maps the final diagnosis of phase two with the behavioral report. By this CAD report, we attempt to provide physicians with a comprehensive clinical report explaining the rationale behind the suggested diagnosis. The main motivations behind using solely morphological features which are extracted from the brain cortex, while neglecting the subcortical structures, are (i) segmentation of subcortical structures is more challenging and prone to error more than the cortex segmentation, and (ii) most of the significant findings in the literature are achieved by utilizing the surface based morphometry (SBM) methods [2, 17, 52, 74, 75, 79–83, 85]. The main motivations behind utilizing the SRS instead of other gold standards are: (i) It is cost-efficient and less time consuming than other gold standards, and (ii) The dataset we are utilizing in this study has less missing data for SRS than the gold standard (TD subjects don't have any ADOS score in ABIDE II dataset). In this study, we avoided the limitations existed in the aforementioned literature such as (i) working on a small sample size [79,85], we utilized ABIDE II dataset which comprises 487 individuals with ASD and 557 TD [146], (ii) neglecting the heterogeneity nature of ASD [7, 85, 92–94], we split ASD into its behavioral components as described by SRS, and within each behavioral component, we classify subjects according to their severity, and (iii) neglecting age as a confounding variable affecting the morphological features of the cortex [7, 79, 85, 92–94], we adjusted volume and surface area features according to cortical growth curves as described by Coupé et al. [116]. To the best of our knowledge, we are the first team to build a multi-level classification system to classify ASD in terms of its behavioral components.

We hypothesize that the optimum classification results of autism can be achieved by splitting ASD into behavioral components, predict each behavioral component, and aggregate the behavioral predication into a final classification. **The main contributions of the proposed work can be summarized as follows: (i) modeling ASD as a multidimensional disorder such that each behavioral component is an independent dimension, and the severity of ASD is the magnitude of ASD in that dimension, (ii) building a comprehensive ML pipeline to find morphological features and brain regions that are correlated with different severity levels of each behavioral category of the SRS, (iii) finding the anomalous neuro-circuits caused by different severity levels of autism (e.g. neuro-atlases), and (iv) proposing an explanatory CAD report which maps extracted cortical features to the behavioral classification to the final diagnosis.**

2 Materials and Methods

A novel comprehensive ML framework (Fig. 15), comprises two classification phases, is proposed in this study. First phase is to select morphological features and find the anomalous neuro-circuits associated with every behavior defined in SRS. Moreover, the first phase contains a ML module that utilizes the selected morphological features to classify subjects to different severity levels within each behavioral group. Second phase aggregates/stacks the models that are trained in the first phase to provide a final classification to every subject in the training and testing dataset. In order to build the neuroatlases associated with each behavioral category, we repeat the first phase 51 times, while shuffling training-validation dataset, to find cortical regions which discrimination power of statistical significance $p < 0.01$. The framework starts with downloading the sMRI volumes of ASD and TD subjects provided by ABIDE II dataset [146], then the preprocessing of the sMRI volumes is performed by Freesurfer v6.0 [97–100]. Preprocessing consists of three stages which are: (i) intensity normalization, (ii) skull stripping, and (iii) brain segmentation. Each of the aforementioned stages comprises a set of substages which are going to be briefly discussed in the following sections. After preprocessing, the first classification phase begins with subjects being categorized according to behavioral modules of SRS, then within each module subjects are labeled according to their standard SRS score either to be TD, mild, moderate, or sever. For each subject, features are extracted and summarized in the form of eight numerical representations for each morphological feature for each brain region. For each behavioral module, a data matrix, and a target vector are created and passed to feature selection algorithm to select the candidate imaging markers. Dimensional reduced data matrices, based on the candidate imaging markers and the target vector, are then passed to the ML algorithms to select the best ML model that can be used for classifying ASD and TD subjects.

Figure 15 demonstrates the general block diagram of the proposed framework comprising three parts: the two classification phases post data preprocessing and feature extraction. The first part comprises data preprocessing, feature extraction, and categorizing subjects based on behavioral modules of SRS, we name subjects within each behavioral category as behavioral group. The second part is the beginning of the first classification phase (see Fig. 15, and Fig. 19), where each behavioral group is split according to their normalized severity scores into TD, mild, moderate, and sever such that $\boldsymbol{TD} \leq 59$, $60 \leq \boldsymbol{Mild} \leq 65$, $66 \leq \boldsymbol{Moderate} \leq 75$, and $\boldsymbol{Sever} \geq 76$. The third part is the beginning of the the second phase of classification (see Fig. 15, and Fig. 24), where the selected features of phase one for each behavioral group is selected from that group feature space creating a thinner matrix to be utilized in training of six different classifiers and perform a randomized grid search hyper parameter optimization to achieve the highest possible classification accuracy within each group. Each of the aforementioned phases will be described in detail in the following sections.

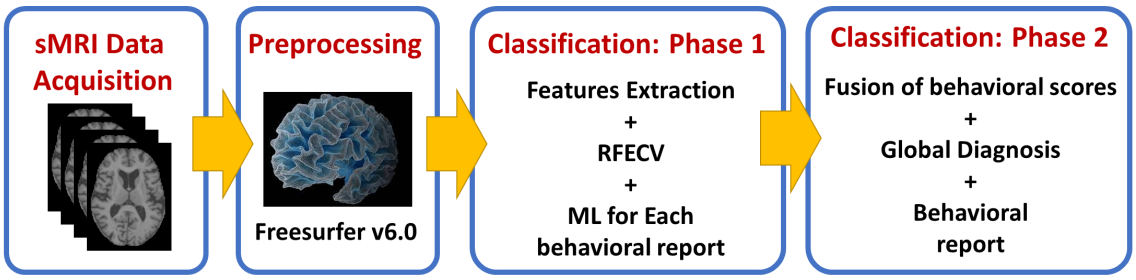


Figure 15. Overview of the proposed system starting from acquiring MRI volumes up to the diagnosis.

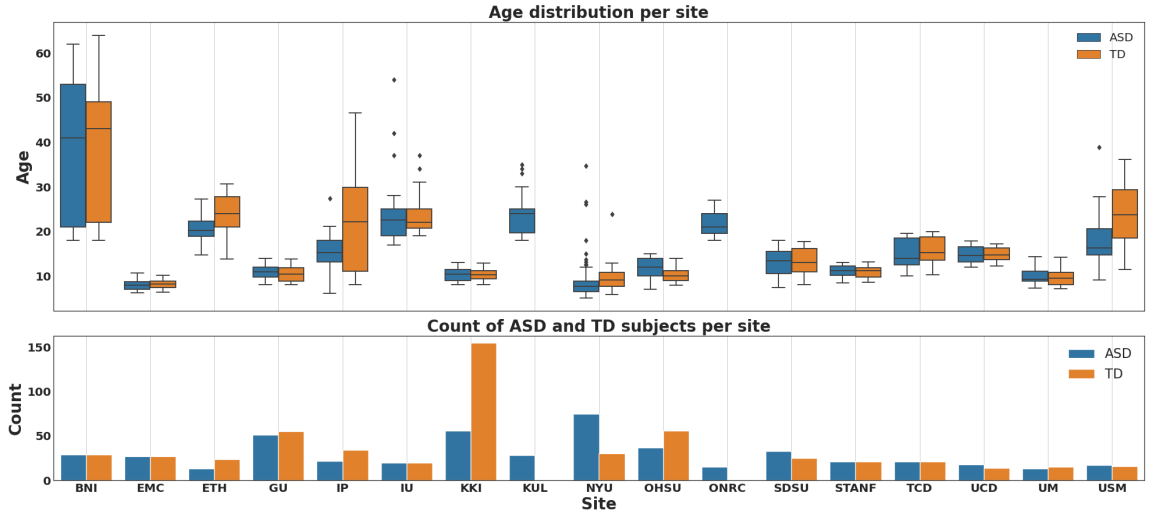


Figure 16. (Top) boxplot of subjects' age distribution on the vertical axis versus sites' names on the horizontal axis for each phenotype. (Bottom) number of ASD and TD subjects on the vertical axis versus sites' names on the horizontal.

Data

ABIDE II is the dataset utilized in this study. It is the second iteration of the Autism Brain Imaging Data Exchange (ABIDE), which aims to enhance the scope of brain connectomics research in ASD. It is worth noting that the first iteration of ABIDE was ABIDE I which was released in 2012 with 1112 subjects who were acquired from different clinical sites [146]. ABIDE II dataset comprises 521 individuals with ASD and 593 TD subjects. After preprocessing and quality control, we ended up with including 496 individuals with ASD and 542 TD subjects from total of 17 sites. The following figure 16 describes the summary statistics of the included cohort in this study.

-

All included subjects are then grouped into behavioral groups according to the SRS. Figure 17 demonstrates the missing values of each behavioral module provided by ABIDE II dataset such that each bar represents the number of missing values for that given behavioral test. The shortest six bars belong to the SRS tests. Figure 17

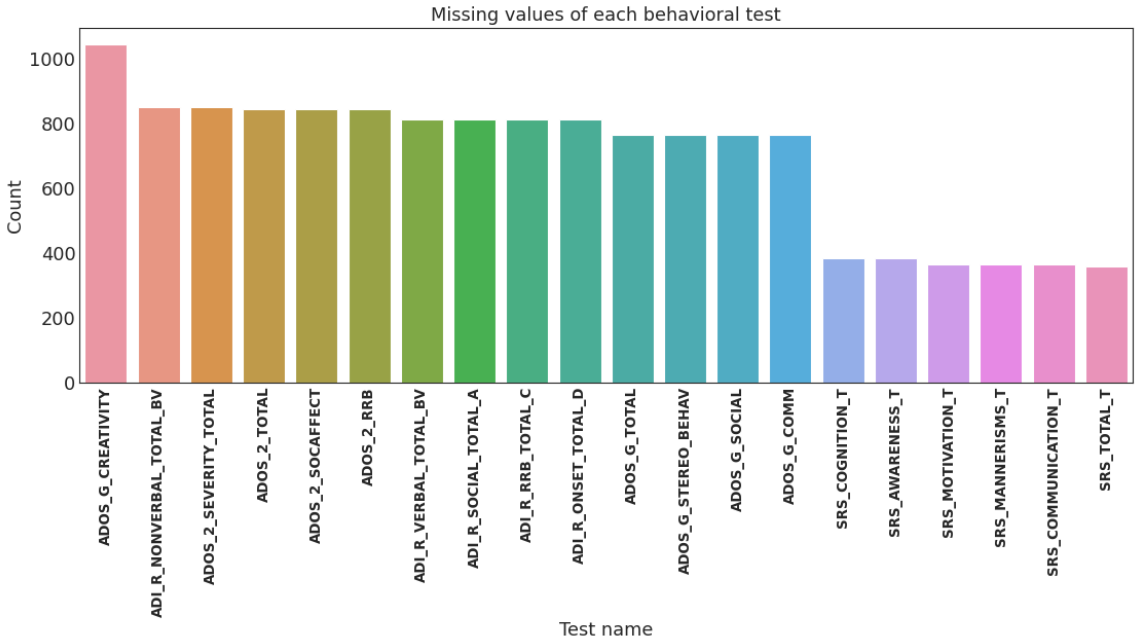


Figure 17. Bar plot demonstrating the number of missing values for each behavioral module included in the ABIDE II dataset

is the main motive behind selecting SRS to be the behavioral ground truth for this study since it possess the least number of missing values.

After grouping the subjects into behavioral groups, each group is considered as an independent dataset to be utilized in solving a multi-class classification problem between TD, and the severity of ASD. Figure 18 demonstrates the counts of TD and different severity levels of ASD within each behavioral category. As it is obvious, the data is imbalanced especially in the case of mild vs TD. In later sections, we will discuss how we proposed a way to achieve an unbiased model.

In the following section, we will briefly cover the preprocessing pipeline applied on the sMRI volumes of each subject to extract the cortical morphological features.

Pre-processing

Preprocessing is a crucial requirement to eliminate the between-subjects variability that may stem from data acquisition, different scanners, artifacts, or partial volume effects. Moreover, brain MRI scans usually contain non-brain tissues as it is shown in Figure **??**; Preprocessing section. FreeSurfer performs multiple steps on each sMRI volume to extract the morphological features. Those steps are intensity normalization, brain extraction and skull stripping, brain segmentation and area labeling, tessellation of the gray-white matter boundary, surface inflation and spherical atlas registration, and eventually cortical surface parcellation to the Desikan-Killiany (DK) atlas. The preprocessing is identical of that we performed in the last chapter III:1: Pre-processing

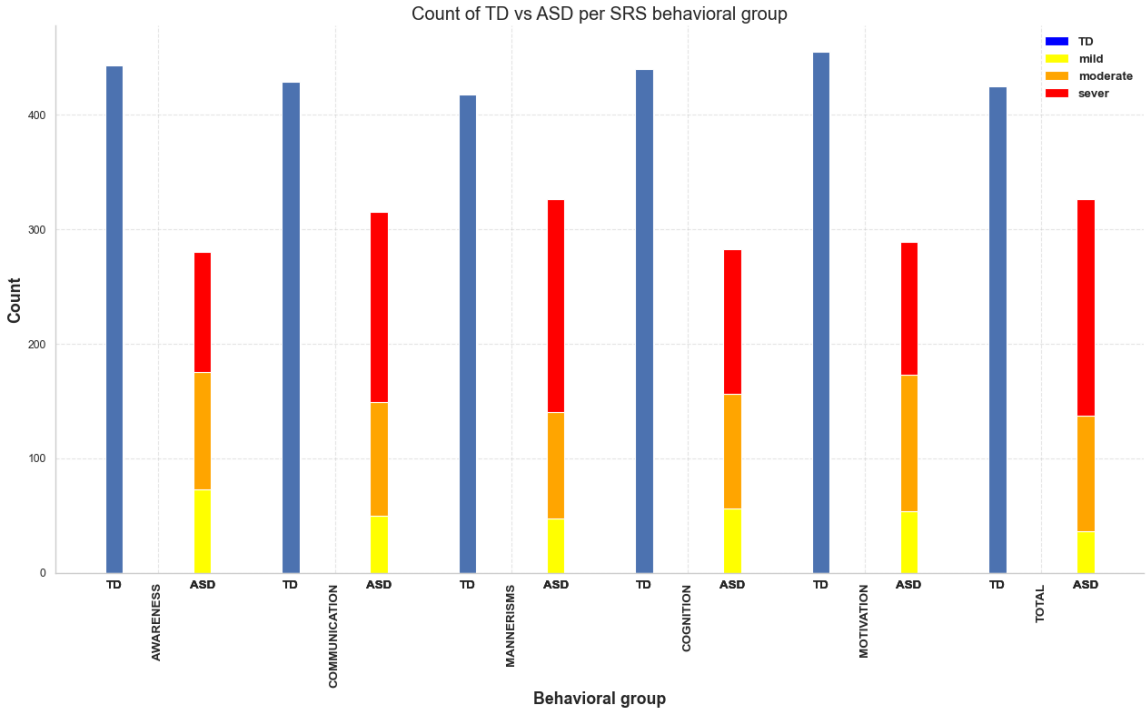


Figure 18. Bar plot demonstrating the counts of TD vs ASD severity group stacked on each others for each behavioral category

Phase I

In this section, we will discuss the first phase of the proposed framework. As shown in Fig. 19, the input to phase I is a matrix containing the values of the morphological features extracted using FreeSurfer for each subject along with a target matrix containing the severity score of each SRS test. The output of phase I, as shown in figure 19, is an optimized trained classifier and a set of selected features for each behavioral group. In the following sections, we will discuss each step of phase I in detail.

Feature Extraction

There are two outputs of FreeSurfer which are (i) a set of volumes for each subject describing each step of the pipeline (normalization, skull stripping and segmentation) as shown in Figure 15; preprocessing, and (ii) surfaces parcellated to DK atlas and containing the morphological features values at each point on a predefined mesh grid created on the brain.

In this study, we utilized the following morphological features to represent the brain of each subject: (i) surface area, (ii) volume, (iii) thickness, and (iv) curvature (see Figure 20). It is worth noting that thickness is calculated as the closest distance from the gray/white matter boundary to the gray/CSF boundary at each vertex on the tessellated surface [113], while curvature is measured as the average of the reciprocal of the principal radii [114]. For each of those features, we calculated the 20^{th}, 40^{th}, 60^{th}, 80^{th} percentiles within each brain region parcellated to the DK

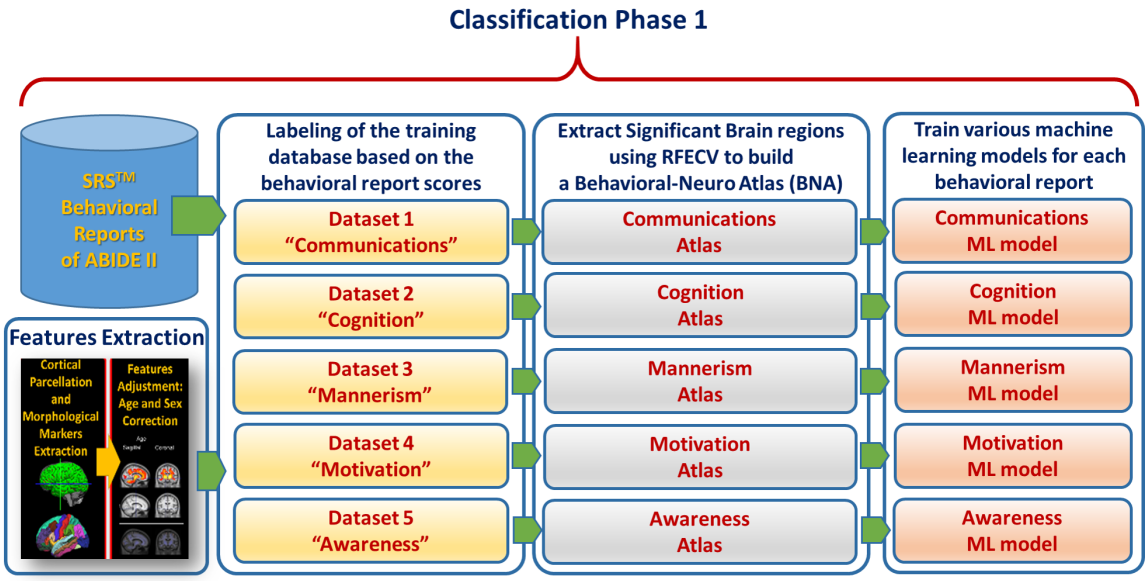


Figure 19. The pipeline of the proposed Phase 1 method of Classification.

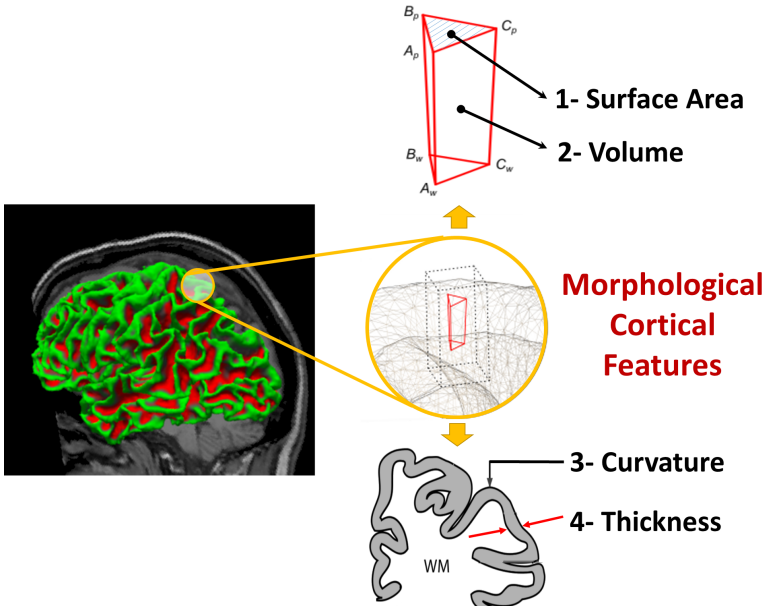


Figure 20. Morphological features extracted from brain surfaces by freesurfer.

atlas. The main motives behind choosing those four percentiles to represent each morphological feature of each brain region instead of mean and standard deviation, which is already provided by FreeSurfer are (i) to exclude outlier and preprocessing errors within each brain regions, (ii) and to have more samples representing the distribution of the morphological values within each brain regions. DK atlas parcellates the brain into 68 brain regions; 34 brain regions on the left hemisphere, and 34 brain regions on the right hemisphere. Therefore, each subject is represented by a vector

of 68 brain regions ×4 features ×4 = 1088 elements of a feature vector.

For each of the behavioral group, we subdivide the ASD subjects according to their severity into mild, moderate, and sever. The main point behind this dividing ASD within each behavioral group is to address the magnitude of the disorder.

For each of the behavioral groups, a data matrix is created (D_L) such that $D_L = \{d_t : d_t \in^{M,1088} \ \& \ 1 \leq t \leq 6\}$ each corresponding to one of the sites; d_t denotes the data matrix corresponding to behavioral group t and t = {communication, mannerism, motivation, cognition, awareness, total}. Total is an SRS behavioral module denoting the overall score of a subject. Each d_t has the size of $M \times 1088$, such that M denotes the number of subjects within site t; sequentially, y_t denotes the diagnosis vector corresponding to site t, and y_L denotes the set containing all the y_t for all sites. Equation 3 shows the shape and the symbolic representation for the data matrix and the target vector of each behavioral group.

$$d_t = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,1088} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,1088} \\ \vdots & \vdots & \ddots & \vdots \\ f_{M,1} & f_{M,2} & \cdots & f_{M,1088} \end{bmatrix}, \qquad y_t = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_M \end{bmatrix}, \qquad D_L = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_6 \end{bmatrix}, \qquad y_L = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_6 \end{bmatrix} \quad (3)$$

Feature Adjustment & Normalization

The feature vector corresponding to every subject contains the 20^{th}, 40^{th}, 60^{th}, and 80^{th} percentile of each morphological feature for every region. Morphological features don't share the same units of measurement; for instance, surface area is measured in mm^2, while volume is measured in mm^3. Consequently, we anticipate having different ranges of values, which might adversely affect the performance of the classifiers [117].

In this study, we utilized minimum-maximum normalization between 0 to 1 as it is one of the most common normalization methods used for biomedical data [118]. Consequently, each column in the data matrix D is normalized between 0 to 1 using the equation 4.

$$\tilde{f}_{i,j} = \frac{f_{i,j} - \min_i\{f_{i,j}\}}{\max_i\{f_{i,j}\} - \min_i\{f_{i,j}\}} \quad (4)$$

Where $\tilde{f}_{i,j}$ and $f_{i,j}$ denote the normalized feature value j and the original feature value j corresponding to the subject i, and $\min(f_j)$ and $\max(f_j)$ correspond to minimum and maximum values of the feature vector j respectively. The output normalized matrix is denoted by d_{tn} for the behavioral groups where $1 \leq t \leq 6$.

Feature selection

To implement a Computer-Aided Diagnosis System (CAD) for accurate diagnosis of autism, we have to use a neuro-atlas tailored to the specific developmental patterns of the brain in autism. Unfortunately, as far as our knowledge, ASD literature does not contain any ASD specific neuro-atlas, or even a behavior-specific neuro-atlas that can be utilized in training ML classifiers and build a CAD system. Therefore, in

this study, we take the first steps toward building behavioral based neuro-atlases that can be used to classify subjects into different ASD behavioral groups. In other words, we are proposing a neuro-atlas that can be used to objectively predict the severity score of a given subject if that subject was assigned to one of the SRS tests. To achieve this goal, we utilize recursive feature elimination with cross-validation (RFECV) method to select the set of brain morphological features which maximizes the balanced accuracy score of classifying ASD within each behavioral group. Since, RFECV is one of supervised feature selection algorithm, we split the every behavioral group into k-folds, $k = 5$ in our case, and for each fold (i) train a predetermined classifier using the training set, (ii) evaluate the performance of the trained classifier on the validation set, (iii) save the classifier's score on the validation set, (iv) find the least significant feature according to the trained classifier, (v) remove the least significant feature from the model, (vi) repeat the whole processes until you end up with only one feature. Again, repeat the whole process for each fold, calculate the average performance of the k-fold cross-validation (CV) when: using all features to train the classifier; using all features but one, and so forth, to the point of classification on a single feature. Find the number of the features at which the classifier has the maximum performance score, assuming it is N_f features. N_f is the optimum number of features to be selected. Perform the whole algorithm again over all the subjects to find the most N_f significant features. The algorithm is summarized in Fig. 21. For further details regarding the algorithm and its implementation, the reader is suggested to read Guyon et al. [119] and Pedregosa et al. [120], respectively.

For each of the behavioral groups, RFECV is implemented using four different classifiers building up total of four models to find the set of features that maximizes the balanced accuracy score of classifying subjects into TD vs mild ASD, TD vs moderate ASD, and TD vs sever ASD. Therefore, for every behavioral group, there are 12 selected set of features. Those 12 sets of features are divided into 4 sets of features act as neuro-atlas candidates to describe and differentiate between TD and each severity level of ASD (mild, moderate, and sever). Each of the selected set of brain morphological features maximizes the balanced accuracy score according to the underlying hypothesis of the utilized RFECV model. Each of the four selected models, to integrate with RFECV, posses one of the following underlying hypotheses: (i) models that assume that subjects are non-linearly separable and thus select a set of feature that constructs a feature space where the subjects are separated via non-linear kernel, and (ii) models that assume that subjects are linearly separable and thus search for a set of features that constructs a feature space where the subjects are linearly separable in . RFECV+RF, RFECV+Light Gradient Boosting Machines (LGBM) implements the first group of models with the underlying assumption that the classification accuracy can be maximized with a non-linear kernel in the selected feature space. On the other hand, RFECV+logistic regression (lr), RFECV+linear SVM (lsvm) implements the second group of models with the underlying that the classification accuracy can be maximized with a linear kernel in the selected feature space. Figure 22 illustrates the processes of feature selection using each of the aforementioned classifiers resulting into four different sets, each set comprises six sets of selected features for each behavioral category. Each of the RFECV models is exe-

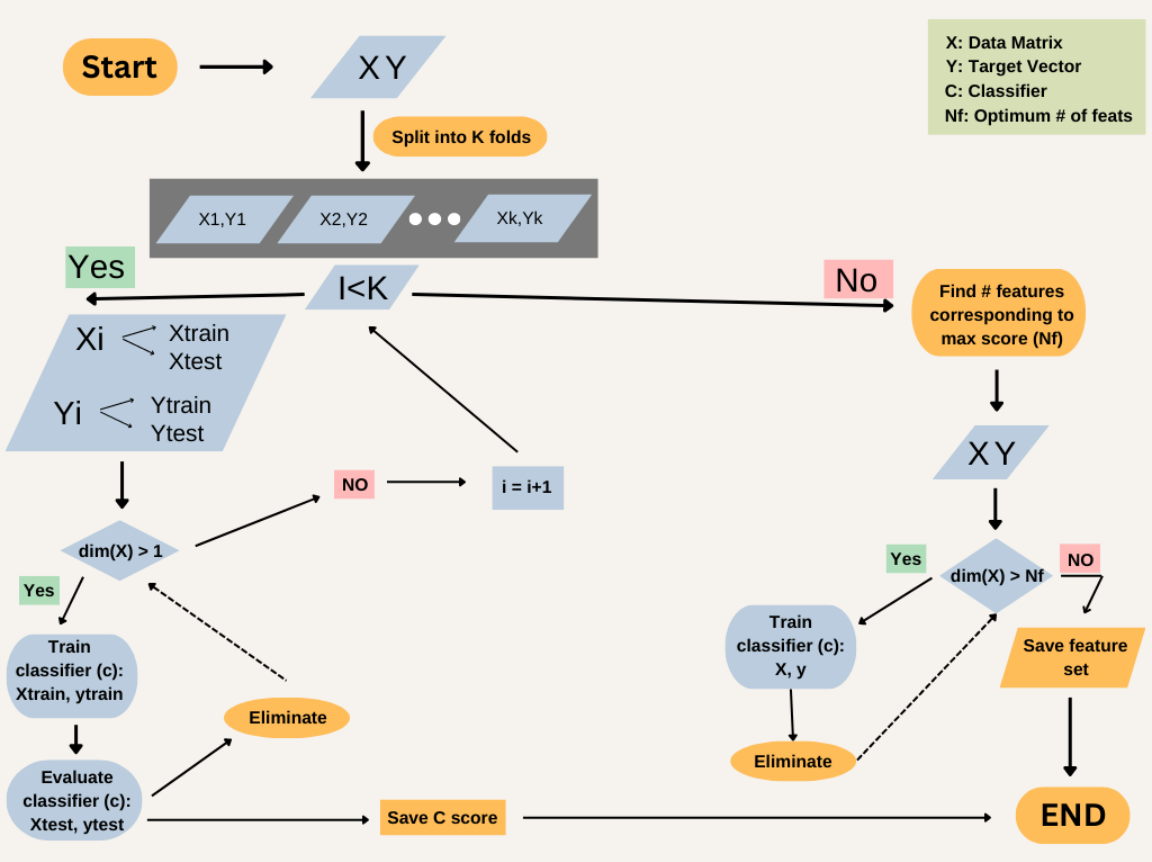


Figure 21. Flowchart illustrating RFECV algorithm for any classifier.

cuted with 5-fold CV; such that we iterate over all the 1088 features, removing one feature at a time, perform 5-fold CV on the current sample, and calculate the average balanced accuracy score.

This step outputs four sets of features for each severity level within each of the behavioral group with total number of 4 RFECV models $\times$ 3 severity levels $\times$ 6 behavioral groups = 72 sets of features. Those 72 sets of features are used to reduce the dimension of each of the data matrices of each behavioral group from 1088 columns/features to s_{knm}, where the size of s_{knm} is $n \times m$ such that n is the number of subjects within behavioral group k, and $m \leq 1088$. Consequently, we define a neuroatlas for a specific behavioral group as the brain regions included with a selected set of morphological features which achieves the maximum value of the average of the balanced accuracy score and f1-score, as defined in [147], in the following ML step. More details are provided in the ML section. It is worth noting that f1-score is only utilized at the ML step as a confirmation step of the ML models since all of the 30 selected set of features are chosen using solely the balanced accuracy score.

Scoring metrics

The first metric that we utilized in this study is the balanced accuracy score that was introduced in 2010 to solve the optimistic estimate occurs when a biased classifier

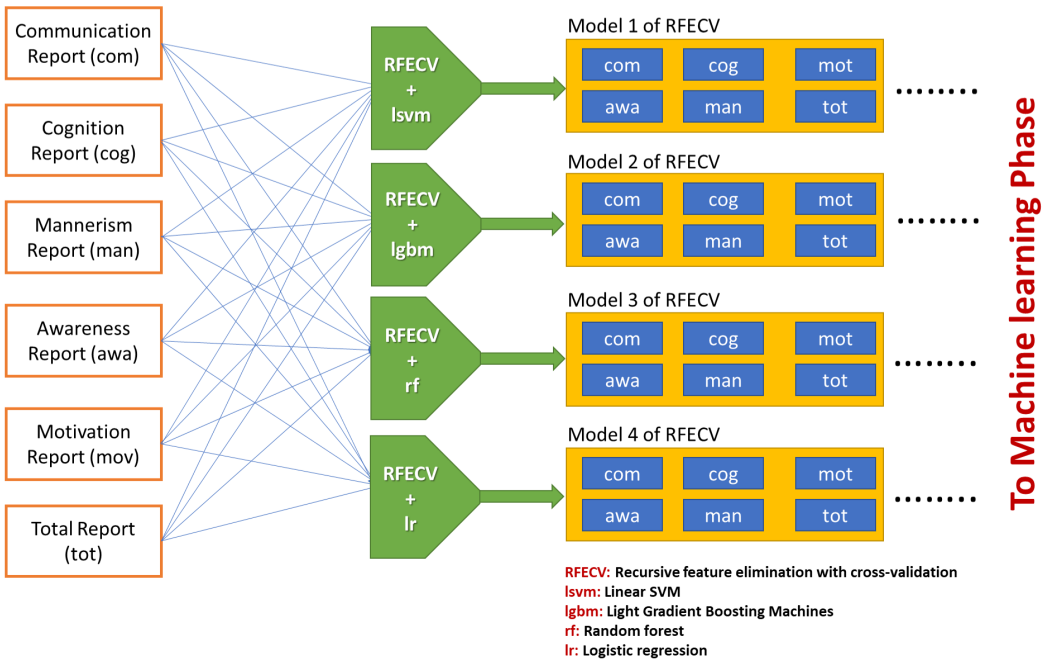


Figure 22. The steps of Recursive Feature Elimination CV and Building Neuro-Atlas

is tested on an imbalanced dataset [121]. The balanced accuracy score is defined by equation 6:

$$bacc = \frac{sensitivity + specificity}{2} \tag{5}$$

Where $score$ denotes the balanced accuracy score, $sensitivity$ is calculated as $\frac{TP}{TP+FN}$ such that TP denotes true positives, FN denotes false negatives, and $specificity$ is calculated as $\frac{TN}{TN+FP}$ such that TN denotes true negative, and FP denotes false positives.

The second metric that we utilized in this study is F-measure (F1), which is widely used in the context of classification of imbalanced datasets [148,149]. It was originally introduced to evaluate the ranking of documents retrieved based on a query [150]. It is interpreted as the harmonic mean of the two degrees of freedom of a confusion matrix.

$$F1 = 2 \times (\frac{PR}{P+R}) \tag{6}$$

where P denotes precision which is calculated as $\frac{TP}{TP+FP}$, and R denotes recall which is calculated as $\frac{TP}{TP+FN}$.

ML classifiers

For every s_{kn} for every behavioral group k, a ML model is initialized to be trained to classify TD vs mild ASD, TD vs moderate ASD, and TD vs sever ASD. In this study,

Table 7. hyperparameters range of each classifier

Classifier	hyperparameters	var_name	values
Linear SVM	Regularization parameter	C	0.1, 1, 5, 10
	Norm used in the penalization	penalty	"l1"l2"
	Loss function	loss	hinge ßquared_inge"
Ridge Classifier	Regularization parameter	α	0.1, 0.2, 0.3, ..., 5
	Normalization	normalize	(True, False)
Logistic Regression	Norm used in the penalization	penalty	"l1 "l2 ëlasticnet"
	Regularization parameter	C	0.1, 1, 5, 10
	Algorithm to use in optimization problem	solver	newton-cg "lbfgs "liblinear ßag ßaga"
Light Gradient Boosting Machines	l1 regularization	*reg_alpha*	0, 0.1, 0.2, ..., 5
	l2 regularization	*reg_lambda*	0, 0.1, 0.2, ..., 5.
	Number of estimators	*n_estimators*	100, 150, 200, ... 5000.
	learning rate	*learning_rate*	0.01,0.02, 0.03, ..., 1.
	Maximum depth for each tree	*max_depth*	1, 2, 3, ..., 10
	Number of leaves	*num_leaves*	3, 4, 5, 6, ..., 32
Random Forest	Number of trees in the forest	n_stimators	50, 100, 200, 500, 1000
	Function to measure the quality of a split	criterion	giniëntropy"
	Number of features to consider when looking for best split	max_eatures	sqrt"log2All features
	Minimum number of samples required to split an internal node	min_amples_plit	1, 2, 5, 10
	Whether to use bootstrap samples while building the tree or use the whole training set	bootstrap	True, False
Non-linear SVM	kernel used	kernel	poly"rbfßigmoid"
	Regularization parameter	C	0,0.1,1,5, 10
	Degree (when kernel is polynomial)	degree	2,3,4,5,6
	Kernel coefficient	gamma	scaleäuto"
	Independent term in kernel function	coef0	0.0,0.01,0.1,0.5,1,5,10,50,100

Algorithm 1 ML-train with hyperparameters optimization

Feature data matrix and target vector trained classifier object (c) Let X be the data matrix comprises M rows, and N_f selected columns from RFECV; let y be the target vector of length M such that $y \in \{0,1\}^M$ Split X, y into k folds. i.e $X_f = \{X_1, X_2, \ldots, X_k\}$ & $y_f = \{y_1, y_2, \ldots, y_k\}$ where $X_i \in \mathbb{R}^{\lfloor M/k \rfloor \times N_f}$ and $y_i \in \{0,1\}^{\lfloor M/k \rfloor}$ Let C be the set of the R classifiers utilized in the study such that $C = \{c_1, c_2, ..., c_R\}$ For each $c_i \in C$, let H_i be the tuple of associated hyperparameters as defined in table7 For each H_i let $v_{i,j}$ denote the range of allowable values of element j $c_i \in C$ HP_set$_i$ $\in$ $\{H_i \times v_i\}$ $(X_i, y_i) \in (X_f, y_f)$ Set the hyperparameters of classifier c_i to the values in HP_set_i Train classifier c_i using (X_{train}, y_{train}) Calculate the balanced accuracy score of c_i using X_{test}, y_{test} Save the trained classifier c_i and it is corresponding balanced accuracy score Save the calculated average balanced accuracy score of all the iterations along with the corresponding HP_set_i Save the maximum of all average balanced accuracy scores and the corresponding HP_set_i, along with classifier c_i $c_i \in C$ return the maximum score and the corresponding HP_set_i

we decided to perform one vs one classification for multiple classes and to classify each of the ASD behavioral severity vs TD. The rationale behind this decision is that we are interested to see how the morphology of TD brain differs from the morphology of ASD with different severity scores without over analyzing the discrepancies within the ASD itself. Throughout this study we are focusing on our primary objective which is building ML based neuro-atlases for the ASD. ML classifiers, with linear hypothesis and non linear hypothesis, are selected to classify subjects within every behavioral group. The training process is repeated three times for classifying subjects into TD and each of the severity levels, similar to the experiment design of feature selection. The utilized six ML classifiers are split into two main divisions: (i) Linear classifiers, (ii) Non-linear classifiers. The linear classifiers set comprises LR, LSVM, and Ridge classifier. The non-linear classifiers set comprises RF, SVM with radial basis function (SVM-RBF), and LGBM. The reduced data matrix s_{kn} of each behavioral group k is fed to every classifier for training. Figure 23 illustrates how the ouput of the feature

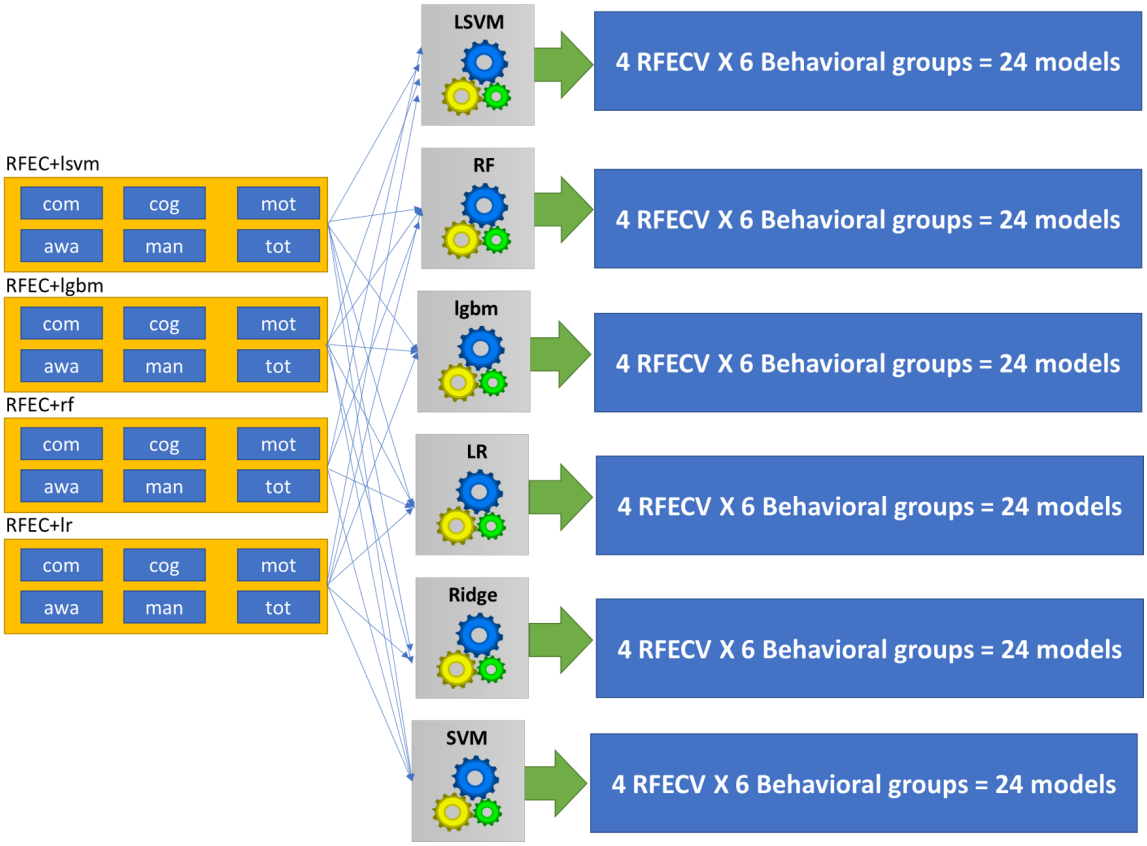


Figure 23. ML Model

selection step is given as the input to each of the aforementioned classifiers to build total of 144 classification model, 24 models for each behavioral group (4 RFECV and 6 classifiers). Every classifier is optimized using randomized grid search over 5-folds cv. For each classifier, the hyperparameters and their ranges are defined in Table 7. The random sampling of the hyper parameter space is repeated 500 times for each classifier. The hyper parameters of each classifier corresponding to the maximum average cross validation balanced accuracy score is selected as the output of this stage. The detailed algorithm is shown in algorithm 1.

There is an extra step is added to the end of the ML pipeline in order to make sure that the selected ML model is not overfitted. Each of the optimized ML classifiers is re-initialized with the optimum hyper-paramters's values, retrained using only 80% of the data and tested on the 20% leftout set. The testing balanced accuracy, and F1 score are calculated for each of the retrained classifiers to compare against the cross-validation score.

Top performing models, to passed to phase II of the framework, are selected based on their performance on the 20% testing set in terms of the average value of the balanced accuracy score, and F1 score. Once a ML classifier is selected for a given severity level within a given behavioral group, we perform backtracking to find the selected morphological features used for training that ML classifier.

Phase I optimization / Building Neuroatlases

The complexity of phase I makes it vulnerable to false positive findings because of one or more potential hidden confounding variables that might exist among a group of subjects in the dataset. In general, a complex experiment, especially in biomarker studies, posses many statistical caveats [151]. Moreover, phase I comprises two stochastic processes that would change the final results for different initial conditions. The two stochastic processes are: (i) Shuffling subjects for cross-validation, and (ii) the hyperparameters optimizations via random search. Although we sampled the hyper parameter space of every ML classifier 500 times, there is no guarantee that the resultant model is optimum, at least for the data in hand.

We executed Phase I 51 times with different random seeds for each of the two stochastic processes. The rationale behind repeating phase I is to be more confident with both the selected morphological feaatures, and the optimized model. We claim that a robust discriminative morphological features, which can be generalized outside ABIDE II dataset, should be consistently selected by RFECV across different samples of training-validation sets. Consequently, that robust discriminative set of morphological features should yield the highest possible accuracy scores when utilized with a hyperparameters optimized ML classifier.

Moreover, we were intrigued by the question of whether the set of morphological features, that corresponds to the maximum classification accuracy for each behavioral group, are selected due to chance, or they actually reflect some sort of underlying neuro-physiology that defines ASD. Therefore, we calculated the probability of finding each of the selected morphological features for n number of times out of the 51 trials using binomial distribution as going to be explained in the following section.

Eventually, the set of the morphological features corresponding to the maximum classification accuracy score for each behavioral group, and the optimized ML classifiers corresponds to every behavioral group are passed to phase II of the framework to be used to predict the final classification of subjects. Phase II is explained in detail in the following section.

Neuro-atlas statistical significance

Testing the significance, of each of the selected features within each neuro-atlas, was a challenging task, especially given the way we designed the framework. The framework is primarily designed to build a CAD system and annotate the features that maximize the classification accuracy of that CAD system. However, studying the significance of the findings is as important as the classification accuracy, if it is not more important. Therefore, we set the null hypothesis, and then analytically calculated the probability of including a specific feature in a specific neuro-atlas under the given null hypothesis.

The null hypothesis is that RFECV algorithms randomly select m features of each behavioral group data matrix. Thus, over the 51 repetitions, what is the probability of observing the neuro-atlas features under the assumption of null hypothesis? To answer this question, we thought of RFECV the opposite way of how it works. We though of the process as RFECV selects a feature, instead of eliminating a feature, up

to m features. Therefore, for every repetition, the probability of randomly selecting any feature out of the 1088, through m samplings without replacement, is $m/1088$ such that m is the number that RFECV selected for this behavioral group for this severity level for this repetition out of the 51 repetitions. Let feature f_i such that $1 \leq\leq 1088$ gets selected r times out of the 51 repetitions, then the probability that f_i is randomly selected r times is defined as

$$P(f_i; 51, r) = \binom{51}{r}(\frac{m}{1088})^{\text{r}}(1 - \frac{m}{1088})^{\text{51-r}} \tag{7}$$

Equation 7 is the well-known binomial distribution equation [152]. Although equation 7 is too close to what we are looking for, it is not exactly the correct representation of our experiment. As we described the stochastic part of phase I in the last section, we can not ensure that for the same behavioral group, for the same severity level, the same m will be selected for every repetition. Therefore, a generalization of equation 7 is required to be applied on our case. We can reformulate our question to be: What is the probability of including feature f_i r times in the selected features set in total of 51 experiments given that the probability of selecting f_i varies across experiments? After introducing the variability of the probability of success of each feature based on the experiment, we can no longer calculate the probability $P(f_i; 51, r)$ using $\binom{51}{r}$. Therefore, we implemented a function that counts all the ways such that f_i can be observed r times giving the probability of selecting f_i at every experiment. The function can be found at [153] in directory $notebooks/probabilityOfSuccessComputations.py$. The statistical significance of the selected feature is performed on the neuro-atlas of every behavioral group, and on the aggregated neuro-atlases to inspect the most common features among all of ASD related behavioral disorders.

Although the significance would be calculated exactly using the previous technique, the number of combinations we need to considers grows exponentially, and becomes impossible to track. For instance, the number of combinations required to be considered to calculate the significance of selecting one feature 10 times out of the 51 experiments is $\binom{51}{10} = 12777711870$. Therefore, for every features, we assigned the highest selection probability that a features possessed across the 51 experiments as the p defined in the binomial distribution. Afterwards, we used the binomial distribution equation 7. This crude assumption might result in type II error which would lead to ignoring significant brain regions. Nevertheless, in this work, we attempted to be as conservative as possible while building the neuroatlases. Consequently, those brain regions, which are found to be significant, can be thought of the "core" brain regions that are directly associated with the disorder. A brain region is said to be significantly associated to ASD within a severity group of a behavioral group if and only if at least one of the morphological features of that features is utilized in one of the selected ML models of the 51 experiments, and have a p-value ¡ 0.001.

Eventually, since phase I is a supervised process, we wanted to study the relationship between the top performing models across all experiments testing and the behavioral groups, as well as, the relationship between the top performing models and the severity groups. We define the top performing models as the top 50th percentile

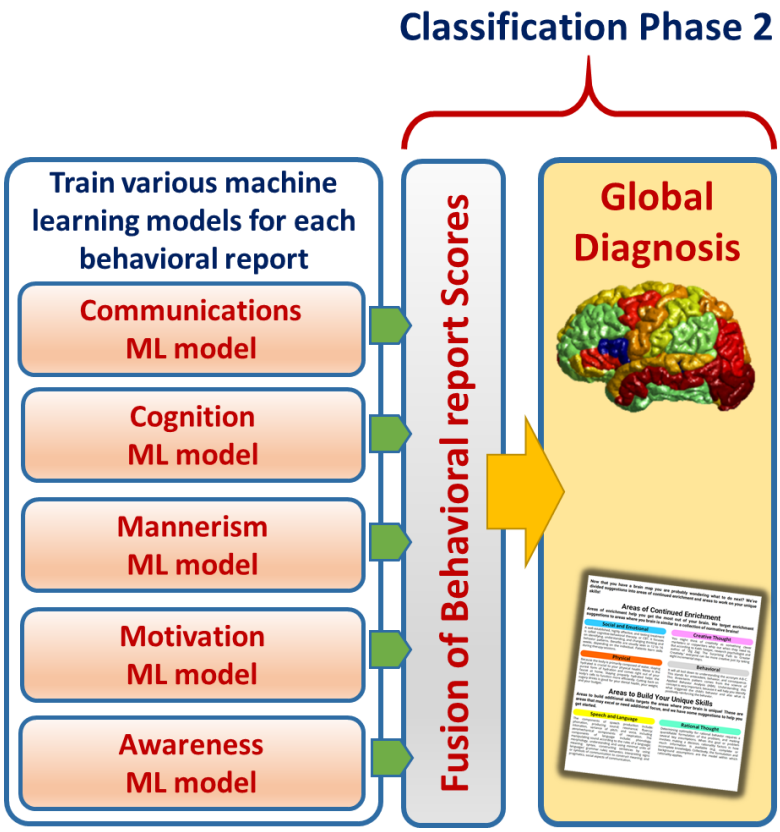


Figure 24. The pipeline of the proposed Phase 2 method of Classification

models with average cross validation balanced accuracy score, testing balanced accuracy score, and testing f1-score. The rationale behind studying the aforementioned relationships is to provide a complementary information to the statistical significance of every neuro-atlas. Substantially, with every neuro-atlas, we provide the information of how consistent and accurate a machine learning model would be if it is trained using a given neuro-atlas.

Phase II

The framework is designed to mimic the behavioral interview conducted between a physician and subject. More or less, we designed phase I to produce an output that looks a bit similar to the behavioral reports with all the scores associated with different behavioral traits, similar to the behavioral reports that can be downloaded via [154]. Consequently, we think of phase II as the aggregation of those behavioral scores to provide a final diagnosis. Therefore, the performance of phase II is totally dependant on the output of phase I. It is worth noting that the classification problem of phase II has totally different labels (target vector) that the classification problem of phase I. Although, they might seem dependent somehow, since the final diagnosis of subject is actually dependent on that subject's behavior scores, in terms of classification problems, the final diagnosis target vector has never been included in phase I.

Similar to phase I, phase II comprises the ML classifiers discussed in section 2.3.5 (Fig. 24). Nevertheless, those classifiers learn to classify a subject into ASD or TD instead of severity levels as it is the case of phase I. Moreover, the input of phase II is no longer cortical morphological features. Instead, it is a matrix of probabilities X of size $N \times 18$. Each row of X denotes a subject, and each column of X denotes a probability that this subject posses a specific severity level of a specific behavior trait. For instance, the first three columns of matrix X at row i denote the probability of subject i to be TD or mild ASD (0 means certainly TD, and 1 means certainly mild ASD), TD or moderate ASD, and TD or sever ASD with respect to communication module of SRS respectively. Similarly for the following three columns except that they denote the probabilities for mannerism module of SRS, and so on.

$$X = \begin{bmatrix} pp_{1,1} & pp_{1,2} & \cdots & pp_{1,18} \\ pp_{2,1} & pp_{2,2} & \cdots & pp_{2,18} \\ \vdots & \vdots & \ddots & \vdots \\ pp_{M,1} & pp_{M,2} & \cdots & pp_{M,18} \end{bmatrix}, \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_M \end{bmatrix}, \tag{8}$$

Equation 8 shows the input feature matrix X and the target vector y given to the classifiers in phase II. The target vector y is the final diagnosis provided by ABIDE II for every subject.

The training process of phase II classifiers is similar to the training process took place for every behavioral group. A set of six different ML classifiers representing both linear and non-linear hypotheses is selected to test the behavioral imaging markers candidates. The utilized six ML classifiers are split into two main divisions: (i) Linear classifiers, and (ii) Non-linear classifiers. The linear classifiers set comprises LR with l1 and l2 regularization. The non-linear classifiers set comprises RF, SVM-RBF, LGBM, nn, and ExtraGradient Boost Trees (xgboost) [155]. To optimize the hyper-parameters of each classifier, the data is split into 5-folds. Four folds of data are utilized for training each classifier with a randomly sampled hyper parameters, and the performance is evaluated using the 5^{th} fold. the classifier with the highest balanced accuracy score, Bagging the classifier of the highest performance, and stacking all the aforementioned trained classifiers to find the model that would be used as the final classification layer. Those three models are then evaluated on a 20% hold-out set to evaluate the generalizability of the framework.

Eventually, we randomly select one ASD subject, and one TD subject from the hold-out 20% dataset, as our case studies. We utilize the top performing classifiers of phase one to classify the severity of the behavior trait of each subject, then feeding the output probability matrix to phase two classifier in order to provide us with the final diagnosis of each subject. To add more interpretability to the results, we utilize local interpretable model-agnostic explanations (LIME) [122] which is a novel explanation technique that explains the prediction of any classifier. The main idea behind LIME is that it builds linear models around the predictions of an opaque model to explain it.

3 Results

In this section, we will cover the output of phase I and phase II of the framework. We focus on justifying the results of neuro-atlases since this is the core of the entire framework, as well as, we are hoping that those neuro-atlases would help more scientists to understand ASD etiology. Moreover, we will focus on two case studies of the hold-out set to demonstrate the process of the entire framework in a production-like environment.

Phase I

In this section, we will demonstrate the morphological features selected for each behavioral neuro-atlas within each severity group. Also, we will provide the summary statistics of the top performing experiments, and the morphological features, brain regions, and hemispheres selection frequency. Eventually, we will demonstrate the results of the selected models for each behavioral group within each severity group. Those selected models are going to build the input of Phase II as described in the previous Phase II section.

Neuro-atlases

In this section, we will go through the results which lead us to build multiple neuro atlases for each behavioral disorder. First, features are selected via the aforementioned RFECV algorithms. Figures 25 and 26 demonstrate the results of each of the RFECV algorithms to select the subset of features which the maximum classification accuracy between sever-cognition ASD and TD. A vertical line indicates the size of features' set corresponding to the maximum classification balanced accuracy.

The following figures demonstrate the neuro-atlas which defines each severity-behavioral group, as well as the aggregated atlases that defines the brain region which associate with each severity level across all behavioral groups, and each behavioral group across all severity levels. We chose to demonstrate only the statistically significant behavioral neuroatlases over all severity scores, while adding to the supplementary materials the severity neuro-atlases, and behavioral per severity neuro-atlases.

Figure 27 demonstrate the color code of each brain region defined by DK atlas. The color codes are the standard color codes defined by Freesurfer.

Figure 28 demonstrates the brain regions which posses a statistical significant $p < 0.001$ contribution to the classification between ASD subjects with mild awareness score, and TD subjects. The **left hemisphere brain regions** shown in figure 28 are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, entorhinal, fusiform, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, middletemporal, parahippocampal, parsopercularis, parstriangularis, pericalcarine, postcentral, precentral, precuneus, rostralanteriorcingulate, supramarginal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are caudalanteriorcingulate, caudalmiddlefrontal, cuneus, entorhinal,

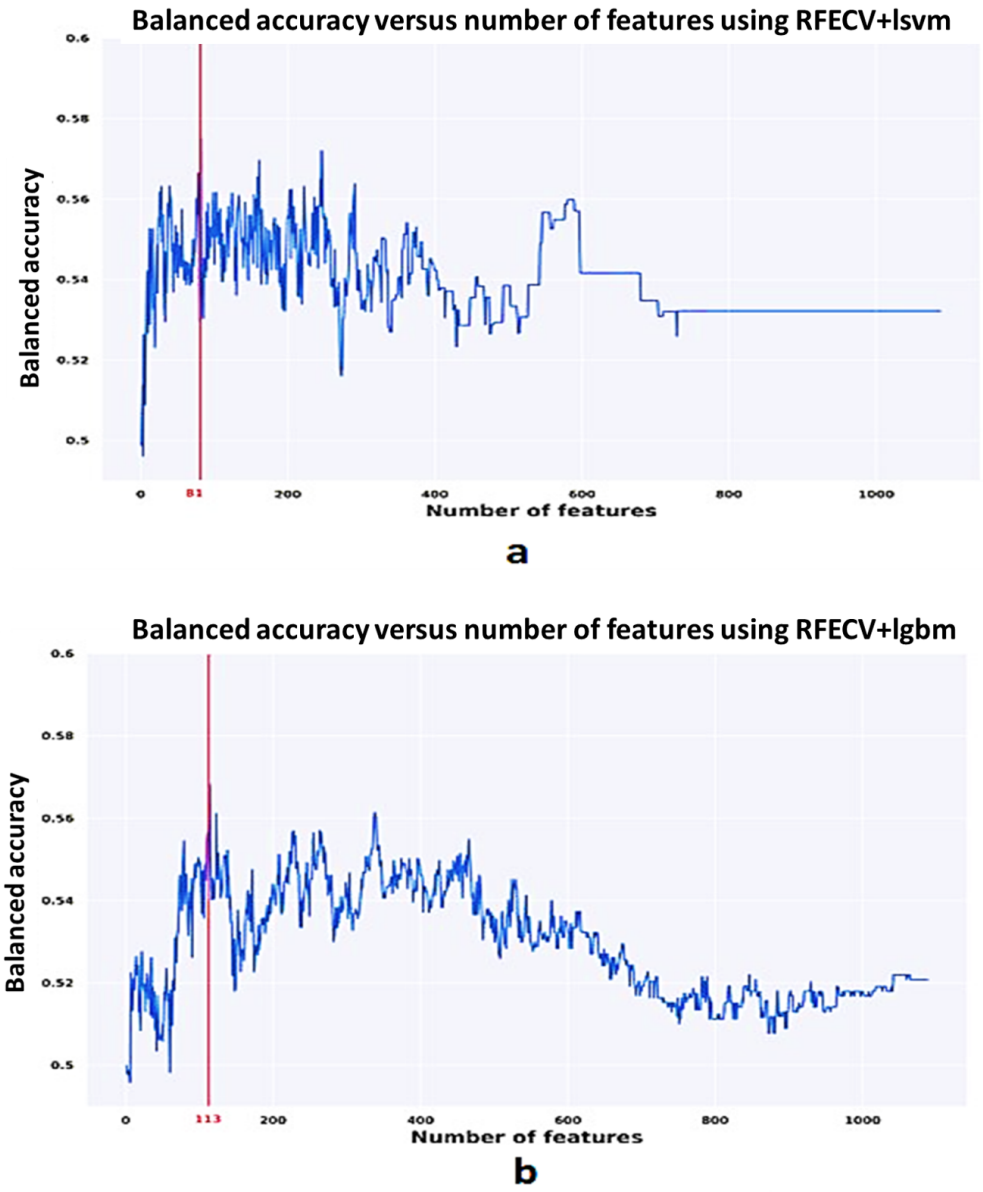


Figure 25. Balanced accuracy versus the number of selected features using each of RFECV classifiers. (a) lsvm, (b) lgbm.

fusiform, inferiorparietal, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, middletemporal, paracentral, parahippocampal, parsopercularis, parsorbitalis, parstriangularis, postcentral, posteriorcingulate, precuneus, rostralanteriorcingulate, superiorfrontal, superiorparietal, superiortemporal, supramarginal, temporalpole, transversetemporal.

Figure 29 demonstrates the brain regions which posses a statistical significant $p < 0.001$ contribution to the classification between ASD subjects with mild cognition score, and TD subjects. The **left hemisphere brain regions** shown in figure 29 are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, cuneus, entorhinal,

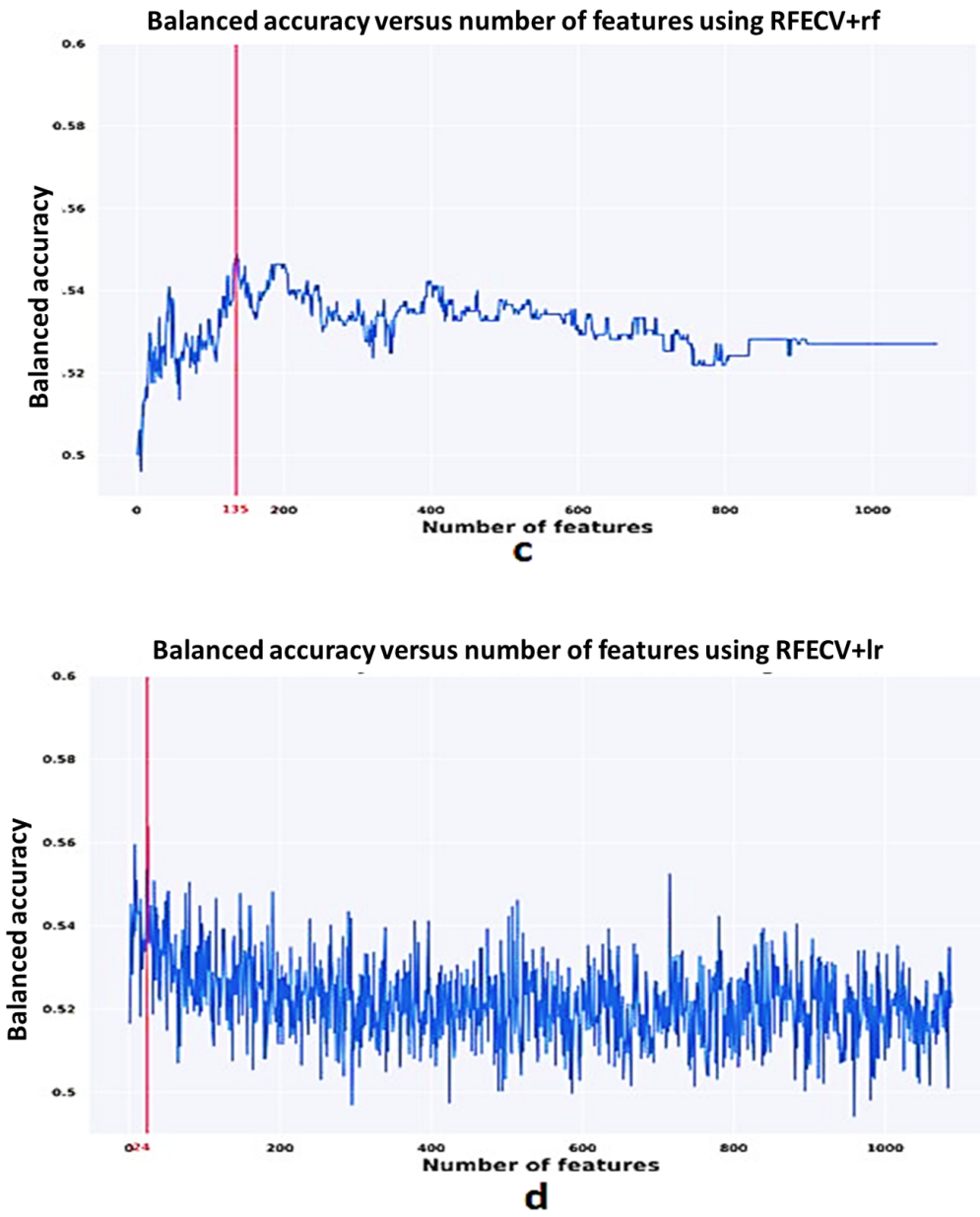


Figure 26. Balanced accuracy versus the number of selected features using each of RFECV classifiers. (c) rf, and (d) lr.

frontalpole, fusiform, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, middletemporal, parahippocampal, parsopercularis, parstriangularis, pericalcarine, postcentral, posteriorcingulate, precentral, precuneus, rostralmiddlefrontal, superior-parietal, superiortemporal, supramarginal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, cuneus, entorhinal, frontalpole, fusiform, inferiorparietal, inferiortemporal, lateralorbitofrontal, lingual, medialorbitofrontal, paracentral, parahippocampal, parsopercularis, parstriangularis, postcentral, posteriorcingulate, precen-

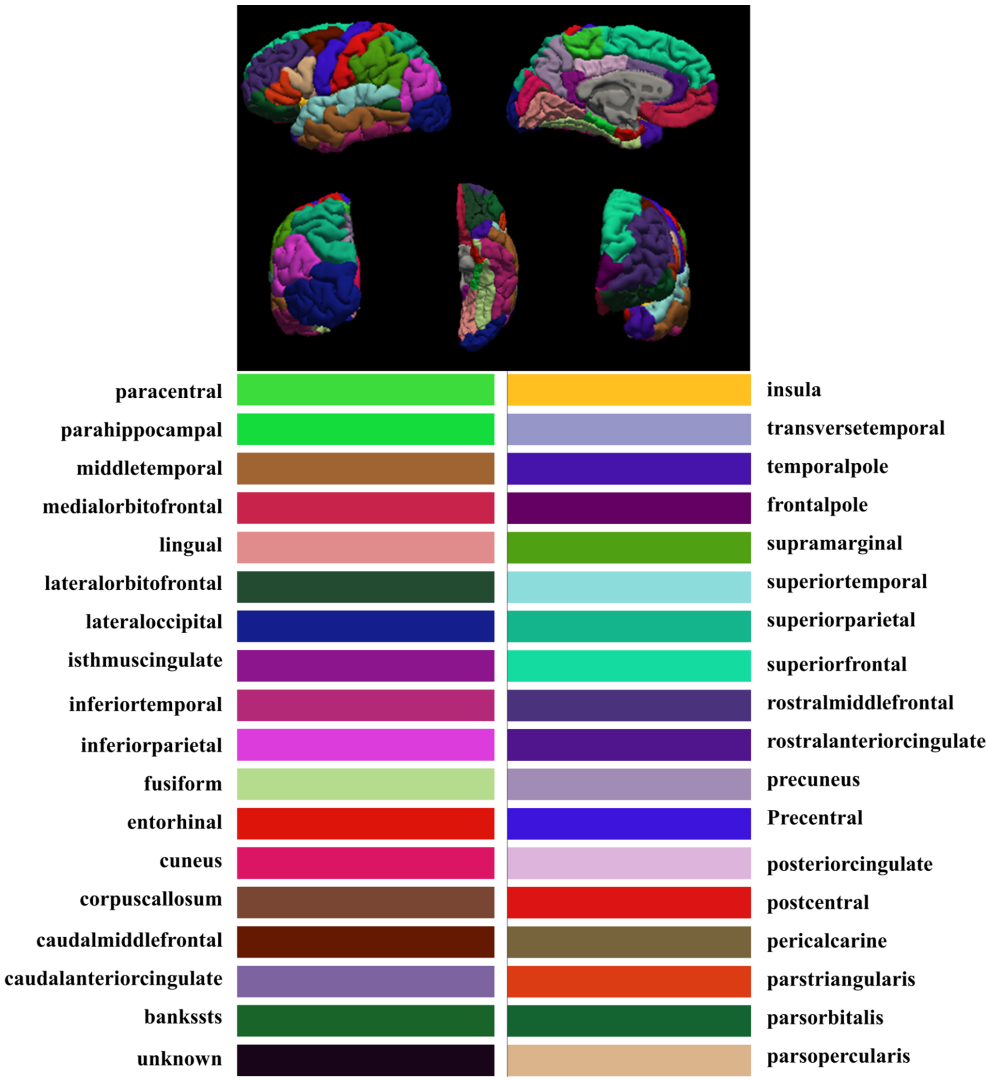


Figure 27. The color-coded brain areas.

tral, precuneus, rostralanteriorcingulate, superiorfrontal, superiorparietal, superiortemporal, supramarginal, temporalpole, transversetemporal.

Figure 30 demonstrates the brain regions which posses a statistical significant $p < 0.001$ contribution to the classification between ASD subjects with mild communication score, and TD subjects. The **left hemisphere brain regions** shown in figure 30 are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, entorhinal, fusiform, inferiortemporal, insula, isthmuscingulate, lateralorbitofrontal, lingual, medialorbitofrontal, paracentral, parahippocampal, parstriangularis, pericalcarine, posteriorcingulate, rostralanteriorcingulate, rostralmiddlefrontal, superiorfrontal, superiortemporal, supramarginal, transversetemporal. While the **right hemisphere brain regions** are bankssts, caudalanteriorcingulate, cuneus, frontalpole, fusiform, inferiorte mporal, insula, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, parsopercularis, parsorbitalis, parstriangularis, postcentral, posteriorcingulate, precuneus, rostralmiddlefrontal, superiortemporal, temporalpole, transversetemporal.

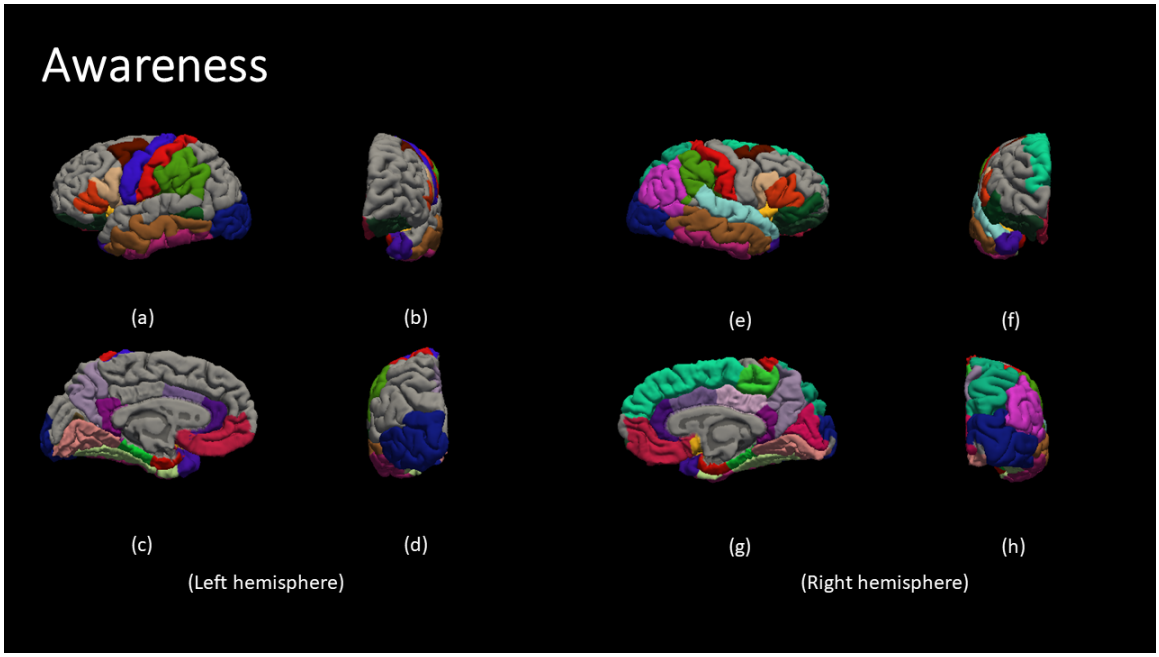


Figure 28. The results for Mild/TD group: Awareness Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

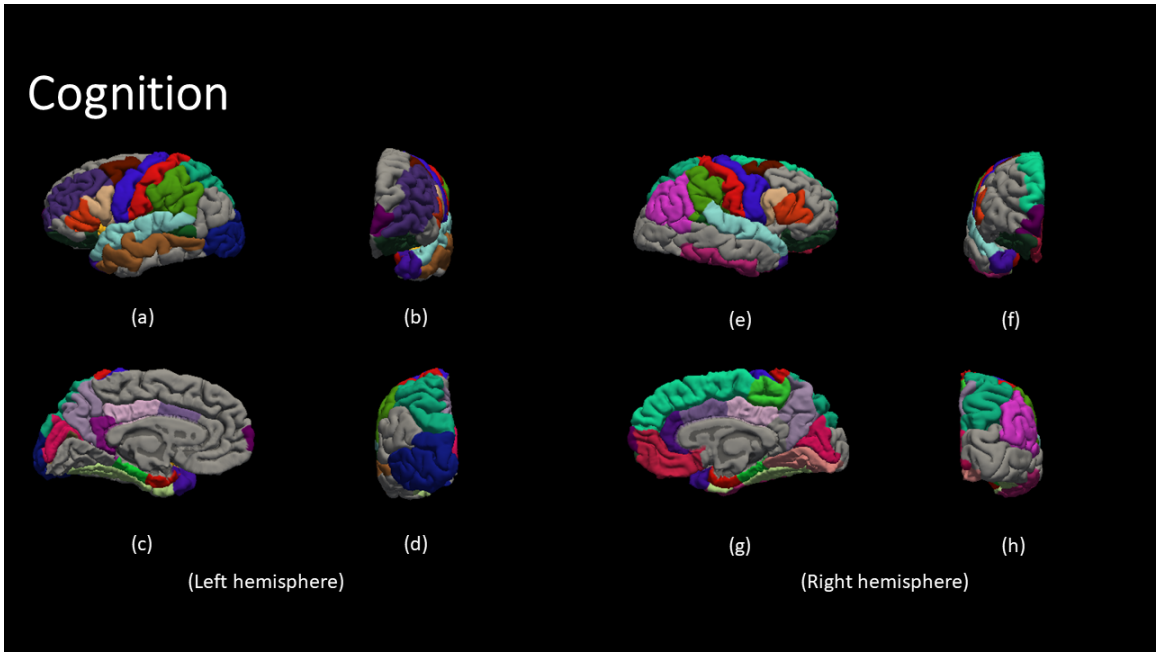


Figure 29. The results for Mild/TD group: Cognition Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

Figure 31 demonstrates the brain regions which posses a statistical significant $p < 0.001$ contribution to the classification between ASD subjects with mild mannerism score, and TD subjects. The **left hemisphere brain regions** shown in figure 31 are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, cuneus, frontalpole,

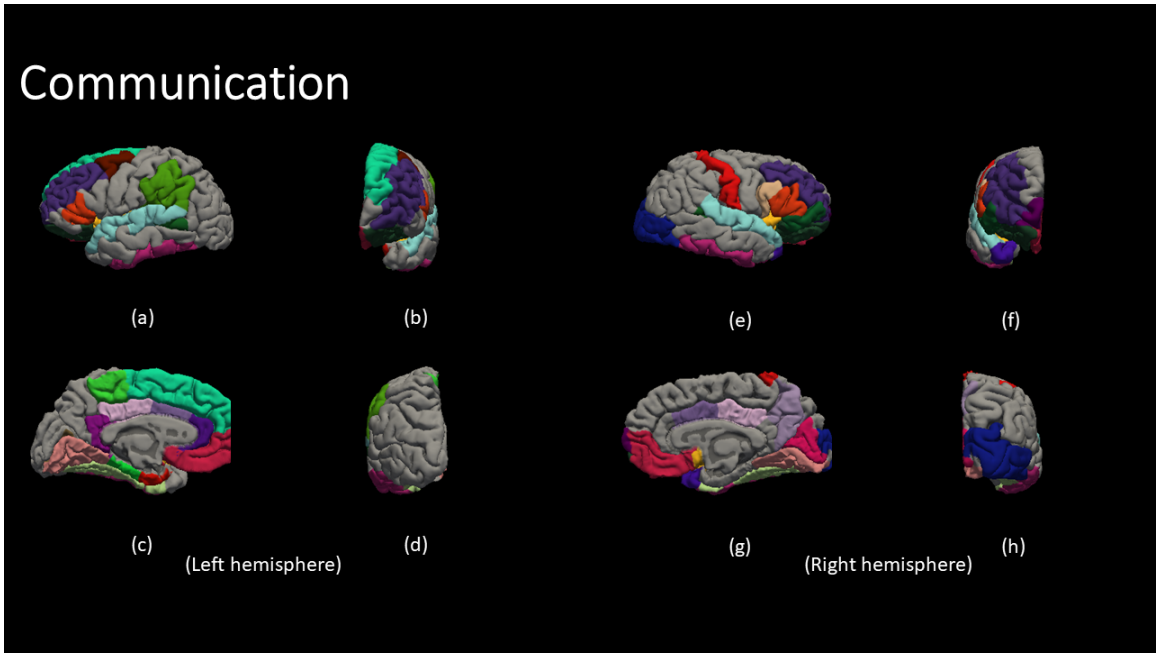


Figure 30. The results for Mild/TD group: Communication Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

inferiorparietal, insula, lateraloccipital, lateralorbitofrontal, lingual, middletemporal, paracentral, parsorbitalis, parstriangularis, pericalcarine, posteriorcingulate, precuneus, rostralanteriorcingulate, superiorfrontal, superiorparietal, superiortemporal, supramarginal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are bankssts, entorhinal, frontalpole, inferiorparietal, inferiortemporal, isthmuscingulate, lateralorbitofrontal, lingual, medialorbitofrontal, middletemporal, posteriorcingulate, precentral, superiorfrontal, superiortemporal, supramarginal, temporalpole, transversetemporal.

Figure 32 demonstrates the brain regions which posses a statistical significant $p < 0.001$ contribution to the classification between ASD subjects with mild motivation score, and TD subjects. The **left hemisphere brain regions** shown in figure 32 are caudalmiddlefrontal, cuneus, entorhinal, fusiform, inferiorparietal, inferiortemporal, isthmuscingulate, lateralorbitofrontal, lingual, medialorbitofrontal, middletemporal, paracentral, parahippocampal, parsopercularis, parsorbitalis, pericalcarine, postcentral, posteriorcingulate, precuneus, rostralanteriorcingulate, rostralmiddlefrontal, superiorfrontal, superiortemporal, transversetemporal. While the **right hemisphere brain regions** are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, cuneus, frontalpole, inferiorparietal, inferiortemporal, insula, lingual, medialorbitofrontal, middletemporal, paracentral, parsopercularis, parsorbitalis, parstriangularis, pericalcarine, postcentral, posteriorcingulate, precentral, precuneus, rostralanteriorcingulate, superiorfrontal, superiorparietal, supramarginal, temporalpole, transversetemporal.

Figure 33 demonstrates the brain regions which posses a statistical significant $p < 0.001$ contribution to the classification between ASD subjects with moderate aware-

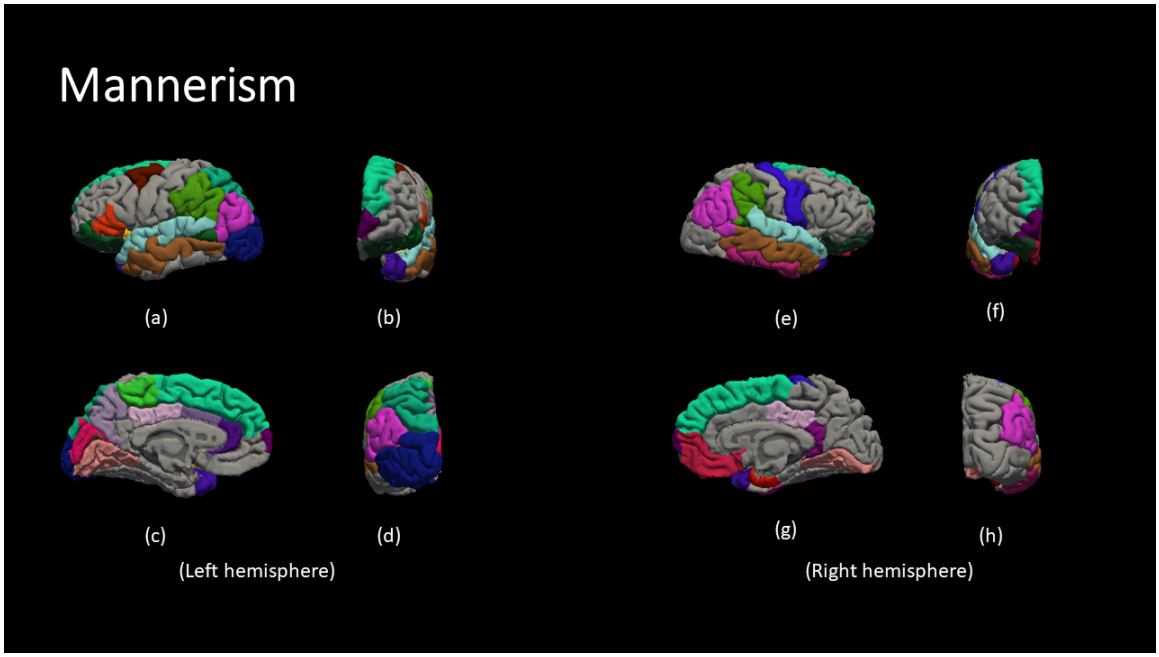


Figure 31. The results for Mild/TD group: Mannerism Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

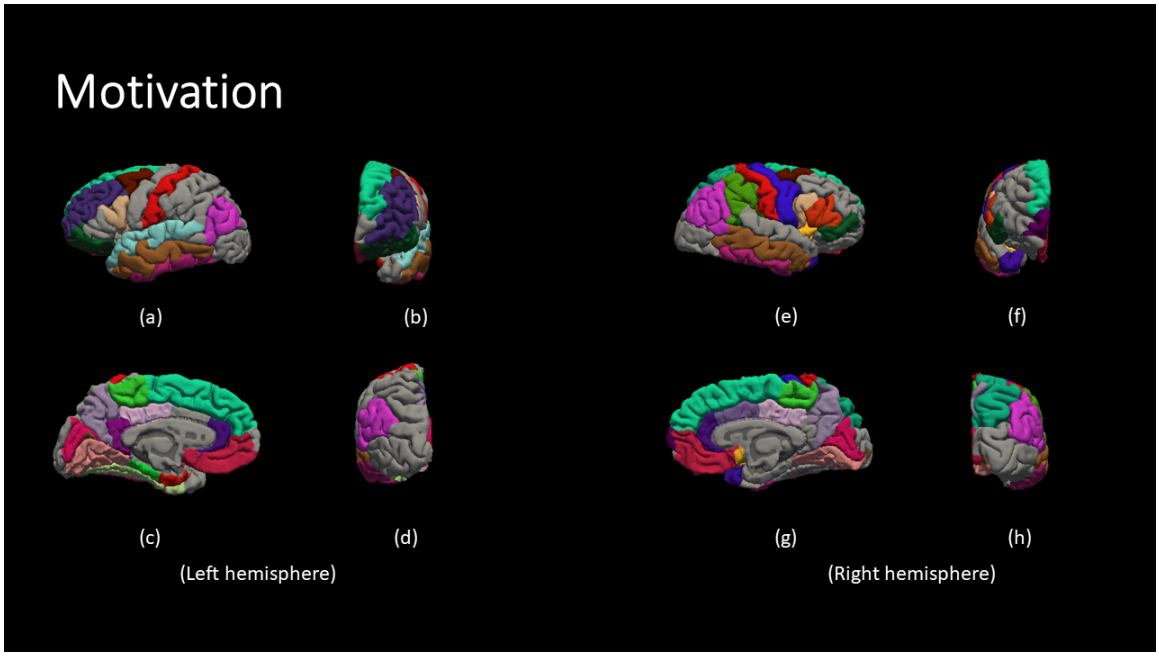


Figure 32. The results for Mild/TD group: Motivation Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

ness score, and TD subjects. The **left hemisphere brain regions** shown in figure 33 are bankssts, caudalmiddlefrontal, cuneus, frontalpole, fusiform, inferiorparietal, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, middletemporal, paracentral, parsorbitalis, parstriangularis, per-

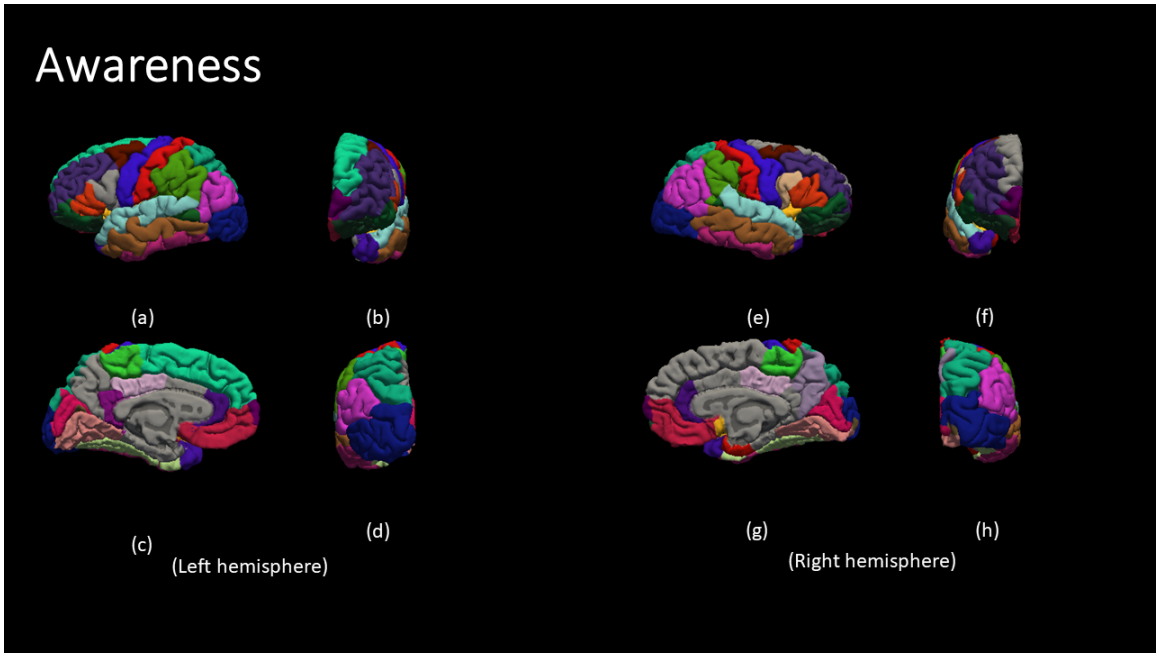


Figure 33. The results for Moderate group: Awareness Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

icalcarine, postcentral, posteriorcingulate, precentral, rostralanteriorcingulate, rostralmiddlefrontal, superiorfrontal, superiorparietal, superiortemporal, supramarginal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are bankssts, caudalmiddlefrontal, cuneus, entorhinal, frontalpole, fusiform, inferiorparietal, inferiortemporal, insula, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, middletemporal, paracentral, parsopercularis, parsorbitalis, parstriangularis, pericalcarine, postcentral, posteriorcingulate, precentral, precuneus, rostralanteriorcingulate, rostralmiddlefrontal, superiorparietal, superiortemporal, supramarginal, temporalpole, transversetemporal.

Figure 34 demonstrates the brain regions which posses a statistical significant $p < 0.001$ contribution to the classification between ASD subjects with moderate cognition score, and TD subjects. The **left hemisphere brain regions** shown in figure 34 are caudalanteriorcingulate, caudalmiddlefrontal, frontalpole, fusiform, inferiorparietal, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, paracentral, parahippocampal, parsorbitalis, parstriangularis, posteriorcingulate, precuneus, rostralmiddlefrontal, superiorfrontal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are caudalmiddlefrontal, cuneus, entorhinal, fusiform, inferiorparietal, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, paracentral, parahippocampal, parsopercularis, parstriangularis, pericalcarine, posteriorcingulate, precuneus, rostralanteriorcingulate, rostralmiddlefrontal, superiorfrontal, superiorparietal, superiortemporal, supramarginal, transversetemporal.

Figure 35 demonstrates the brain regions which posses a statistical significant

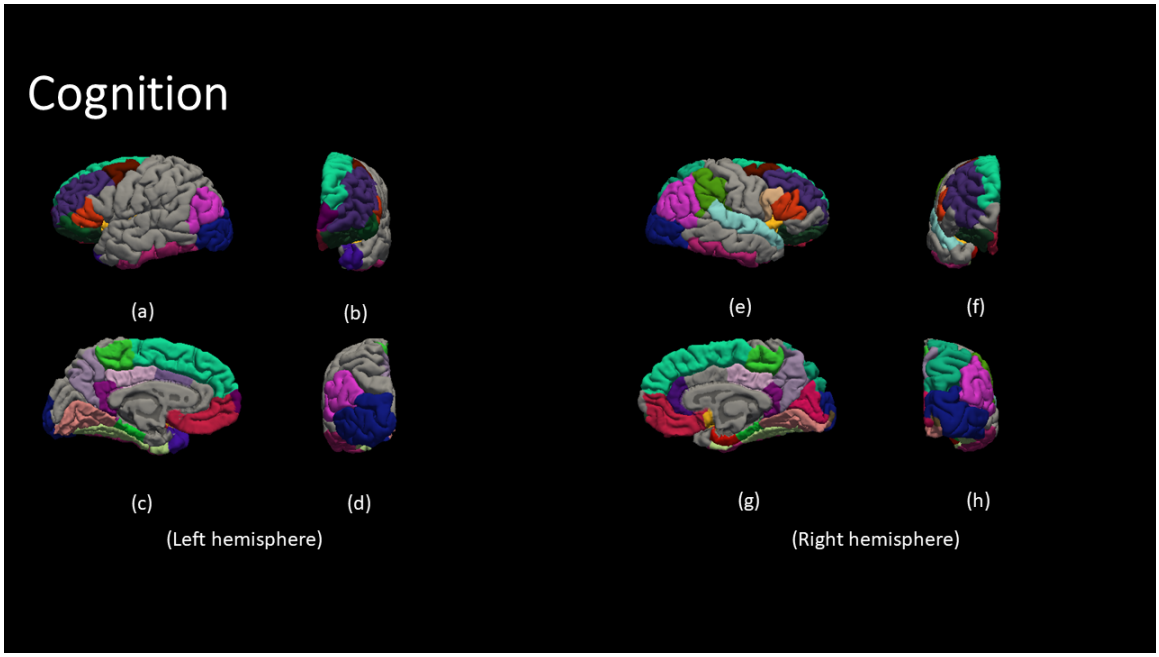


Figure 34. The results for Moderate group: Cognition Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

$p < 0.001$ contribution to the classification between ASD subjects with moderate communication score, and TD subjects. The **left hemisphere brain regions** shown in figure 35 are caudalanteriorcingulate, caudalmiddlefrontal, frontalpole, fusiform, inferiorparietal, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, paracentral, parahippocampal, parsorbitalis, parstriangularis, posteriorcingulate, precuneus, rostralmiddlefrontal, superiorfrontal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are caudalmiddlefrontal, cuneus, entorhinal, fusiform, inferiorparietal, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, paracentral, parahippocampal, parsopercularis, parstriangularis, pericalcarine, posteriorcingulate, precuneus, rostralanteriorcingulate, rostralmiddlefrontal, superiorfrontal, superiorparietal, superiortemporal, supramarginal, transversetemporal.

Figure 36 demonstrates the brain regions which posses a statistical significant $p < 0.001$ contribution to the classification between ASD subjects with moderate mannerism score, and TD subjects. The **left hemisphere brain regions** shown in figure 36 are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, cuneus, entorhinal, frontalpole, fusiform, inferiorparietal, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, middletemporal, paracentral, parahippocampal, parsopercularis, parsorbitalis, postcentral, posteriorcingulate, precentral, precuneus, rostralanteriorcingulate, rostralmiddlefrontal, superiorfrontal, superiorparietal, superiortemporal, supramarginal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, cuneus, entorhinal, frontalpole, fusiform, inferior-

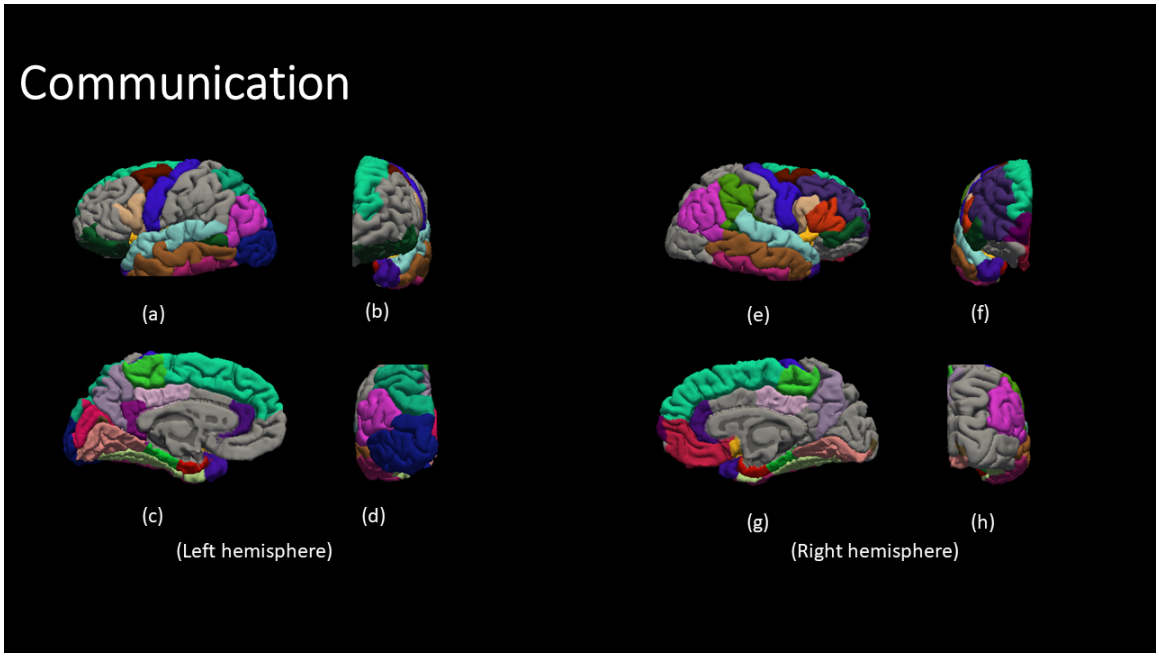


Figure 35. The results for Moderate group: Communication Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

parietal, inferiortemporal, insula, isthmuscingulate, lateralorbitofrontal, lingual, medialorbitofrontal, paracentral, parahippocampal, parsopercularis, parsorbitalis, parstriangularis, pericalcarine, postcentral, posteriorcingulate, precuneus, rostralanteriorcingulate, rostralmiddlefrontal, superiorfrontal, superiorparietal, superiortemporal, supramarginal, temporalpole, transversetemporal.

Figure 37 demonstrates the brain regions which posses a statistical significant $p < 0.001$ contribution to the classification between ASD subjects with moderate motivation score, and TD subjects. The **left hemisphere brain regions** shown in figure 37 are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, cuneus, entorhinal, frontalpole, fusiform, inferiorparietal, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, middletemporal, paracentral, parahippocampal, parsopercularis, parsorbitalis, parstriangularis, pericalcarine, postcentral, posteriorcingulate, rostralanteriorcingulate, rostralmiddlefrontal, superiorfrontal, superiorparietal, superiortemporal, supramarginal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are banks sts, caudalanteriorcingulate, caudalmiddlefrontal, cuneus, entorhinal, frontalpole, inferiorparietal, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, middletemporal, paracentral, parsopercularis, parsorbitalis, parstriangularis, pericalcarine, postcentral, posteriorcingulate, precuneus, rostralanteriorcingulate, rostralmiddlefrontal, superiorfrontal, superiortemporal, supramarginal, temporalpole, transversetemporal.

Figure 38 demonstrates the brain regions which posses a statistical significant $p < 0.001$ contribution to the classification between ASD subjects with sever awareness score, and TD subjects. The **left hemisphere brain regions** shown in fig-

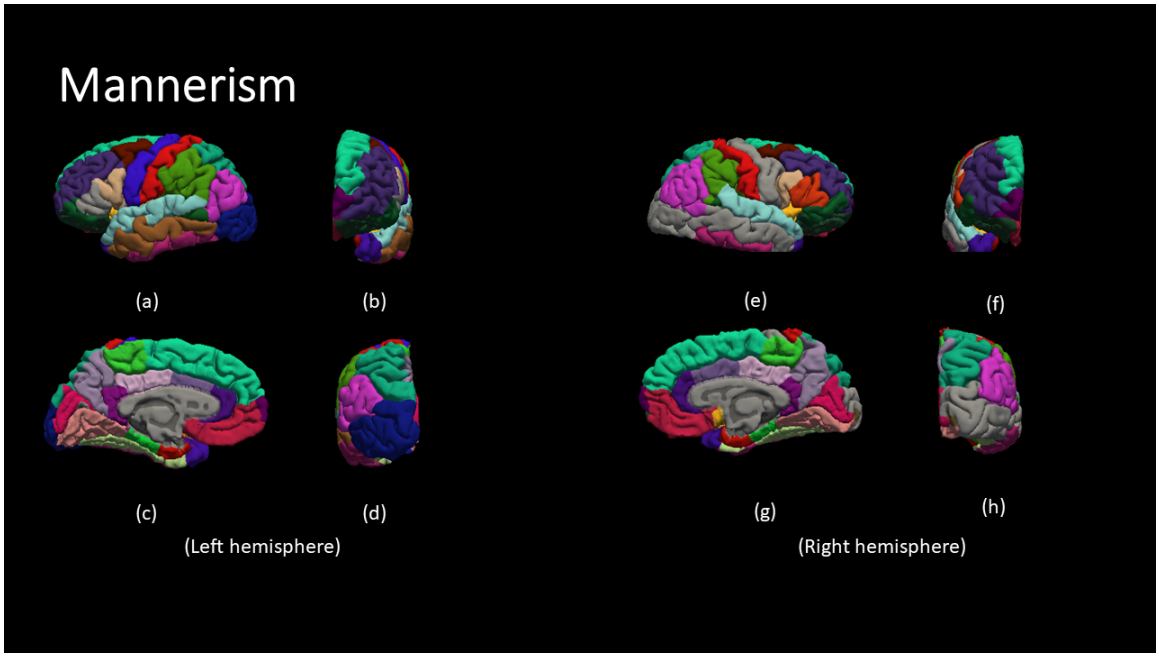


Figure 36. The results for Moderate group: Mannerism Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

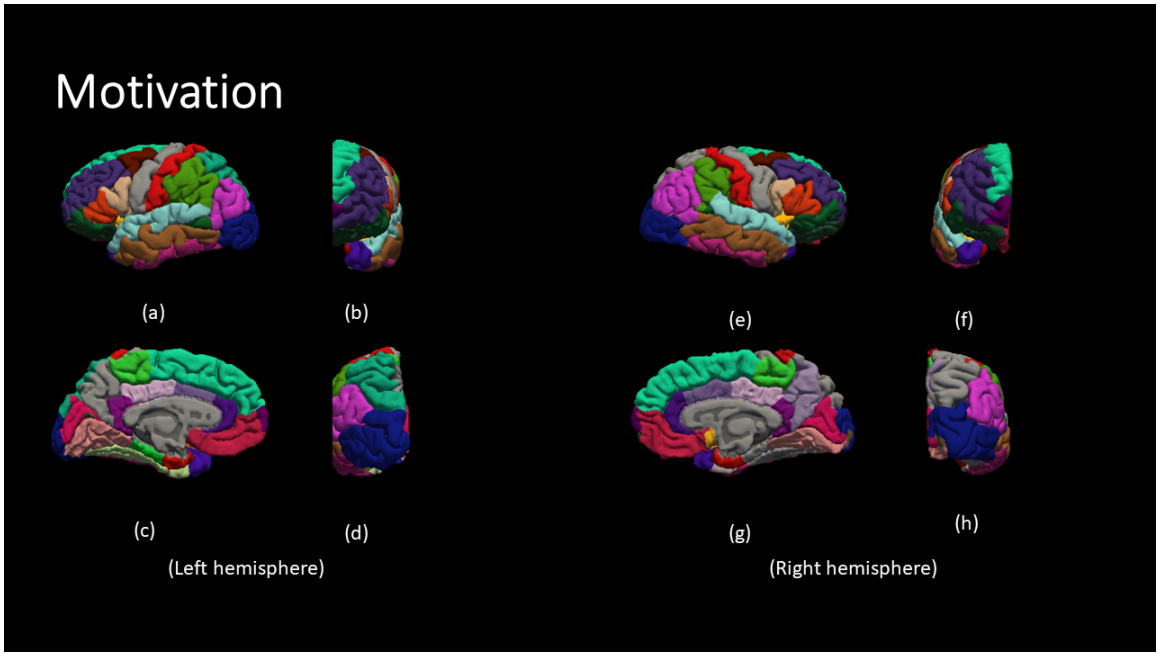


Figure 37. The results for Moderate group: Motivation Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

ure 38 are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, cuneus, entorhinal, frontalpole, fusiform, inferiorparietal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, middletemporal, paracentral, parahippocampal, parsopercularis, parsorbitalis, parstriangularis, pericalcarine, postcentral,

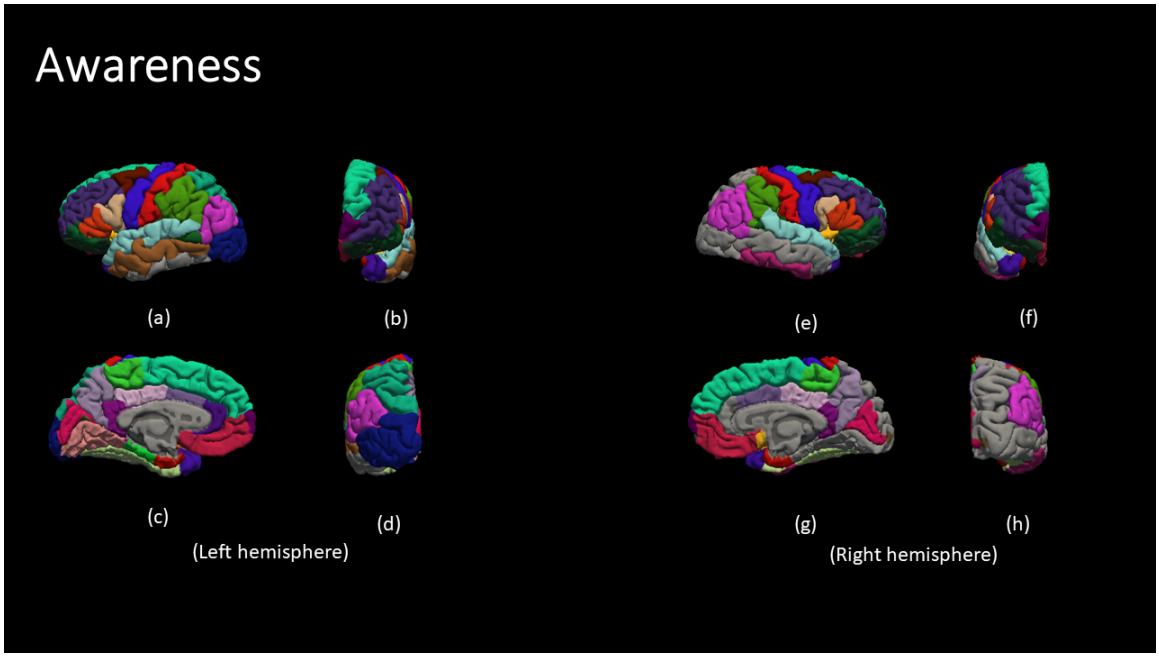


Figure 38. The results for Severe group: Awareness Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

posteriorcingulate, precentral, precuneus, rostralanteriorcingulate, rostralmiddlefrontal, superiorfrontal, superiorparietal, superiortemporal, supramarginal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, cuneus, entorhinal, frontalpole, fusiform, inferiorparietal, inferiortemporal, insula, isthmuscingulate, lateralorbitofrontal, medialorbitofrontal, paracentral, parsopercularis, parsorbitalis, parstriangularis, pericalcarine, postcentral, posteriorcingulate, precentral, precuneus, rostralmiddlefrontal, superiorfrontal, superiortemporal, supramarginal, temporalpole, transversetemporal.

Figure 39 demonstrates the brain regions which posses a statistical significant $p < 0.001$ contribution to the classification between ASD subjects with sever cognition score, and TD subjects. The **left hemisphere brain regions** shown in figure 39 are cuneus, frontalpole, fusiform, inferiorparietal, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, middletemporal, parahippocampal, parsopercularis, parsorbitalis, parstriangularis, postcentral, posteriorcingulate, precentral, precuneus, rostralanteriorcingulate, superiortemporal, supramarginal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, cuneus, entorhinal, frontalpole, fusiform, inferiortemporal, insula, isthmuscingulate, lateralor bitofrontal, lingual, middletemporal, parahippocampal, parsopercularis, parsorbitalis, parstriangularis, pericalcarine, posteriorcingulate, precentral, precuneus, rostralanteriorcingulate, rostralmiddlefrontal, superiorfrontal, superiorparietal, superiortemporal, supramarginal, temporalpole, transversetemporal.

Figure 40 demonstrates the brain regions which posses a statistical significant $p < 0.001$ contribution to the classification between ASD subjects with sever communica-

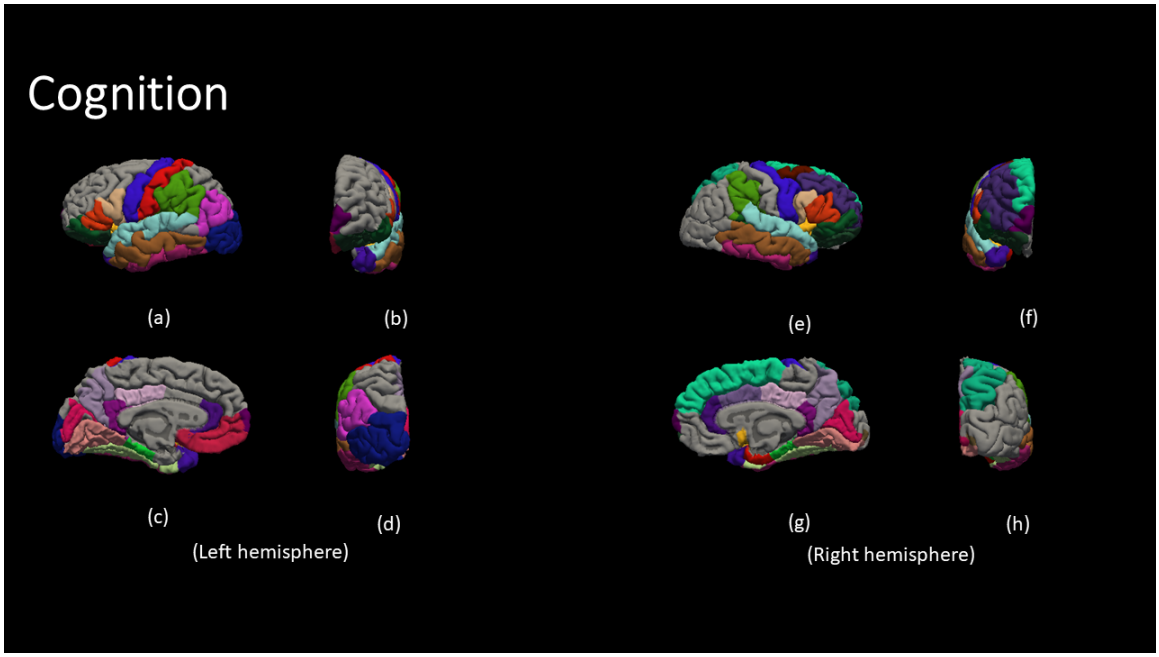


Figure 39. The results for Severe group: Cognition Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

tion score, and TD subjects. The **left hemisphere brain regions** shown in figure 40 are caudalanteriorcingulate, caudalmiddlefrontal, cuneus, entorhinal, frontalpole, inferiorparietal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, middletemporal, paracentral, parahippocampal, parsorbitalis, parstriangularis, postcentral, posteriorcingulate, precentral, superiorparietal, superiortemporal, supramarginal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are caudalanteriorcingulate, caudalmiddlefrontal, cuneus, fusiform, inferiorpari etal, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofron tal, lingual, medialorbitofrontal, middletemporal, paracentral, parahippocampal, parsopercularis, parstriangularis, pericalcarine, postcentral, precentral, precuneus, rostralmiddlefrontal, superiorfrontal, superiorparie tal, superiortemporal, supramarginal, temporalpole, transversetemporal.

Figure 41 demonstrates the brain regions which posses a statistical significant $p < 0.001$ contribution to the classification between ASD subjects with sever mannerism score, and TD subjects. The **left hemisphere brain regions** shown in figure 41 are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, cuneus, frontalpole, fusiform, inferiorparietal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, paracentral, parsopercularis, parsorbitalis, parstriangularis, pericalcarine, postcentral, posteriorcingulate, precentral, rostralanteriorcingulate, superiorfrontal, superiorparietal, superiortemporal, supramarginal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are caudalmiddlefrontal, cuneus, entorhinal, frontalpole, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, middletemporal, pa racentral, parahippocampal, parsopercularis, parstriangularis, pericalcarine, postcen-

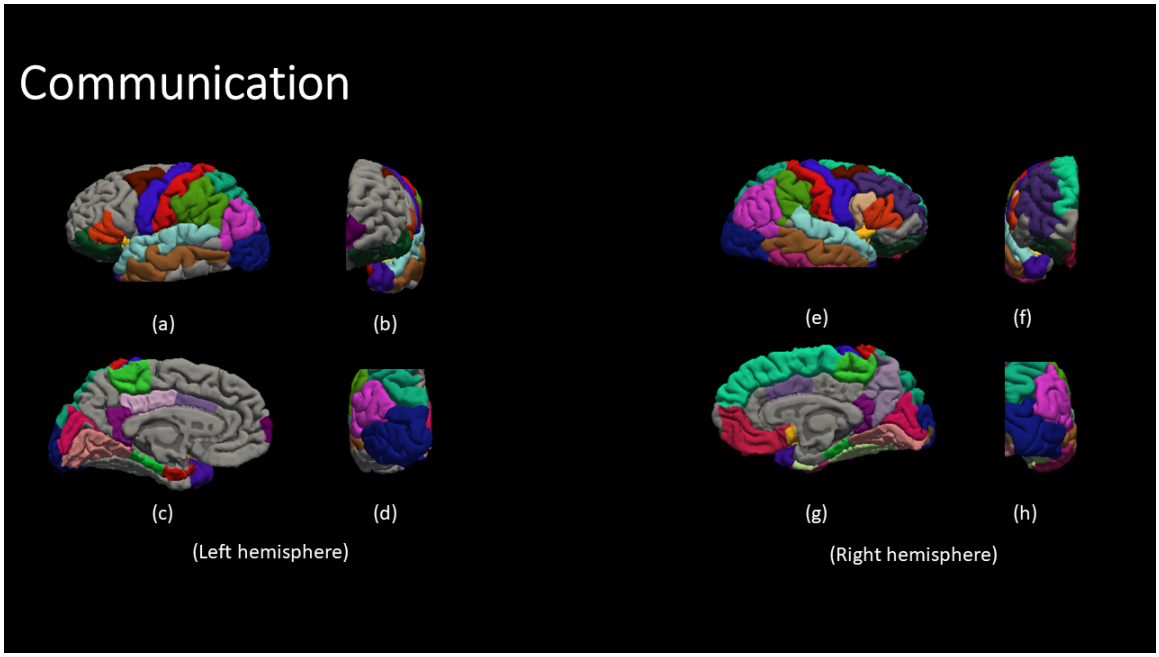


Figure 40. The results for Severe group: Communication Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

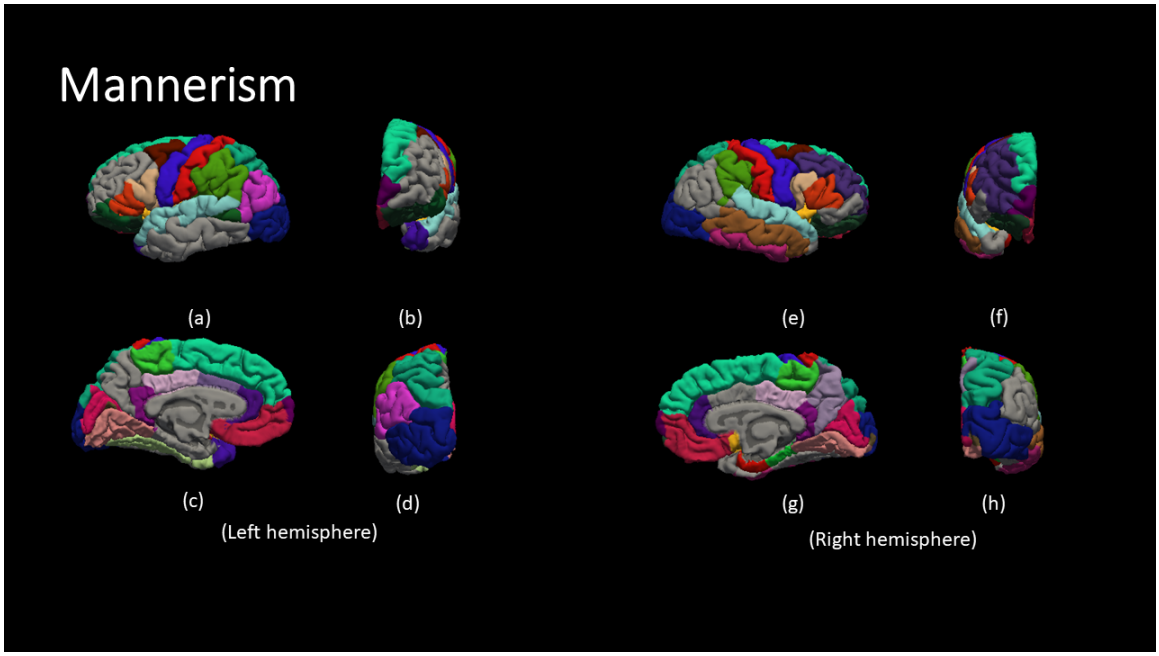


Figure 41. The results for Severe group: Mannerism Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

tral, posteriorcingulate, precentral, precuneus, rostralanteriorcingulate, rostralmiddlefrontal, superiorfrontal, superiorparietal, superiortemporal, supramarginal, transversetemporal.

Figure 42 demonstrates the brain regions which posses a statistical significant

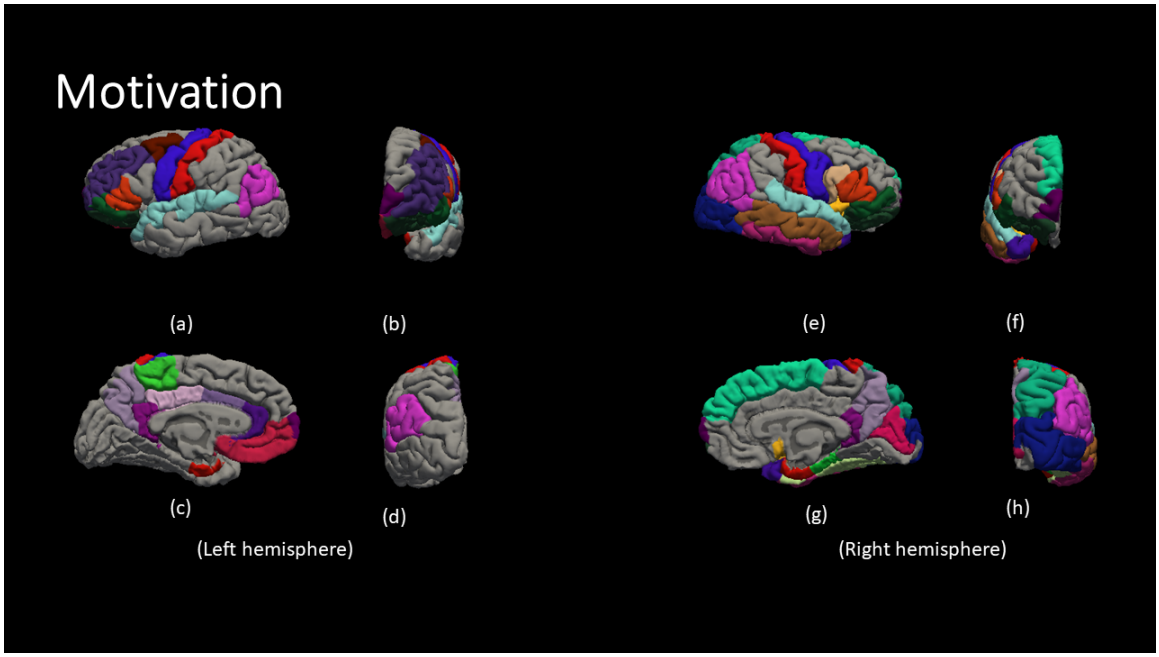


Figure 42. The results for Severe group: Motivation Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

$p < 0.001$ contribution to the classification between ASD subjects with sever motivation score, and TD subjects. The **left hemisphere brain regions** shown in figure 42 are caudalanteriorcingulate, caudalmiddlefrontal, entorhinal, frontalpole, inferiorparietal, isthmuscingulate, lateralorbitofrontal, medialorbitofrontal, paracentral, parsorbitalis, parstriangularis, postcentral, posteriorcingulate, precentral, precuneus, rostralanteriorcingulate, rostralmiddlefrontal, superiortemporal, transversetemporal. While the **right hemisphere brain regions** are cuneus, entorhinal, frontalpole, fusiform, inferiorparietal, inferiortemporal, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, middletemporal, parahippocampal, parsopercularis, parsorbitalis, parstriangularis, postcentral, precentral, precuneus, superiorfrontal, superiorparietal, superiortemporal, temporalpole, transversetemporal.

Figure 43 demonstrates the most frequent brain regions over all severity levels for the communication behavior report. This set of features are selected from the from the aforementioned neuroatlases. This neuro-atlas is thought of as the neurocharacterization of the communication attribute in the SRS report. The **left hemisphere brain regions** shown in figure 43 are caudalmiddlefrontal, entorhinal, insula, isthmuscingulate, lateralorbitofrontal, lingual, paracentral, parahippocampal, posteriorcingulate, superiortemporal, transversetemporal. While the **right hemisphere brain regions** are fusiform, inferiortemporal, insula, lingual, medialorbitofrontal, parsopercularis, parstriangularis, precuneus, rostralmiddlefrontal, superiortemporal, temporalpole, transversetemporal.

Figure 44 demonstrates the most frequent brain regions over all severity levels for the cognition behavior report. This set of features are selected from the from the aforementioned neuroatlases. This neuro-atlas is thought of as the neuro-

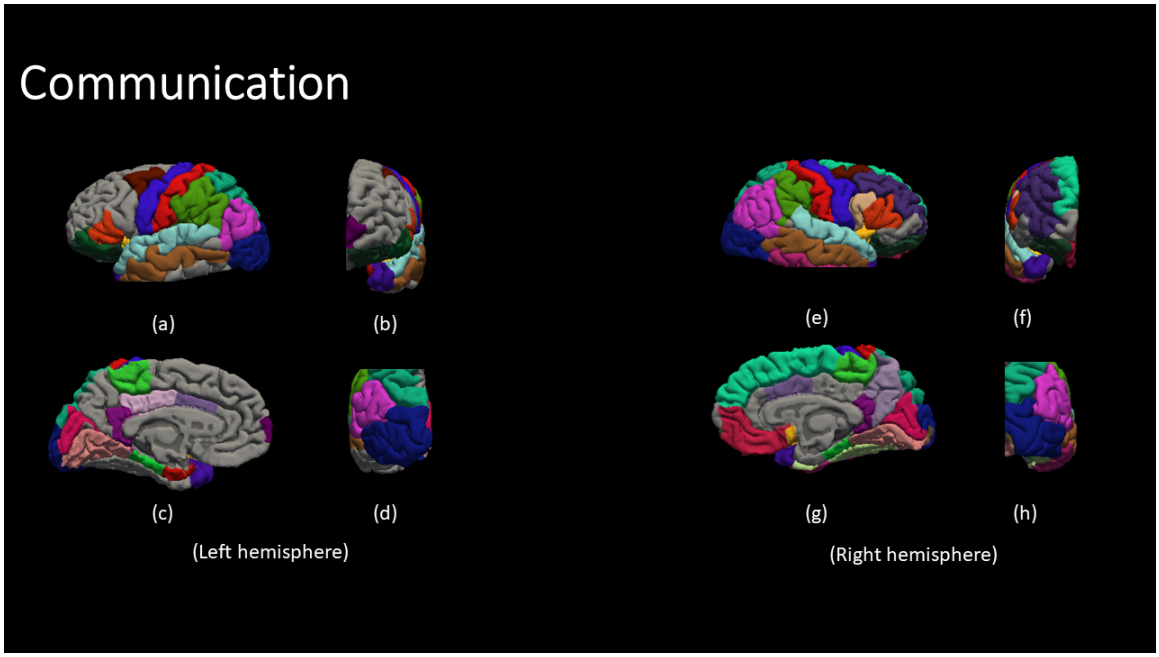


Figure 43. The results for most frequent brain region over all Severe group for the communication behavioral Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

characterization of the cognition attribute in the SRS report. The **left hemisphere brain regions** shown in figure 44 are frontalpole, fusiform, insula, isthmuscingulate, lateraloccipital, lateralorbitofrontal, parahippocampal, parstriangularis, posteriorcingulate, precuneus, temporalpole, transversetemporal. While the **right hemisphere brain regions** are caudalmiddlefrontal, cuneus, entorhinal, fusiform, inferiortemporal, lateralorbitofrontal, lingual, parahippocampal, parsopercularis, parstriangularis, posteriorcingulate, precuneus, rostralanteriorcingulate, superiorfrontal, superiorparietal, superiortemporal, supramarginal, transversetemporal.

Figure 45 demonstrates the most frequent brain regions over all severity levels for the mannerism behavior report. This set of features are selected from the from the aforementioned neuroatlases. This neuro-atlas is thought of as the neurocharacterization of the mannerism attribute in the SRS report. The **left hemisphere brain regions** shown in figure 45 are bankssts, caudalanteriorcingulate, caudalmiddlefrontal, cuneus, frontalpole, inferiorparietal, insula, lateraloccipital, lateralorbitofrontal, lingual, paracentral, parsorbitalis, posteriorcingulate, rostralanteriorcingulate, superiorfrontal, superiorparietal, superiortemporal, supramarginal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are entorhinal, frontalpole, inferiortemporal, isthmuscingulate, lateralorbitofrontal, lingual, medialorbitofrontal, posteriorcingulate, superiorfrontal, superiortemporal, supramar ginal, transversetemporal.

Figure 46 demonstrates the most frequent brain regions over all severity levels for the awareness behavior report. This set of features are selected from the from the aforementioned neuroatlases. This neuro-atlas is thought of as the neuro-

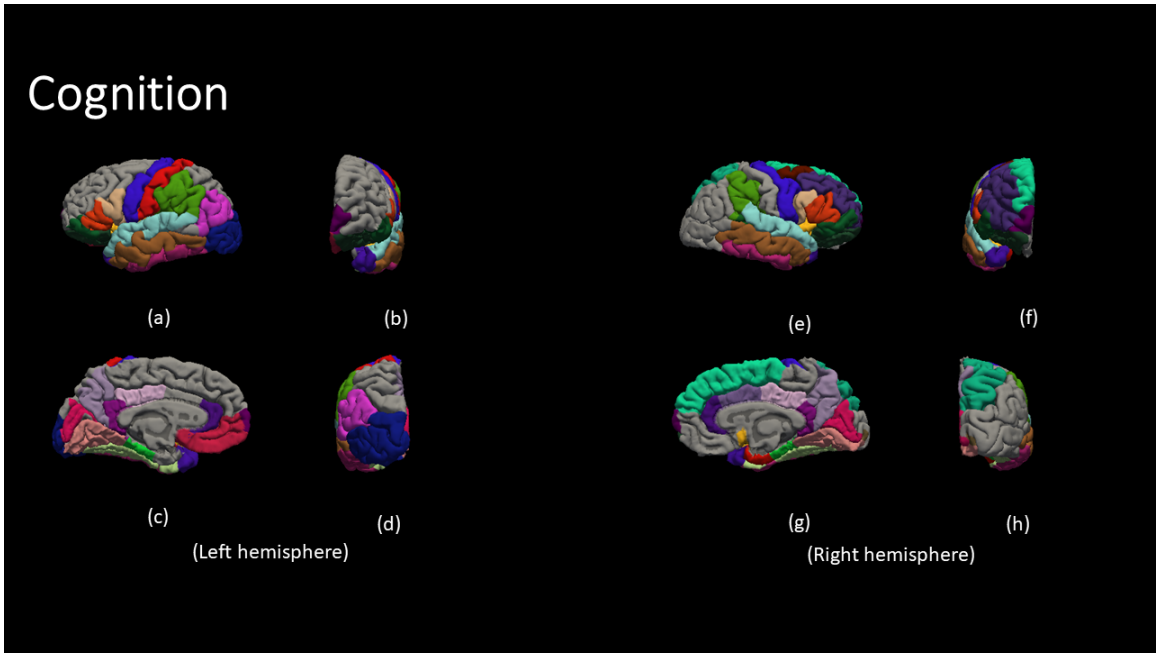


Figure 44. The results for most frequent brain region over all Severe group for the cognition behavioral Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

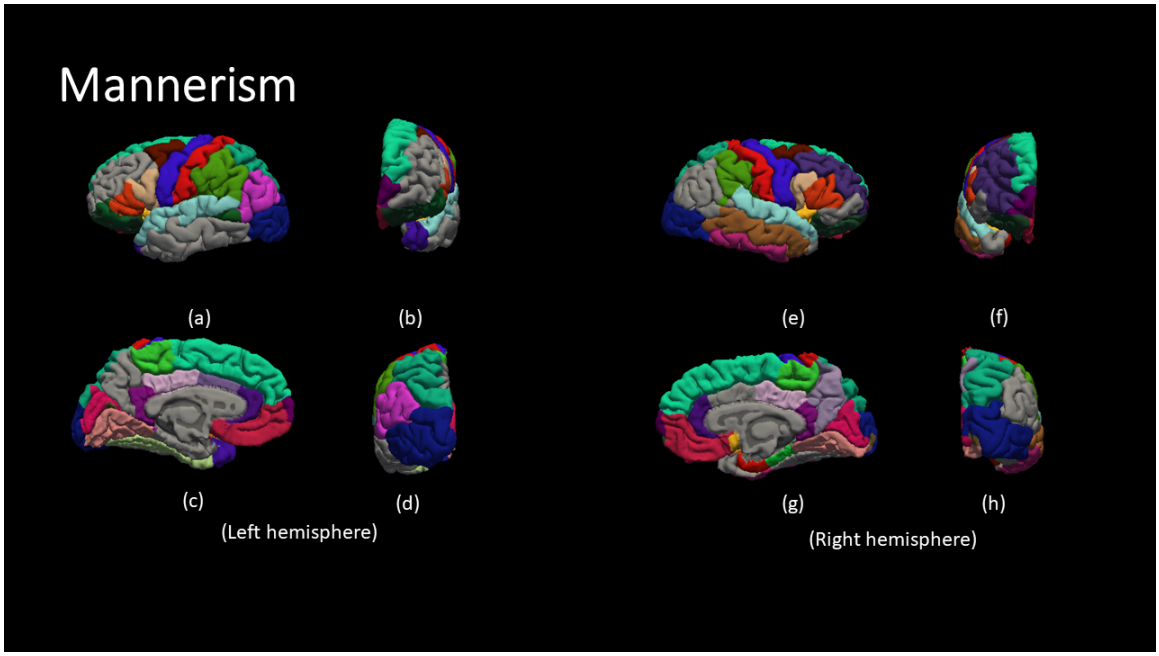


Figure 45. The results for most frequent brain region over all Severe group for the mannerism behavioral Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

characterization of the awareness attribute in the SRS report. The **left hemisphere brain regions** shown in figure 46 are bankssts, caudalmiddlefrontal, fusiform, insula,

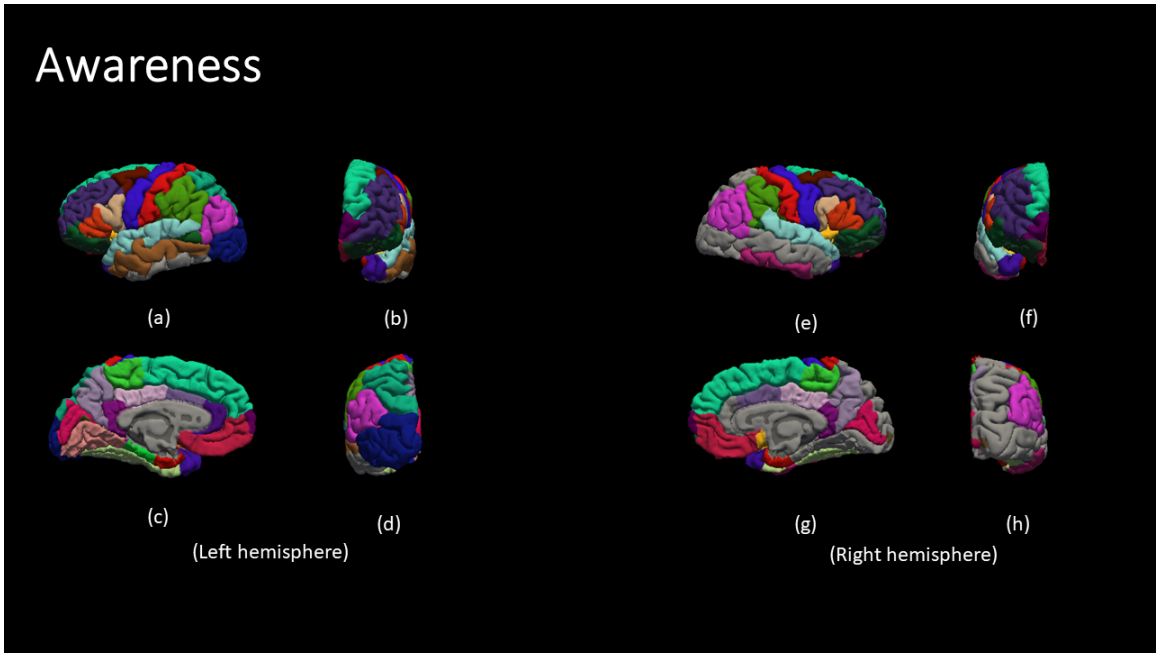


Figure 46. The results for most frequent brain region over all Severe group for the awareness behavioral Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

isthmuscingulate, lateraloccipital, lateralorbitofrontal, lingual, medialorbitofrontal, middletemporal, parstriangularis, pericalcarine, postcentral, precentral, rostralanteriorcingulate, supramarginal, temporalpole, transversetemporal. While the **right hemisphere brain regions** are caudalmiddlefrontal, cuneus, entorhinal, fusiform, inferiorparietal, inferiortemporal, insula, lateralorbitofrontal, medialorbitofrontal, par acentral, parsopercularis, parsorbitalis, parstriangularis, postcentral, posteriorcingulate, precuneus, superiortemporal, supramarginal, temporalpole, transversetemporal.

Figure 47 demonstrates the most frequent brain regions over all severity levels for the motivation behavior report. This set of features are selected from the from the aforementioned neuroatlases. This neuro-atlas is thought of as the neurocharacterization of the motivation attribute in the SRS report. The **left hemisphere brain regions** shown in figure 46 are caudalmiddlefrontal, entorhinal, inferiorparietal, isthmuscingulate, lateralorbitofrontal, medialorbitofrontal, paracentral, parsorbitalis, postcentral, posteriorcingulate, rostralanteriorcingulate, rostralmiddlefrontal, superiortemporal, transversetemporal. While the **right hemisphere brain regions** are cuneus, frontalpole, inferiorparietal, inferiortemporal, insula, middletemporal, parsopercularis, parsorbitalis, parstriangularis, postcentral, precuneus, superiorfrontal, temporalpole, transversetemporal.

ML classifiers

The results introduced in this section are for the experiments that yielded the maximum classification performance between TD, and each of the severity-behavioral

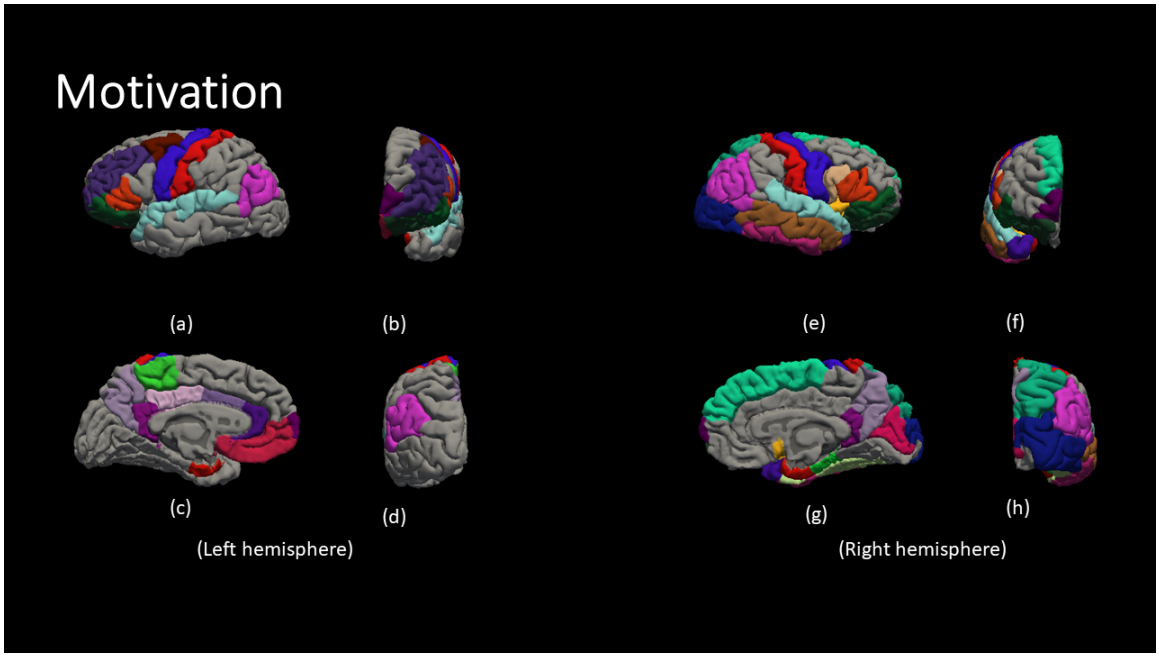


Figure 47. The results for most frequent brain region over all Severe group for the motivation behavioral Report. (a, e) lateral view, (b, f) anterior view, (c, g) medial view, (d, h) posterior view.

groups. Unlike the previous section, the models are solely based on their performance on the hold-out set without performing any statistical analysis. The selected features, which are used as the input, for each of the selected ML models are not introduced in this section. Since it is difficult to assume the generalizability of any single experiment, we did not weigh the input features to those models as much as the neuro-atlases which are selected based on statistical significance across all the 55 experiments.

Each of those models outputs a probability that a given subject belongs to a severity-behavioral group. Table 8 shows the classification accuracy of each of the selected ML models.

Table 9 shows the summary statistics of each severity group over all behavioral groups. The summary statistics are provided for the number of input features for each ML model, hold-out set balanced accuracy, and hold-out set F1-score.

Table 10 shows the summary statistics of each behavioral group over all severity groups. The summary statistics are provided for the number of input features for each ML model, hold-out set balanced accuracy, and hold-out set F1-score.

Phase II

In this section, we will demonstrate the the final output of the proposed framework. Phase II provides the final diagnosis/classification of each subject given the results of Phase I. Table 11 demonstrates the cross-validation results of training the six classifiers LR (l1-norm, l2-norm), xgboost, lgbm, NN, SVM, and RF using 80% of

Table 8. Phase I classification results

Severity	Behavior	RFECV	Classifier	# features	Bacc	F1
Mild	Awareness	lsvm	svm	183	0.98	0.92
	Communication	lsvm	svm	70	0.99	0.99
	Cognition	lsvm	lsvm	96	0.99	0.95
	Motivation	lsvm	svm	101	0.97	0.85
	Mannerism	lsvm	lsvm	137	0.99	0.99
	Total	lsvm	lsvm	130	0.99	0.99
Moderate	Awareness	lsvm	lsvm	240	0.92	0.8
	Communication	lsvm	lsvm	132	0.96	0.88
	Cognition	lsvm	lr	191	0.98	0.94
	Motivation	lsvm	lsvm	212	0.96	0.91
	Mannerism	lsvm	svm	184	0.98	0.94
	Total	lsvm	lsvm	167	0.97	0.91
Sever	Awareness	lsvm	lr	154	0.93	0.82
	Communication	lsvm	svm	241	0.91	0.84
	Cognition	lsvm	lsvm	254	0.95	0.90
	Motivation	lsvm	lsvm	166	0.96	0.91
	Mannerism	lsvm	lsvm	172	0.95	0.91
	Total	lsvm	lsvm	213	0.92	0.86

Table 9. Severity Classification Summary Statistics

Severity	**# Features** ($mean \pm std$)	**Bacc** ($mean \pm std$)	**F1** ($mean \pm std$)
Mild	119.5±39.4	0.99±0.01	0.95±0.05
Moderate	187.6±31.1	0.96±0.02	0.90±0.05
Sever	200±42	0.94±0.02	0.87±0.03

Table 10. Behavior Classification Summary Statistics

Behavior	**# Features** ($mean \pm std$)	**Bacc** ($mean \pm std$)	**F1** ($mean \pm std$)
Awareness	192.3±41.7	0.95±0.03	0.85±0.06
Communication	147.6±86.6	0.96±0.04	0.91±0.07
Cognition	180.33±79.5	0.97±0.01	0.93±0.03
Motivation	159.6±55.7	0.97±0.003	0.89±0.03
Mannerism	164.33±24	0.98±0.02	0.95±0.04
Total	170±41.5	0.96±0.04	0.93±0.06

Table 11. 5-fold cross validation results

Classifier	mean	std
LR	0.91	0.03
xgboost	0.86	0.03
SVM	0.94	0.03
NN	0.91	0.03
lgbm	0.91	0.04
RF	0.66	0.13

Table 12. Confusion matrix of the Bagging Classifier on the 20% hold-out data.

		Predicted	
		ASD	TD
Truth	ASD	54	4
	TD	1	86

Table 13. Confusion matrix of the Stacking Classifier on the 20% hold-out data.

		Predicted	
		ASD	TD
Truth	ASD	53	5
	TD	1	86

the dataset. It is worth mentioning that the input of each classifier is the probability matrix which is the output from phase I classification as shown in equation 8.

Trained SVM is then used to predict the final diagnosis of the hold-out set given their probability matrices. Tables 12, 13, 14 demonstrate the confusion matrix the bagging classifier built-up with SVM, initialized with the optimized hyperparameters as selected by the randomized grid-search cross-validation results, the stacking classifier built-up with all of six classifiers used in phase II, each is initialized with its optimized hyperparameters in the same way as SVM, and eventually SVM classifier. Each of those three classifiers is trained on the 80% dataset, and tested on the 20% hold-out set.

Table 14. Confusion matrix of the optimized hyperparameters SVM Classifier on the 20% hold-out data.

		Predicted	
		ASD	TD
Truth	ASD	54	4
	TD	1	86

Case Studies

In this section, we simulate a clinical environment by randomly selecting one TD subject, and one ASD out of the 20% hold-out set. The two selected subjects are processed by the proposed framework to output both the behavioral report and the final classification. We illustrate the output in the form of a proposed diagnosis report.

Case I

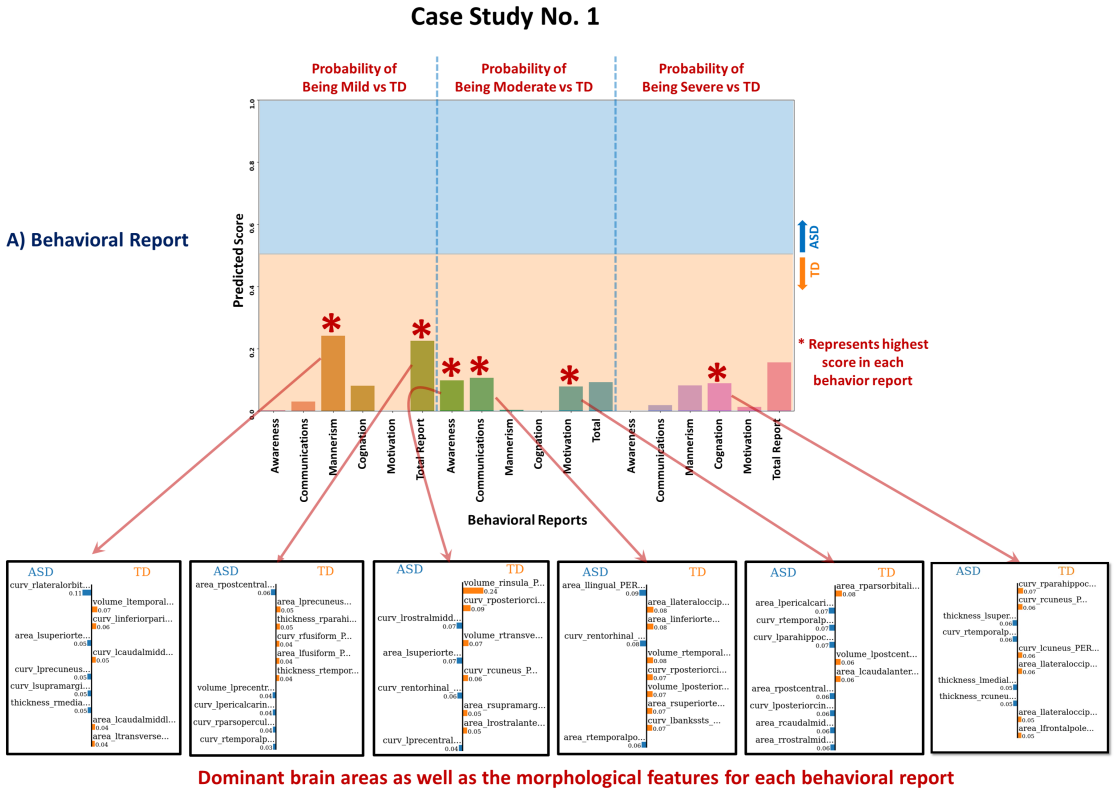
Case I is a TD subject from GU_1 site with id 28830. Based on the SRS_T scores of that subject, the case is diagnosed to be TD. Figure 48 demonstrates the proposed diagnostic report. Figure 48: A demonstrates the behavioral diagnosis such that each value on the horizontal axis represents a behavioral category of the SRS report, the dotted blue vertical lines separate the three severity groups starting from mild to severe, from left to right. The orange lower region denotes the region at which a subject would be classified as TD, it the probability crosses over to the blue region, then that subject is classified as ASD for that specific behavioral trait at that specific severity. Furthermore, we investigate the highest probability for each behavioral trait across the severities to understand on what basis the classifier assigned that high probability. The lower part of fig 48 demonstrates the cortical morphological features which influenced the classifier decision. The contribution/weight of each morphological feature is denoted by its length.

Figure 48: B demonstrates the final diagnosis of the subject based on the results of the behavioral report in fig 48. In the same way, the behavioral results are linked to the underlying cortical morphological features, the final diagnosis is linked to the underlying behavioral results via LIME.

Case II

Case II is an ASD subject from $OHSU_1$ site with id 28952. Based on the SRS_T scores of that subject, the case is diagnosed to be severe ASD. Figure 49 demonstrates the proposed diagnostic report. Figure 49: A demonstrates the behavioral diagnosis such that each value on the horizontal axis represents a behavioral category of the SRS report, the dotted blue vertical lines separate the three severity groups starting from mild to severe, from left to right. The orange lower region denotes the region at which a subject would be classified as TD, it the probability crosses over to the blue region, then that subject is classified as ASD for that specific behavioral trait at that specific severity. Furthermore, we investigate the highest probability for each behavioral trait across the severity levels to understand on what basis the classifier assigned that high probability. The lower part of fig 49 demonstrates the cortical morphological features which influenced the classifier decision. The contribution/weight of each morphological feature is denoted by its length.

Figure 49: B demonstrates the final diagnosis of the subject based on the results of the behavioral report in fig 49. In the same way, the behavioral results are linked



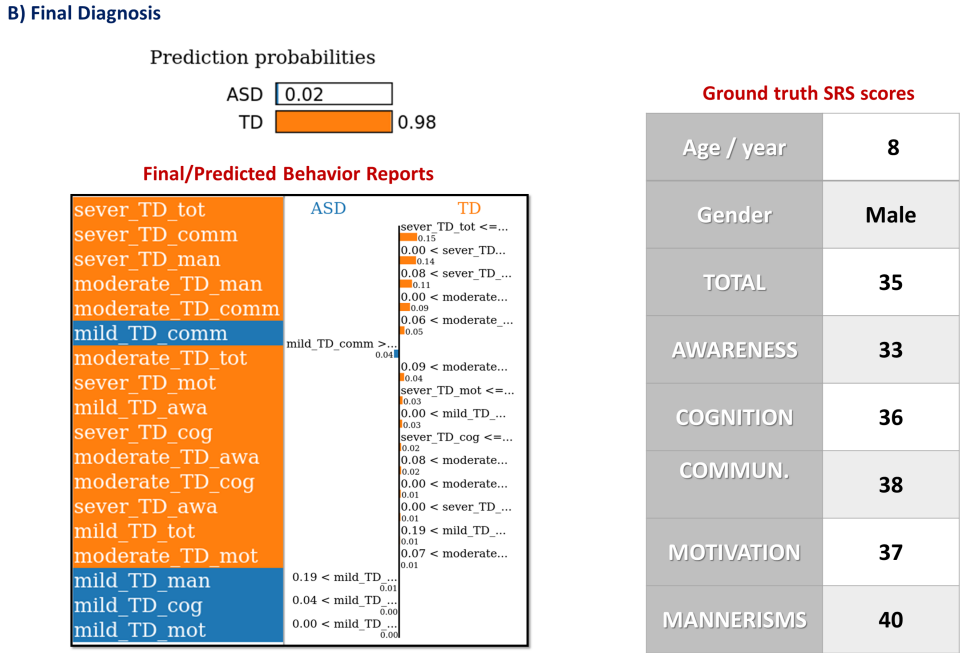
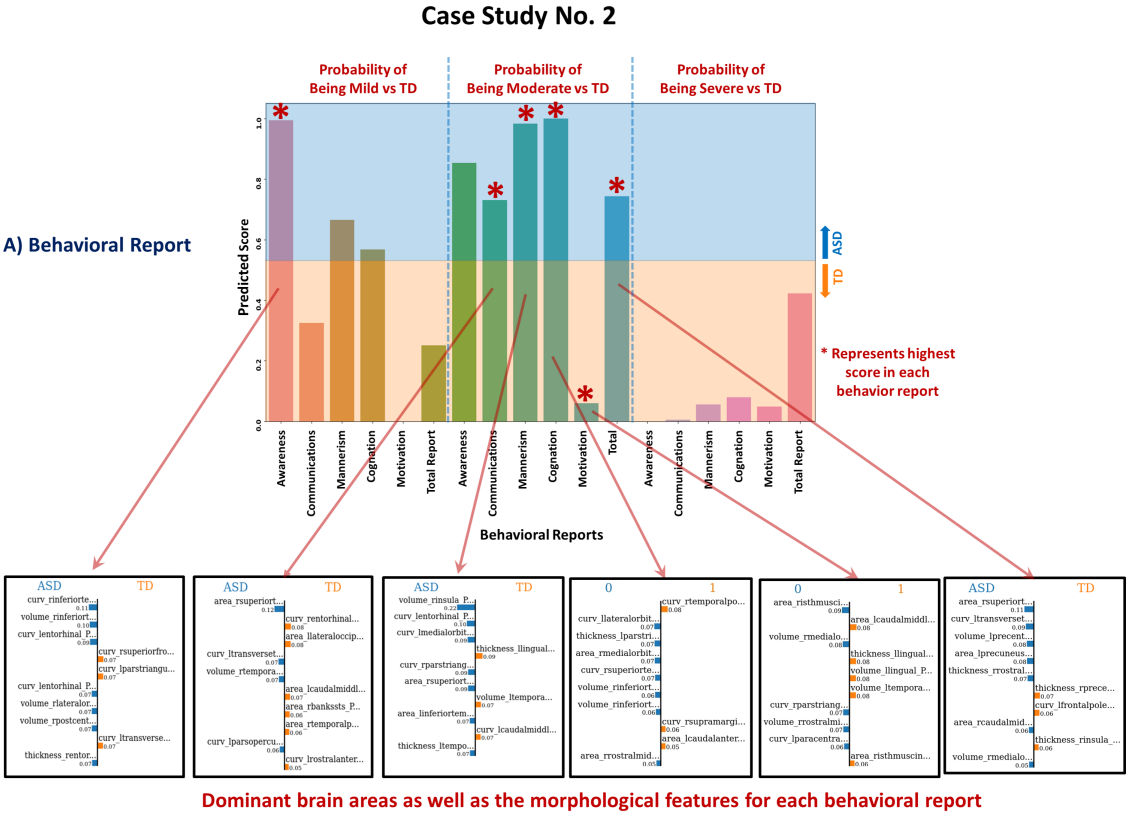
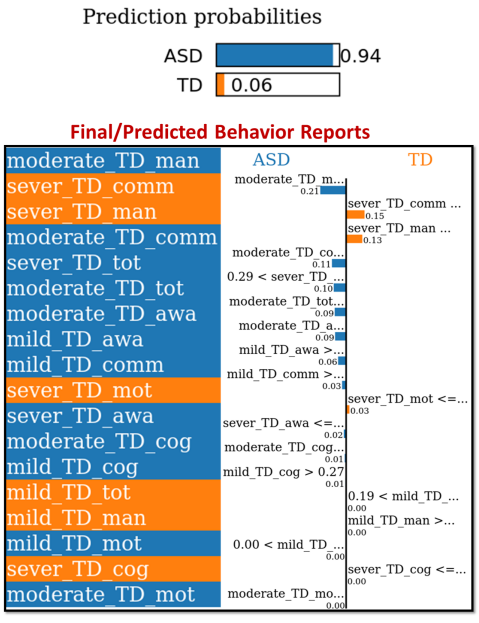


Figure 48. Proposed diagnostic report. (A) Behavioral Report (B) Final Diagnosis

to the underlying cortical morphological features, the final diagnosis is linked to the underlying behavioral results via LIME.



B) Final Diagnosis



Ground truth SRS scores

Age / year	12
Gender	Male
TOTAL	69
AWARENESS	60
COGNITION	70
COMMUN.	69
MOTIVATION	56
MANNERISMS	75

Figure 49. Proposed diagnostic report. (A) Behavioral Report (B) Final Diagnosis

4 Chapter Discussion

In the proposed study, we focus on three aims: (i) Build an ML pipeline that mimics the clinical diagnosing process of ASD, (ii) Define behavioral neurocircuits that affect alteration in the subject's behavior according to SRS module, (iii) Build a diagnostic report to connect the final diagnosis with the behavioral severity classification to the alteration in morphological features. In the following subsection, we will discuss the findings of each of the aforementioned three aims, and how this work takes a leap toward an objective psychological disorders' diagnosis.

Phase I Classification

As we previously explained, Phase I is meant to serve two tasks: (i) Find the morphological cortical features that may affect the behavior of a subject, and (ii) Train an ML model to classify the severity of a subject's behavior based on the selected morphological cortical features. RFECV with four kernels (LR, RF, LGBM, and LSVM) are employed to find the subset of cortical morphological features that maximize the balanced accuracy score of classifying the severity of ASD according to every SRS behavioral module. Each of the RFECV four kernels is initialized with the default parameters as specified in scikit-learn package, therefore, we are not expecting to achieve the highest possible classification accuracy while searching for the subset of morphological features, however, we are looking for the subset of morphological features that will correspond to the maximum classification accuracy given the utilized kernel. Thus, the maximum accuracy observed in figure 25 is far below those in table 10. This is due to the fact that the ML models utilized at the end of Phase I classification are optimized using a random search grid.

By the end of Phase I, we have an optimized ML model trained on a selected subset of features that results in the highest classification accuracy between TD and a severity-behavioral ASD group. Although this subset of morphological features can be thought of as the atlas/neuro-circuit, which affects this behavior trait, we still do not have any statistical significance to support our findings. Therefore, we decided to create the Phase I optimization step at which we repeated Phase I 51 times while randomly sampling the training-validation dataset every time in order to confirm the findings on different subsets of data. The statistically significant morphological cortical features, which demonstrated significance at $\alpha = 0.001$, are considered to be the neuro-atlas that defines the behavioral group at a given severity level for ASD. We aggregated the behavioral group findings over different severity levels to get a cortical atlas that defines the behavioral spectrum of ASD, as shown in figures from 43 to 47.

Phase I classification defines and classifies ASD as a behavioral spectrum. In the proposed study, that behavioral spectrum is defined in terms of SRS module, however, it can be defined in terms of any other ASD diagnostic reports such as ADI-R, or ADOS. Phase I classification is a standalone CAD system that places a subject into the ASD spectrum. ASD spectrum is thought of as a multidimensional space in which

each dimension represents a severity of a given behavioral trait, and the diagnosis of a subject depends on the location of that subject within that space.

The numerical results of phase I are summarized in table 8, table 9, and table 9. Table 8 demonstrates the results of the top performing severity-behavior models which are used to predict the probability of a subject to be within that severity-behavior group vs TD. The cortical morphological features utilized by each group as well as the hyper parameters set selected for each classifier can be found in supplementary material 1. Table 9 demonstrates the mean, and the standard deviation of the number of features, the balanced accuracy score, and F1-score of classifying each severity against TD over all behaviors. We assumed that the easiest to classify would be sever vs TD, since it is the one, we assumed, with highest contrast. However, phase I required more number features to successfully classify sever ASD vs TD with mean bacc of 94%. Moreover, mild ASD vs TD is the easiest to classify utilizing only 120 features on average, across all behavior groups, with bacc of 99%. We claim that the rationale behind this counter-intuitive result is the fact that a sever ASD subject would has more cortical morphological alteration, which will lead the feature selection step to select more features that discriminate between TD and sever ASD. With more number of features, a ML model requires more number of subjects to be able to achieve higher cross-validation results. From a different perspective, we hypothesize that sever ASD is merely a label for a heterogeneous ASD traits on the spectrum. Therefore, a clinical solution would be to further study sever ASD and split it into further simpler traits which are more homogenous and can be described with less number of cortical features. Table 10 demonstrates the mean, and the standard deviation of the number of features, the balanced accuracy score, and F1-score of classifying each ASD behavioral group against TD over all severity. Table 10 demonstrates that mannerism has the highest classification accuracy on average followed by cognition, and motivation, followed by communication, and total, and eventually awareness. Again we observe similar results to those in table 9, the more the number of features selected to classify behavioral ASD group and TD, the less the mean classification accuracy.

Although Phase I classification is considered to be a standalone CAD system, it is a novel approach that requires some sort of validation. The validation that we propose in this paper is via proposing a second phase of classification that yield a final diagnosis that confirms the feasibility of Phase I to mimic the behavioral report in the process of clinically diagnosing ASD.

Phase II Classification

Phase II classification is designed to mimic the final diagnosis of a physician, based on the behavioral report scores, as well as a validation of Phase I results. Phase I outputs every subject as a point in a multidimensional severity-behavior space. Given the location of each subject in that space, which we are calling "the spectrum", phase II predict whether the overall diagnosis of that subject is ASD or TD. Similar to phase I, the decision is interpreted in terms of the input data, which is the behavior report results. Therefore, the final decision is understood in terms of the behavioral-

severity report, i.e the spectrum, and the behavioral report is interpreted in terms of the cortical morphological features, which eventually can be mapped to neuro-physiological alterations. Unlike phase I, phase II does not require a normalization, or a feature selection step. Since the output of phase I is a vector of probabilities for each subjects, so all the values are guaranteed to be between 0 and 1. Moreover, there is no need to select features out of 18 features space, as we do not suspect any cure of dimensionality issues at that scale.

Numerical results of phase II are summarized in table 11, table 12, table 13, and table 14. Table 11 demonstrates the cross-validation results of utilizing multiple hyperparameters-optimized classifiers on the output of Phase I to assign a final classification of each subject whether to be ASD or TD. The results demonstrate that SVM achieves the highest classification accuracy with 94% on average. A further analysis is carried out by utilizing a bagging classifier using the hyperparameters-optimized SVM, stacking classifier using all the hyperparameters-optimized classifier (SVM, LR, xgboost, NN, lgbm, RF), and only the hyperparameters-optimized SVM. The three newly created set of classifiers are tested on a hold-out set to test the generalizability of the model, and based on the confusion matrices of tables 12, 13, and 14, we found that just utilizing the optimized SVM is sufficient to achieve the best possible results.

By the end of Phase II, we achieve a functioning CAD system that mimics a 2-step clinical process which comprises: (i) Assigning behavioral scores for a subject, (ii) Give a final diagnosis to a subject based on his behavioral scores. And to top it all off, we propose a clinical report that can be provided by the proposed model to assist physicians in diagnosing a subject, and understanding the underlying rationale behind a subject's classification.

Case Studies

We propose case studies of two randomly selected subjects out of the hold-out test set in order to demonstrate the proposed clinical report associated with the proposed model. The proposed clinical report can be found for the two case studies in figure 48 and figure 49 which are a TD subject and ASD subject respectively. The proposed clinical encapsulate most of the information coming out of phase I and phase II. The first part of the report, which is labeled as (A), illustrates what we call "the spectrum" at which each subject is defined by his/her probability of being with a given severity at a behavioral category vs being TD. A fine illustration of why we call the first part of the report "the spectrum" is shown in figure 49:A. Subject $OHSU\_1\_28952$'s behavioral report scores are within the moderate range according to his/her SRS results; this can be easily visualized from the distributions of the bars in figure 49:A where the highest bars are located in the moderate zone with an overlap with the mild zone. Each of these high probability bars are then mapped to the cortical features using LIME.

Significant Brain Regions

In this section, we cover the significance of the detected brain regions, in phase I, according to the literature to be associated with specific behavioral disorders. For each behavioral trait, we demonstrate how the associated brain regions are also nominated in the literature as relevant to that behavior.

The left bank of the superior temporal sulcus (STS) is a brain region that is thought to be involved in social cognition and the perception of social and emotional information. The STS might also play a role in helping us understand other people's intentions and emotions, as well as recognize and interpret facial expressions and other nonverbal cues that are used in social interactions, which would be classified as "mannerisms." The STS is also involved in the ability to perceive, understand, and be conscious of one's surroundings and experiences, and this is its association with "awareness". The STS is also involved in the perception of speech and language and may play a role in helping us to understand the meanings of words and sentences. [156–158].

The caudal anterior cingulate cortex (cACC) is a brain region in the medial frontal lobe. It is thought to be involved in various functions anatomically and functionally related to frontal executive functions, parietal sensorimotor systems, and limbic intentional or emotional processes, including the regulation of emotional and behavioral responses, attention and cognitive control, and pain perception, which is all related to "awareness." It is also believed to play a role in integrating information from different sources, such as sensory, motor, and affective inputs [159].

The left caudal middle frontal region of the brain, also known as Broca's area, is associated with literacy and language processing, which includes the ability to understand and produce written language, as well as the planning and coordination of movements required for writing. This ability to understand and produce written language is associated with awareness, communication, and mannerism in several ways: Awareness: Literacy allows individuals to access and process written information, allowing them to understand the world around them and be aware of their surroundings. Communication: Literacy enables individuals to understand and use written language to communicate with others, to read and comprehend written texts, and to express themselves through writing. Mannerism: The planning and coordination of movements required for writing is a form of mannerism, which is the control and execution of movements, gestures, and facial expressions used to communicate [160].

The right caudal middle frontal region is associated with attention and problem-solving, specifically numerical and spatial processing. It is often referred to as the "number sense" region because it involves our ability to understand and manipulate numerical concepts. This ability to understand and manipulate numerical concepts is closely related to cognitive processes such as attention, problem-solving, and decision-making. It allows individuals to perform mathematical operations, understand and manipulate numbers, and perform spatial reasoning tasks. This ability to perform mathematical operations and spatial reasoning are essential for daily life, such as navigation, budgeting, and understanding scientific concepts. In terms of awareness, this region also plays a role in attentional processes, which enables individuals to

selectively attend to specific information in the environment and filter out irrelevant information. It also helps decision-making by allowing individuals to make informed decisions based on numerical and spatial information [160].

The cuneus is a brain region located in the posterior (back) part of the cerebral cortex. It is part of the occipital lobe, which is involved in processing visual information. The left cuneus is involved in the processing of visual information from the right visual field, as well as in various higher-level visual processes such as color perception and visual attention. It may also play a role in the integration of visual information with other sensory modalities, such as touch and sound. Studies have shown no direct association with mannerism, however, it is possible that color perception could play a role in the interpretation of certain types of mannerisms. For example, changes in skin color (e.g. blushing) can be used as an indicator of embarrassment or anxiety, and thus, color perception could be involved in recognizing this type of mannerism [161]. The right cuneus is also involved in the processing of visual information, specifically from the left visual field. Like the left cuneus, it is involved in higher-level visual processes such as color perception and visual attention, and may also play a role in the integration of visual information with other sensory modalities. In general, the cuneus is thought to play a role in the interpretation of visual stimuli, particularly those that are complex or abstract in nature [161].

The left entorhinal region of the brain is involved in spatial memory and navigation. It is a part of the medial temporal lobe, which is a region of the brain that is important for learning and memory (spatial and semantics). The entorhinal region is specifically involved in spatial memory, which is the ability to remember the layout of a space and to navigate through it. This region is also involved in the consolidation of memories from short-term to long-term storage, and it plays a role in the processing of sensory information from the environment. . It is not directly associated with communication, but it can play a role in communication in certain contexts [162]. While the left entorhinal region is specifically associated with spatial memory, the right entorhinal region is more involved in memory for events and experiences. The right entorhinal brain region is associated with awareness, cognition, and mannerism in several ways: Awareness: The right entorhinal cortex (ERC) plays a crucial role in spatial awareness, allowing individuals to navigate and orient themselves within their environment. Damage to the ERC can result in spatial disorientation and difficulty navigating familiar environments. Cognition: The ERC is also involved in cognitive processes such as memory, attention, and decision-making. Damage to the ERC can result in memory impairment and difficulty with attentional tasks. Mannerism: The ERC is also involved in the regulation of habits and automatic behaviors. Damage to the ERC can result in the loss of control over habitual behaviors, leading to changes in mannerism. This can manifest in a variety of ways, such as repetitive movements or compulsions. Overall, the right entorhinal brain region plays a critical role in a wide range of cognitive and behavioral processes, and damage to this region can have significant impacts on an individual's ability to navigate, remember, make decisions, and control their behavior [162].

The left fusiform region of the brain is involved in the recognition of words and letters. It is a part of the fusiform gyrus, which is a region of the brain located in

the temporal lobe and is important for visual processing. The left fusiform region is specifically involved in the recognition of familiar words and letters, and it is thought to play a role in the "how" pathway of the ventral visual stream, which is responsible for the identification of the shapes and forms of objects. The left fusiform region is also involved in object and face recognition, but it is more specifically associated with the recognition of words and letters. The left fusiform brain region is associated with awareness and cognition in several ways: Awareness: The left fusiform gyrus (FFG) is involved in facial recognition and the processing of facial expressions. Damage to the left FFG can result in difficulty recognizing and interpreting facial expressions, which can impact social interactions and communication. Cognition: The left FFG is also involved in language processing and reading. Damage to the left FFG can result in difficulty with reading and language comprehension. Additionally, the left fusiform gyrus is also associated with object recognition, meaning that it plays a role in recognizing familiar objects and objects that are similar to each other. Damage to the left FFG can result in difficulty in recognizing objects [163]. The right fusiform region is also involved in object and face recognition, but it is more specifically associated with the recognition of familiar objects and faces. It is thought to play a role in the "what" pathway of the ventral visual stream, which is responsible for the recognition of objects and their features. The right fusiform brain region is associated with awareness, cognition, and communication in several ways: Awareness: The right fusiform gyrus (FFG) plays a role in spatial awareness and perception, allowing individuals to perceive and understand the spatial relationships between objects in their environment. Damage to the right FFG can result in difficulty with spatial perception and navigation. Cognition: The right FFG is also involved in the recognition of objects and patterns, particularly those that are non-verbal such as shapes, colors, and textures. Damage to the right FFG can result in difficulty recognizing and interpreting non-verbal information. Communication: The right FFG is also involved in the perception of non-verbal cues such as facial expressions, body language, and tone of voice. Damage to the right FFG can result in difficulty interpreting and understanding non-verbal cues, which can impact social interactions and communication [163].

The right inferior temporal region is involved in visual object recognition and the processing of visual information, but it is more specifically associated with the recognition of faces and facial expressions, unlike the left inferior temporal region is specifically involved in the recognition of familiar objects and is thought to play a role in the "what" pathway of the ventral visual stream, which is responsible for the recognition of objects and their features. The right inferiortemporal brain region is associated with awareness, cognition, communication, and mannerism in several ways: Awareness: The right inferiortemporal cortex (ITC) plays a role in visual perception and the recognition of objects and scenes. Damage to the right ITC can result in difficulty recognizing and perceiving objects and scenes. Cognition: The right ITC is also involved in the processing of visual information and the formation of visual memories. Damage to the right ITC can result in difficulty forming and retrieving visual memories. Communication: The right ITC is also involved in the interpretation of non-verbal cues such as facial expressions and body language, which are critical for

effective communication. Damage to the right ITC can result in difficulty interpreting and understanding non-verbal cues. Mannerism: The right ITC is also involved in the regulation of habits and automatic behaviors. Damage to the right ITC can result in the loss of control over habitual behaviors, leading to changes in mannerism. This can manifest in a variety of ways, such as repetitive movements or compulsions [164].

The left inferior parietal brain region plays a critical role in a wide range of cognitive and behavioral processes, including motor control, perception, attention, and language, and damage to this region can have significant impacts on an individual's ability to coordinate and control their movements, perceive and understand their environment, pay attention and communicate effectively, leading to changes in mannerism [165]. The right inferior parietal lobe is also involved in the processing of sensory information, particularly information related to spatial awareness and the perception of one's own body in space. In addition, the right inferior parietal lobe is thought to play a role in mathematical and spatial reasoning, and in the manipulation of mental representations of objects and their spatial relationships. Some research suggests that the right inferior parietal lobe may be more involved in the processing of visuospatial information than the left inferior parietal lobe. It is associated with Awareness in several ways: The right inferior parietal lobule (IPL) is involved in spatial awareness and perception, allowing individuals to navigate and orient themselves within their environment. It also plays a crucial role in mathematical and spatial reasoning, allowing individuals to understand and manipulate mathematical concepts and spatial relationships. Damage to the right IPL can result in spatial disorientation and difficulty navigating familiar environments, as well as difficulty with mathematical tasks such as solving equations or understanding geometric concepts [165].

The Left insula is associated with negative emotions and interoception, the ability to perceive internal bodily sensations. It's thought to play a role in cognitive processes such as attention, decision-making, and memory. It's involved in the processing of emotions and social cues in language, and may be important for understanding the emotions of others. The Left insula also plays a role in the regulation of bodily movements and postures which may be important for social interactions and communication [166]. The right insula is thought to play a role in various cognitive and emotional processes, specifically in the following behavioral domains: Awareness: The right insula is thought to play a role in exteroception, which is the ability to perceive external stimuli such as touch and temperature. It also thought to be involved in the sense of body ownership and self-awareness, Communication: The right insula is thought to be involved in the processing of nonverbal cues, such as facial expressions and body language, which may be important for social interactions and communication. Additionally, the right insula is thought to play a role in the perception of music and prosody, which are important elements of communication in speech. Motivation: The right insula is thought to be involved in regulating motivation, particularly with regards to reward processing, and decision-making. Studies have found that the right insula is activated when individuals are presented with pleasurable stimuli, such as food or monetary rewards, and that it may play a role in the experience of pleasure and desire. Positive emotions: Research suggests that the right insula may be more closely associated with positive emotions such as happiness and pleasure [166].

The left isthmuscingulate cortex, also known as the anterior cingulate cortex (ACC), is thought to play a role in various cognitive and emotional processes, including awareness, communication, cognition, and motivation. Awareness: The ACC is involved in attention and awareness, particularly in conflict and error detection. Communication: The ACC is involved in processing social and emotional information, such as facial expressions, and may play a role in empathy and theory of mind. Cognition: The ACC is involved in cognitive control and decision-making, particularly in goal-directed behavior and conflict resolution. Motivation: The ACC is involved in regulating motivation and emotion, particularly in pleasure and desire. Memory: The ACC is involved in encoding and retrieval of memories and formation of new memories. Pain processing: The ACC is involved in processing and regulating pain perception [167]. Studies have found that the right ACC is activated when individuals engage in repetitive behaviors, such as tic disorders or obsessive-compulsive disorder (OCD), and it is thought that this activation may be related to the regulation of these behaviors [167].

The left lateral occipital (LO) region is a key area in the brain that is involved in visual processing and object recognition, including the recognition of the shape and form of objects. This region is thought to play a role in awareness, mannerisms, and cognition by helping us to quickly and efficiently recognize objects in our environment and understand their shape and form [168].

The left lateral orbitofrontal cortex (lOFC) is a brain region that is involved in several aspects of behavior and cognition, including awareness, communication, mannerism, cognition, and motivation. In particular, the lOFC has been found to play a role in processing information about punishers and non-rewards, which can affect our awareness, communication, mannerisms, cognition, and motivation [169]. the right lateral orbitofrontal cortex (OFC) could be associated with: Awareness: As the OFC is involved in conscious perception and attention, allowing us to be aware of our surroundings. Mannerisms: As the OFC is involved in regulating behavior, including mannerisms and social behavior. Cognition: As the OFC is involved in a variety of cognitive processes, including decision making, emotional regulation, and reinforcement learning. These processes are important for overall cognitive function and behavior [169].

The lingual brain region is a part of the occipital lobe, which is located at the back of the brain and is involved in visual processing. The lingual brain region is located near the back of the occipital lobe and is thought to play a role in the processing of visual information, including the recognition of faces and other objects. Some research suggests that the lingual brain region may be involved in the encoding and retrieval of memories related to visual information, particularly the recognition of faces. It may also be involved in the generation of visual images during dreaming, as well as in various other aspects of visual perception and processing [170]. It is not clear whether the left and right lingual brain regions have any specific functions that are distinct from each other. However, it is likely that the two brain regions work together to perform various functions related to visual processing and recognition. More research is needed to fully understand the specific functions of the lingual brain region and its role in the development and maintenance of mental health conditions

such as depression [170].

The medial orbitofrontal cortex is a brain region in the frontal lobe that is thought to be involved in various functions, including emotion regulation and decision-making. And this is closely tied to awareness and motivation. Emotion regulation: Emotion regulation is controlling or modifying our emotional responses to events to align them with our goals and values. Emotion regulation is closely tied to awareness because it requires us to be aware of our emotional states and understand the emotional significance of events. When we can regulate our emotions effectively, we can better respond to events in a way that is consistent with our goals and values, which contributes to our overall awareness and understanding of the world around us. Decision-making: Decision-making is the process of choosing between alternative courses of action, closely tied to awareness and motivation. When making decisions, we must be aware of each option's potential outcomes and our own goals and values, and we must be motivated to choose an option that will bring us closer to our goals. The decision-making process requires us to be aware of the options available to us, weigh the potential outcomes of each option, and make a choice that is aligned with our goals and values [169]. It is not clear whether the left and right medial orbitofrontal cortex (mOFC) brain regions have any specific functions that are distinct from each other. However, the right mOFC might be less involved in regulating positive emotions than the left mOFC and plays a more significant role in regulating negative emotions. Overall, emotion regulation and decision-making are closely tied to awareness, communication, and mannerism, as effective decision-making can also improve communication by allowing individuals to respond to social cues and make choices that are consistent with their goals and values. Regarding mannerism, The right mOFC is thought to play a role in mannerism, which refers to how we express our emotions and attitudes through nonverbal cues, such as facial expressions, posture, and gestures [169].

The left MTG is associated with awareness through its role in language processing (particularly with respect to semantics (the meanings of words and phrases)), memory, and attention. By helping us to understand the meaning of words, recall verbal information, and attend to language-related information, the left MTG contributes to our overall understanding of the world and our ability to navigate it effectively. By processing language and other verbal information, the left MTG also plays a critical role in our ability to communicate effectively and to interact with others in a meaningful way [171]. The right middle temporal brain region is also thought to be involved in language processing, particularly with respect to the sounds and rhythms of language (phonology and prosody). This brain region is often active when people are listening to or producing language, and it is thought to play a role in helping us to recognize and produce the sounds and rhythms of language. It is also thought to be involved in memory and attention, and may play a role in helping us to pay attention to and remember important information which could be how it's associated with motivation. Additionally, the right middle temporal brain region is thought to be involved in spatial processing and may play a role in helping us to navigate and orient ourselves in the environment [171].

The left paracentral lobule (PCL) is a brain region that is involved in the con-

trol of movement and associated with communication, mannerism, and motivation through its role in the control of movement. By helping us to execute movements with precision and accuracy, regulate posture and mannerisms, and regulate the drive to move and take action, the left PCL contributes to our ability to communicate effectively, to interact with others in a mannerly and appropriate way. It's not clear how it could be associated with motivation but it could be through the regulation of motivational processes, particularly in the regulation of the drive to move and to take action. This region is involved in our ability to initiate and sustain movement, which is essential for effective communication and interpersonal interactions, as well as for pursuing goals and ambitions [172]. The right PCL is associated with awareness through its role in the control of movement and in the regulation of attention to sensory inputs related to movement and bodily sensations. By helping us to execute movements with precision and accuracy, maintain awareness of the position and movements of our body, and attend to and process sensory information related to our own movements and sensations, the right PCL contributes to our ability to be aware of our own actions and movements, to regulate our behavior effectively, and to be self-aware [172].

The left parahippocampal is associated with cognition and communication through its involvement in learning, memory, and spatial navigation. By contributing to our ability to encode and recall information, orient ourselves in space, and navigate effectively, the left PHG helps us to process and respond to environmental cues, to communicate effectively, and to engage in meaningful interactions with the world around us. The right parahippocampal is thought to have the same functionality as the left parahippocampal, it's not clear why it might not be associated with communication. [173].

The right parsopercularis (PO) is associated with awareness, communication, cognition, and motivation through its involvement in language processing, speech production, and cognitive control. By contributing to our ability to understand language, articulate words and sentences effectively, regulate our behavior, and process and respond to environmental cues, the right PO helps us to communicate effectively, acquire new knowledge and skills, and engage in meaningful interactions with the world around us [174].

The left parstriangularis is associated with awareness and cognition, especially through its role in semantic processing of language. By contributing to our ability to understand the meaning of words and sentences, this region helps us to process and respond to environmental cues, to engage in meaningful interactions with the world around us, and to build a coherent understanding of the world around us [175]. The right parstriangularis is the same as the left with regards to awarness and cognition. It's not clear whether there's a difference between the right and the left regions, however, this study may suggest that there's some slight structural differences between them. Overall, it could be also assoicated with communication through its involvement in the processing of linguistic information, which is crucial for effective communication. By contributing to our ability to understand the meaning of words and sentences, the right PT helps us to engage in meaningful interactions with others, and to convey and receive information effectively. And it's not clear whether it's

directly releated to motivation but it could be indirectly associated with it through its involvement in language processing and semantic processing, which are essential for effective communication, and this in turn might be related to motivation in certain situations. For example, when someone is motivated to engage in a conversation with others, they might activate their right PT as they process the linguistic information they are receiving and produce responses that are semantically appropriate [175].

The left pericalcarine brain region is important for visual processing and awareness, especially in the identification of objects and shapes, and is part of the neural network responsible for processing visual information and forming a complete representation of the world around us [176].

The left postcentral brain region is important for processing various somatic sensations and is part of the neural network responsible for processing sensory information and forming a complete representation of the body and its surroundings. This region plays a role in our awareness and helps us respond appropriately to various sensory stimuli in our environment. The direct association between the left postcentral brain region and motivation is not well understood. However, motivation can be influenced by various factors, including sensory processing and awareness. For example, if an individual is motivated to engage in a particular activity or task, they may pay greater attention to the sensory information related to that task, and their primary somatosensory cortex, including the left postcentral brain region, may be more active. The right postcentral is s thought to have the same functionality as the left postcentral [177].

The left precentral brain region is associated with awareness through its role in the control of voluntary movements and the processing of sensory information related to movement. By allowing individuals to be aware of their own movements and to adjust them accordingly, the left precentral brain region plays an important role in overall body awareness and movement control [178]. The right precuneus brain region is associated with various cognitive processes including self-awareness, memory, and spatial orientation. Some studies suggest that the right precuneus may play a role in processing self-relevant information and in monitoring one's own thoughts and actions. It is also involved in tasks related to attention and consciousness, and in integrating information from different parts of the brain to create a sense of self and promote self-awareness. There is limited research on the exact role of the right precuneus in motivation, but it may play a role in regulating motivation and drive by integrating information from various brain regions to generate goal-directed behavior [178].

The left posterior cingulate cortex (PCC) is thought to play a role in various cognitive, executive, emotional, and linguistic functions. Its involvement in communication and mannerism is supported by evidence of its role in language processing and semantic processing. Its association with cognition is due to its role in attention, working memory, and executive control processes, as well as its involvement in self-referential thinking and self-awareness. Its involvement in motivation is less clear, but some studies suggest that it may play a role in the regulation of motivation and goal-directed behavior. The right posterior cingulate cortex is thought to have the same functionality as the left posterior cingulate cortex [179].

The left rostral anterior cingulate cortex (rACC) has been linked to various aspects of emotional processing and regulation, attention, and decision-making. For example, research has shown that the left rACC is involved in the regulation of negative emotions such as fear. This brain region has also been found to be involved in attentional processes, including the allocation of attention to emotionally significant stimuli. In addition, the left rACC has been implicated in decision-making, particularly in making decisions that involve emotional or reward-related factors. These findings suggest that the left rACC plays a role in a variety of functions related to awareness, mannerism, and motivation, including emotional regulation, attention, and decision-making [180]. The right rACC has been shown to play a role in cognition, specifically through its involvement in emotional processing, attention regulation, and decision-making. This region has been implicated in processing negative emotions such as fear, as well as attention allocation and resolving conflict between competing responses. Additionally, the rACC has been shown to play a role in integrating information from various sources to support goal-directed behavior and decision-making [180].

The left supramarginal brain region might be associated with awareness and mannerism, especially through the processing of pitch in musical sounds and speech, language perception, and the processing of emotional information. This region is thought to play a role in the integration of sensory and linguistic information, as well as in the processing of emotional and musical stimuli [181]. The right supramarginal brain region is thought to have the same functionality as left supramarginal brain region. Additionaly, the right supramarginal brain region has been associated with various aspects of cognition, including attention, language processing, and perception [181].

The left temporal pole is a brain region located in the temporal lobe of the cerebral cortex. It is involved in a number of functions, including language, perception, and emotion. With respect to awareness, it is believed to play a role in processing information about the meaning of objects, including both semantic and emotional information. It is also thought to be involved in language processing and perception. With regards to mannerisms, the left temporal pole is involved in regulating emotional expression and social behavior. It is thought to play a role in processing information about the emotional content of social cues and experiences. In terms of cognition, the left temporal pole is believed to play a role in various cognitive processes, including memory and attention. It is thought to be involved in encoding and retrieval of memories, as well as the allocation of attention to important stimuli. Additionally, it is thought to play a role in the processing of semantic information and the formation of concepts [182]. The right temporal pole is a complex region of the brain that integrates information from multiple domains to support various aspects of cognition, including awareness, communication, and motivation. In terms of awareness, the RTP is involved in semantic processing and integrating information from multiple sensory modalities to form a coherent perception of the environment. With regards to communication, the RTP is involved in language processing, particularly in the processing of object meaning and semantic memory. It also contributes to the interpretation of nonverbal cues in social interaction. In terms of motivation, the RTP is involved in the processing of emotional and affective information, including the regulation of emotions and the formation of emotional memories. It

also plays a role in the processing of reward-related information, which can influence motivation [182].

The left transverse temporal (LTT) region, also known as the primary auditory cortex or Brodmann area 41, aka left Heschl's gyrus, is associated with the processing of auditory information. In terms of awareness, the LTT is involved in the perception and interpretation of auditory stimuli, including speech and music, and contributes to the conscious experience of sound. With regards to mannerisms, the LTT plays a role in the processing of musical rhythm and melody, which are important components of musical performance and appreciation. In terms of cognition, the LTT is involved in the integration of auditory information with other sensory modalities, such as vision, to form a cohesive perceptual representation of the environment. Regarding communication, the LTT is involved in the processing of speech sounds, including the ability to discriminate between different speech sounds, which is crucial for understanding spoken language. In terms of motivation, the LTT is involved in the processing of emotional and affective information conveyed through auditory stimuli, including the regulation of emotions and the formation of emotional memories. The right transverse temporal is thought to have the same functionality as LTT [183].

The left superior frontal (LSF) region, also known as the dorsal lateral prefrontal cortex (DLPFC), is associated with higher cognitive functions, including working memory (WM). In terms of mannerisms, the LSF region is involved in the regulation and control of behavior, including the ability to perform complex tasks and exhibit refined motor skills. This region is involved in the process of generating voluntary movement, which is important for the expression of mannerisms [184]. The right superior frontal (RSF) is thought to have the same functionality as the LSF. And in terms of motivation, the RSF region is involved in the processing of reward-related information and the regulation of goal-directed behavior. This region helps to guide behavior towards rewarding outcomes and plays a role in the formation of habits and patterns of behavior [184].

In terms of mannerisms, the left superior parietal (LSP) region is involved in the control of fine motor movements, including the precise movements of the hands and fingers. This region is also involved in the generation of voluntary movement, which is important for the expression of mannerisms. Furthermore, the LSP region is involved in the processing of information about the size, shape, and texture of objects, as well as the formation of object representations in working memory. This region plays a key role in the perception of objects in the environment, including the ability to manipulate objects, such as when holding or reaching for objects [185]. The right superior parietal (RSP) is thought to have the same functionality as the LSP region, and these are all high cognitive processes. Moreover, the RSP region is involved in attention and spatial processing, including the ability to attend to and manipulate multiple objects in space. This region plays a key role in the perception of objects in the environment, including the ability to manipulate objects, such as when holding or reaching for objects [185].

The left superior temporal (LST) region is associated with various aspects of auditory processing, including short-term memory and speech comprehension. This region is involved in the processing of speech sounds, including the ability to distinguish be-

tween different speech sounds and to understand speech in noisy environments. In terms of communication, the LST region is critical for the comprehension of spoken language and the integration of linguistic information with other sensory inputs, such as facial expressions and body language. This region is also involved in the processing of prosody, including the rhythm, stress, and intonation patterns of speech, which provide important cues for the interpretation of meaning in spoken language. In terms of mannerisms, the LST region is involved in the processing of auditory information, including the perception of speech sounds and prosody. This region is also involved in the regulation of voluntary movement, which is important for the expression of mannerisms. Regarding motivation, the LST region is involved in the processing of auditory information and the regulation of behavior in response to auditory stimuli. And this is how it could be involved in the processing of reward-related information, contributing to the regulation of goal-directed behavior and motivation [186]. The right superior temporal (RST) is thought to share the same functionality with the LST region. And regarding cognition, the RST region is involved in the processing of auditory information and the integration of this information with other sensory inputs, including visual and somatosensory information, to form a cohesive perceptual representation of the environment. This region is also involved in the processing of musical information and the perception of musical patterns and structures. In terms of awareness, the RST region is involved in the processing of auditory information and the regulation of behavior in response to auditory stimuli. This region is also involved in the processing of attention-related information, contributing to the regulation of attentional focus and awareness of the environment [186].

The left frontal pole (LFP) region is associated with various aspects of higher cognitive function, including consciousness, communication, memory, and attention. In terms of mannerisms, the LFP region is involved in the regulation of voluntary movement, which is important for the expression of mannerisms. The LFP is also involved in the processing of sensory information and the regulation of behavior in response to this information, contributing to the expression of mannerisms. Regarding cognition, the LFP region is involved in various aspects of higher cognitive function, including working memory, executive function, and attentional control. The LFP is also involved in the processing of linguistic information and the regulation of language-related behavior, including speech production and comprehension [187]. The right frontal pole (RFP) is thought to share the same functionality with the LFP region, and its association with motivation is less clear, however, it could aslo be through procession of information that regulate a goal-directed behavior [187].

The left rostral middle frontal cortex (LMF) is associated with several aspects of motivation and behavior regulation, including emotion regulation and working memory. In terms of emotion regulation, the LMF is involved in the appraisal of emotional information, the regulation of emotional responses, and the regulation of attentional focus in response to emotionally-relevant information. This region is also involved in the generation of cognitive strategies that help to regulate emotions, such as reappraisal, distraction, and suppression. In terms of working memory, the LMF is involved in the temporary storage and manipulation of information, which is essential for the performance of various cognitive tasks. The LMF is also involved in

the regulation of attentional focus and the allocation of resources to working memory, which are important for the maintenance of task-relevant information in working memory [188]. The right rostal middle frontal cortex (RMF) is thought o share the same functionality with the LMF region. And regulating those emotions and generating those cognitive strategies can be important in social communication. Overall, the RMF region plays a critical role in the neural network involved in communication, contributing to the regulation of emotional responses, the allocation of attentional resources to support language processing, and the maintenance of task-relevant information in working memory [188].

5 Conclusion

In this study, we present a pioneering method for the diagnosis of autism and other psychological disorders through the replication of the clinical diagnosis process using artificial intelligence. Our proposed framework consists of two crucial stages in diagnosing/classification a subject with Autism Spectrum Disorder (ASD). Firstly, we obtain morphological features from the MRI scans of each subject and identify the most salient features that accurately differentiate ASD within various behavioral domains. Secondly, we categorize each subject as having severe, moderate, mild, or typical development (TD) based on the behavioral domains of the SRS. Finally, we make an overall classification of each subject as either ASD or TD based on the subject's behavioral categorization. It is also important to note that the current diagnostic process for ASD relies heavily on reports based on patient interviews and physician-based scores, which can be time-consuming and susceptible to human error. The ability of our AI-based model to detect functional differences in brain regions using MRI scans alone can not only speed up the diagnostic process but also increase its accuracy, leading to improved outcomes for individuals with ASD.

The proposed framework also provides clarification and interpretation of the classifier's decisions at every step. During the training of the classifiers, we constructed neuroatlases to gain insight into the correlation between brain region morphology and various behavioral traits of each subject. The regions of the brain defined in each behavioral neuroatlas were chosen based on a combination of machine learning classification efficiency and statistical significance. Finally, interpretable methods were employed to demonstrate, for new subjects, the mechanisms and reasons behind their classification/diagnosis. This interpretability phase has been included to assist physicians in comprehending the fundamental causes of ASD and to enable them to offer presonalized medical treatment for each subject.

Our AI-based model's ability to accurately differentiate between the functionalities of specific brain regions, such as the left and right caudal middle frontal regions, in the classification of behavioral severity of Autism Spectrum Disorder is a significant advancement in the field. The model has found that the left caudal middle frontal region is linked to mannerism, awareness, and communication behavioral domains, while the right caudal middle frontal region is primarily associated with awareness and cognition, making it more associated with analytical processes rather than social processes. This is particularly noteworthy because previous research has also

documented a remarkable difference between the left and right caudal middle frontal regions. At the same time, most studies show little to no difference between other left and right brain regions. This shows that our AI model has the ability to accurately identify specific contralateral regional differences or similarities, adding to the validity of the model's findings.

It is important to note that all behavioral domains are complex and multifaceted processes, and many different brain regions are thought to be involved in different aspects of it. It is also important to note that the field of neuroscience is always evolving and the understanding of the function of different brain regions is still under research.

CHAPTER V
CONCLUSIONS AND FUTURE WORK

This dissertation presents a novel approach to understand and diagnose ASD using AI by mimicking the clinical setup. The proposed approach depends on building a sophisticated AI to framework that allows to accurately diagnose ASD while allowing backtracking to the underlying neuro abnormalities. The core of the proposed AI framework primarily consist of four key stages: i) preprocessing, ii) feature calculation/extraction, iii) feature engineering (including feature representation and selection, among others), and iv) machine learning-based classification. The developed models and approaches within this dissertation demonstrate promising overall outcomes, as well as insights into alternative methods. Subsequently, the primary contributions of each module are detailed.

1 Summary of contributions

Homogeneous approach

A ML pipeline is designed and implemented to identify morphological imaging markers of ASD. The proposed pipeline includes data preprocessing, feature extraction, feature normalization and age adjustment, feature selection via four different RFECV models, and classification using hyperparameters optimization of linear and non-linear machine learning models. The most discriminative set of features is formed using RFECV+lg2 model. The resultant features' set of RFECV+lg2 is used to train a set of linear, and non-linear classifiers. The highest balanced accuracy score is achieved by NN for both the global model and the local model with average balanced accuracy score of 71.6% $\pm$ 2% and 97% $\pm$ 2% respectively. The most common features among the global model and each site of the local model are then analyzed to create ASD neurocircuits. The two main steps that helped in achieving the high results are: (i) feature selection step via RFECV, and (ii) Hyperparameter optimization of the classifiers. The overall structural mapping of cognition and behaviors to distinct neuroanatomical and functional linked neural circuits is more likely to not only diagnosed but map a cluster of ASD individuals whose behaviors and characteristics are more similar than different. The proposed pipeline is anticipated to achieve better results than those in the literature because of the way that the morphological feature values are aggregated is less prone to outliers. RFECV implementation with more than one classifier will also cover as many assumptions on the relationship between the features and the target as possible while selecting the features while performing hyper-parameter optimization using grid search on eight classifiers to achieve the optimum results given the selected set of features. Neuroimaging is an attractive

non-invasive technology to facilitate the definition of relationship between genes, environment, and behaviors in ASD. While this study's numbers, design, mitigation of age/sex, pre-processing, etc. lend credence to these results, the truth is that use of sMRI and fMRI data is still a challenge since large datasets from typically developed children from infancy through 8 years of age are still lacking. The current sample size does identify brain regions implicated infants who are at high risk for ASD suggesting that this approach is scalable for use in larger more heterogeneous groups of ASD populations. The higher accuracy of ASD classification in this study also reinforces this hypothesis. Ultimately, the proposed system should provide a complete map explaining what linked brain regions are affected, to what extent impairments are more severe, and thereby could be very useful to a treating physician/provider from a clinical point of view. We hypothesize the difference in the balanced accuracy score of the global model and the local model is due to the high heterogeneity of the disorder. This hypothesis is based on the number of common features among the sites of the local model, as well as the number of features being selected for both the global and local model. Consequently, for the future work, we are planning to do the following: (i) incorporate the ABIDE II dataset along with ABIDE I, and (ii) partition subjects based on behavioral traits in order to subdivide ASD into more types of "homogeneous ASD" where subjects share more traits. Eventually, we proposed a personalized diagnosis method at which we describe the phenotype of each subject in terms of the local imaging markers values. The outcome of this step is a personalized model that describe the affected brain regions which made the classifier decide a specific subject to be ASD. We hypothesize that the affected brain region, giving their feature values, might be correlated with a brain physiological anomaly that might be causing a specific autistic behavior. Thus, by recognizing those affected brain regions, a personalized treatment can be assigned for each subject to help with autistic traits moderation

Heterogeneous approach

In this study, we present a pioneering method for the diagnosis of autism and other psychological disorders through the replication of the clinical diagnosis process using artificial intelligence. Our proposed framework consists of two crucial stages in diagnosing/classification a subject with Autism Spectrum Disorder (ASD). Firstly, we obtain morphological features from the MRI scans of each subject and identify the most salient features that accurately differentiate ASD within various behavioral domains. Secondly, we categorize each subject as having severe, moderate, mild, or typical development (TD) based on the behavioral domains of the SRS. Finally, we make an overall classification of each subject as either ASD or TD based on the subject's behavioral categorization. It is also important to note that the current diagnostic process for ASD relies heavily on reports based on patient interviews and physician-based scores, which can be time-consuming and susceptible to human error. The ability of our AI-based model to detect functional differences in brain regions using MRI scans alone can not only speed up the diagnostic process but also increase its accuracy, leading to improved outcomes for individuals with ASD.

The proposed framework also provides clarification and interpretation of the classifier's decisions at every step. During the training of the classifiers, we constructed neuroatlases to gain insight into the correlation between brain region morphology and various behavioral traits of each subject. The regions of the brain defined in each behavioral neuroatlas were chosen based on a combination of machine learning classification efficiency and statistical significance. Finally, interpretable methods were employed to demonstrate, for new subjects, the mechanisms and reasons behind their classification/diagnosis. This interpretability phase has been included to assist physicians in comprehending the fundamental causes of ASD and to enable them to offer presonalized medical treatment for each subject.

Our AI-based model's ability to accurately differentiate between the functionalities of specific brain regions, such as the left and right caudal middle frontal regions, in the classification of behavioral severity of Autism Spectrum Disorder is a significant advancement in the field. The model has found that the left caudal middle frontal region is linked to mannerism, awareness, and communication behavioral domains, while the right caudal middle frontal region is primarily associated with awareness and cognition, making it more associated with analytical processes rather than social processes. This is particularly noteworthy because previous research has also documented a remarkable difference between the left and right caudal middle frontal regions. At the same time, most studies show little to no difference between other left and right brain regions. This shows that our AI model has the ability to accurately identify specific contralateral regional differences or similarities, adding to the validity of the model's findings.

It is important to note that all behavioral domains are complex and multifaceted processes, and many different brain regions are thought to be involved in different aspects of it. It is also important to note that the field of neuroscience is always evolving and the understanding of the function of different brain regions is still under research.

REFERENCES

[1] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders: DSM-5* (Autor, Washington, DC, 2013), 5th ed.

[2] B. S. Khundrakpam, J. D. Lewis, P. Kostopoulos, F. Carbonell, and A. C. Evans, "Cortical thickness abnormalities in autism spectrum disorders through late childhood, adolescence, and adulthood: a large-scale MRI study," Cerebral Cortex **27**, 1721–1731 (2017).

[3] Z. Kovacs Balint, J. Raper, V. Michopoulos, L. Howell, C. Gunter, J. Bachevalier, and M. Sanchez, "Validation of the social responsiveness scale (srs) to screen for atypical social behaviors in juvenile macaques," Plos one **16**, e0235946 (2021).

[4] J. P. Leigh and J. Du, "Brief report: Forecasting the economic burden of autism in 2015 and 2025 in the united states," Journal of autism and developmental disorders **45**, 4135–4139 (2015).

[5] A. V. Buescher, Z. Cidav, M. Knapp, and D. S. Mandell, "Costs of autism spectrum disorders in the united kingdom and the united states," JAMA pediatrics **168**, 721–728 (2014).

[6] S. Brieber, S. Neufang, N. Bruning, I. Kamp-Becker, H. Remschmidt, B. Herpertz-Dahlmann, G. R. Fink, and K. Konrad, "Structural brain abnormalities in adolescents with autism spectrum disorder and patients with attention deficit/hyperactivity disorder," Journal of Child Psychology and Psychiatry **48**, 1251–1258 (2007).

[7] O. Dekhil, M. Ali, R. Haweel, Y. Elnakib, M. Ghazal, H. Hajjdiab, L. Fraiwan, A. Shalaby, A. Soliman, A. Mahmoud *et al.*, "A comprehensive framework for differentiating autism spectrum disorder from neurotypicals by fusing structural mri and resting state functional mri," in "Seminars in Pediatric Neurology," (Elsevier, 2020), p. 100805.

[8] M. Noriuchi *et al.*, "Altered white matter fractional anisotropy and social impairment in children with autism spectrum disorder," Brain research **1362**, 141–149 (2010).

[9] Y. ElNakieb, M. T. Ali, A. Elnakib, A. Shalaby, A. Soliman, A. Mahmoud, M. Ghazal, G. N. Barnes, and A. El-Baz, "The role of diffusion tensor mr imaging (dti) of the brain in diagnosing autism spectrum disorder: Promising results," Sensors **21**, 8171 (2021).

[10] F. Rafiee, R. Rezvani Habibabadi, M. Motaghi, D. M. Yousem, and I. J. Yousem, "Brain mri in autism spectrum disorder: Narrative review and recent advances," Journal of Magnetic Resonance Imaging **55**, 1613–1624 (2022).

[11] D.-Y. Song, C.-C. Topriceanu, D. C. Ilie-Ablachim, M. Kinali, and S. Bisdas, "Machine learning with neuroimaging data to identify autism spectrum disorder: a systematic review and meta-analysis," Neuroradiology **63**, 2057–2072 (2021).

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

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

[14] S. Risi, C. Lord, K. Gotham, C. Corsello, C. Chrysler, P. Szatmari, E. H. Cook Jr, B. L. Leventhal, and A. Pickles, "Combining information from multiple sources in the diagnosis of autism spectrum disorders," Journal of the American Academy of Child & Adolescent Psychiatry **45**, 1094–1103 (2006).

[15] A. American Psychiatric Association *et al.*, *Diagnostic and statistical manual of mental disorders*, vol. 3 (American Psychiatric Association Washington, DC, 1980).

[16] J. N. Constantino, S. A. Davis, R. D. Todd, M. K. Schindler, M. M. Gross, S. L. Brophy, L. M. Metzger, C. S. Shoushtari, R. Splinter, and W. Reich, "Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised," Journal of autism and developmental disorders **33**, 427–433 (2003).

[17] H. S. Nogay and H. Adeli, "Machine learning (ml) for the diagnosis of autism spectrum disorder (asd) using brain imaging," Reviews in the Neurosciences **1** (2020).

[18] S. Jacob, J. J. Wolff, M. S. Steinbach, C. B. Doyle, V. Kumar, and J. T. Elison, "Neurodevelopmental heterogeneity and computational approaches for understanding autism," Translational psychiatry **9**, 1–12 (2019).

[19] J. N. Constantino and C. P. Gruber, *Social responsiveness scale: SRS-2* (Western Psychological Services Torrance, CA, 2012).

[20] H. Cholemkery, J. Kitzerow, S. Rohrmann, and C. M. Freitag, "Validity of the social responsiveness scale to differentiate between autism spectrum disorders

and disruptive behaviour disorders," European child & adolescent psychiatry **23**, 81–93 (2014).

[21] T. Falkmer, K. Anderson, M. Falkmer, and C. Horlin, "Diagnostic procedures in autism spectrum disorders: a systematic literature review," European child & adolescent psychiatry **22**, 329–340 (2013).

[22] J. Hayes, T. Ford, R. McCabe, and G. Russell, "Autism diagnosis as a social process," Autism p. 13623613211030392 (2021).

[23] K. Gotham, A. Pickles, and C. Lord, "Trajectories of autism severity in children using standardized ados scores," Pediatrics **130**, e1278–e1284 (2012).

[24] P. Moridian, N. Ghassemi, M. Jafari, S. Salloum-Asfar, D. Sadeghi, M. Khodatars, A. Shoeibi, A. Khosravi, S. H. Ling, A. Subasi *et al.*, "Automatic autism spectrum disorder detection using artificial intelligence methods with mri neuroimaging: A review," arXiv preprint arXiv:2206.11233 (2022).

[25] A. Shoeibi, N. Ghassemi, M. Khodatars, P. Moridian, A. Khosravi, A. Zare, J. M. Gorriz, A. H. Chale-Chale, A. Khadem, and U. R. Acharya, "Automatic diagnosis of schizophrenia and attention deficit hyperactivity disorder in rs-fmri modality using convolutional autoencoder model and interval type-2 fuzzy regression," arXiv preprint arXiv:2205.15858 (2022).

[26] D. Li, H.-O. Karnath, and X. Xu, "Candidate biomarkers in children with autism spectrum disorder: a review of mri studies," Neuroscience bulletin **33**, 219–237 (2017).

[27] S. Ogawa, T.-M. Lee, A. R. Kay, and D. W. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation." proceedings of the National Academy of Sciences **87**, 9868–9872 (1990).

[28] P. A. Bandettini, E. C. Wong, R. S. Hinks, R. S. Tikofsky, and J. S. Hyde, "Time course epi of human brain function during task activation," Magnetic resonance in medicine **25**, 390–397 (1992).

[29] J. Belliveau, D. Kennedy, R. McKinstry, B. Buchbinder, R. Weisskoff, M. Cohen, J. Vevea, T. Brady, and B. Rosen, "Functional mapping of the human visual cortex by magnetic resonance imaging," Science **254**, 716–719 (1991).

[30] P. J. Basser, J. Mattiello, and D. LeBihan, "Mr diffusion tensor spectroscopy and imaging," Biophysical journal **66**, 259–267 (1994).

[31] A. L. Alexander, J. E. Lee, M. Lazar, R. Boudos, M. B. DuBray, T. R. Oakes, J. N. Miller, J. Lu, E.-K. Jeong, W. M. McMahon *et al.*, "Diffusion tensor imaging of the corpus callosum in autism," Neuroimage **34**, 61–73 (2007).

[32] V. Rajagopalan, Z. Jiang, J. Stojanovic-Radic, G. Yue, E. P. Pioro, G. Wylie, and A. Das, "A basic introduction to diffusion tensor imaging mathematics and image processing steps," Brain Disord Ther **6**, 2 (2017).

[33] A. A. K. A. Razek, K. A. Baky, and E. Helmy, "Diffusion tensor imaging in characterization of mediastinal lymphadenopathy," Academic Radiology **29**, S165–S172 (2022).

[34] A. A. K. A. Razek, E. M. Helmy, H. Maher, and M. A. Kasem, "Diffusion tensor imaging of the lateral rectus muscle in duane retraction syndrome," Journal of computer assisted tomography **43**, 467–471 (2019).

[35] M. Lazar, L. M. Miles, J. S. Babb, and J. B. Donaldson, "Axonal deficits in young adults with high functioning autism and their impact on processing speed," NeuroImage: Clinical **4**, 417–425 (2014).

[36] M. Hrdlicka, J. Sanda, T. Urbanek, M. Kudr, I. Dudova, S. Kickova, L. Pospisilova, M. Mohaplova, A. Maulisova, P. Krsek *et al.*, "Diffusion tensor imaging and tractography in autistic, dysphasic, and healthy control children," Neuropsychiatric Disease and Treatment **15**, 2843 (2019).

[37] C. Lord, M. Elsabbagh, G. Baird, and J. Veenstra-Vanderweele, "Autism spectrum disorder," The lancet **392**, 508–520 (2018).

[38] S. S. Joudar, A. Albahri, and R. A. Hamid, "Triage and priority-based healthcare diagnosis using artificial intelligence for autism spectrum disorder and gene contribution: A systematic review," Computers in Biology and Medicine p. 105553 (2022).

[39] M. Xu, V. Calhoun, R. Jiang, W. Yan, and J. Sui, "Brain imaging-based machine learning in autism spectrum disorder: methods and applications," Journal of neuroscience methods **361**, 109271 (2021).

[40] R. Bellman, "Dynamic programming," Science **153**, 34–37 (1966).

[41] B. Sen, N. C. Borle, R. Greiner, and M. R. Brown, "A general prediction model for the detection of adhd and autism using structural and functional mri," PloS one **13** (2018).

[42] S. Haar, S. Berman, M. Behrmann, and I. Dinstein, "Anatomical abnormalities in autism?" Cerebral cortex **26**, 1440–1452 (2016).

[43] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the abide dataset," NeuroImage: Clinical **17**, 16–23 (2018).

[44] K. Polat and S. Güneş, "A new feature selection method on classification of medical datasets: Kernel f-score feature selection," Expert Systems with Applications **36**, 10367–10373 (2009).

[45] N. S. Escanilla, L. Hellerstein, R. Kleiman, Z. Kuang, J. Shull, and D. Page, "Recursive feature elimination by sensitivity testing," in "2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)," (IEEE, 2018), pp. 40–47.

[46] M.-L. Huang, Y.-H. Hung, W. Lee, R.-K. Li, and B.-R. Jiang, "Svm-rfe based feature selection and taguchi parameters optimization for multiclass svm classifier," The Scientific World Journal **2014** (2014).

[47] G. R. Gaffney, L. Y. Tsai, S. Kuperman, and S. Minchin, "Cerebellar structure in autism," American journal of diseases of children **141**, 1330–1332 (1987).

[48] E. Courchesne, R. Yeung-Courchesne, J. Hesselink, and T. Jernigan, "Hypoplasia of cerebellar vermal lobules vi and vii in autism," New England Journal of Medicine **318**, 1349–1354 (1988).

[49] S. Eliez and A. L. Reiss, "Annotation: Mri neuroimaging of childhood psychiatric disorders: a selective review," Journal of Child Psychology and Psychiatry **41**, 679–694 (2000).

[50] R. Chen, Y. Jiao, and E. H. Herskovits, "Structural mri in autism spectrum disorder," Pediatric research **69**, 63–68 (2011).

[51] A. Mechelli, C. J. Price, K. J. Friston, and J. Ashburner, "Voxel-based morphometry of the human brain: methods and applications," Current Medical Imaging **1**, 105–113 (2005).

[52] J. G. Levitt, R. E. Blanton, S. Smalley, P. Thompson, D. Guthrie, J. T. McCracken, T. Sadoun, L. Heinichen, and A. W. Toga, "Cortical sulcal maps in autism," Cerebral Cortex **13**, 728–735 (2003).

[53] H. Ozgen, G. Hellemann, M. De Jonge, F. Beemer, and H. Van Engeland, "Predictive value of morphological features in patients with autism versus normal controls," Journal of autism and developmental disorders **43**, 147–155 (2013).

[54] M. Kijonka, D. Borys, K. Psiuk-Maksymowicz, K. Gorczewski, P. Wojcieszek, B. Kossowski, A. Marchewka, A. Swierniak, M. Sokol, and B. Bobek-Billewicz, "Whole brain and cranial size adjustments in volumetric brain analyses of sex- and age-related trends," Frontiers in Neuroscience **14** (2020).

[55] S. Zhang, X. Li, J. Lv, X. Jiang, L. Guo, and T. Liu, "Characterizing and differentiating task-based and resting state fmri signals via two-stage sparse representations," Brain imaging and behavior **10**, 21–32 (2016).

[56] Y. A. Elnakieb, M. T. Ali, A. Soliman, A. H. Mahmoud, A. M. Shalaby, N. S. Alghamdi, M. Ghazal, A. Khalil, A. Switala, R. S. Keynton *et al.*, "Computer aided autism diagnosis using diffusion tensor imaging," IEEE Access **8**, 191298–191308 (2020).

[57] P. J. Basser, J. Mattiello, and D. LeBihan, "Estimation of the effective self-diffusion tensor from the nmr spin echo," Journal of Magnetic Resonance, Series B **103**, 247–254 (1994).

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

[59] S. White, H. O'Reilly, and U. Frith, "Big heads, small details and autism," Neuropsychologia **47**, 1274–1281 (2009).

[60] S. Baron-Cohen, "Autism: the empathizing-systemizing (es) theory," Annals of the New York Academy of Sciences **1156**, 68–80 (2009).

[61] S. Baron-Cohen, "Empathizing, systemizing, and the extreme male brain theory of autism," in "Progress in brain research," , vol. 186 (Elsevier, 2010), pp. 167–175.

[62] M. F. Casanova, D. P. Buxhoeveden, A. E. Switala, and E. Roy, "Minicolumnar pathology in autism," Neurology **58**, 428–432 (2002).

[63] R. A. Carper and E. Courchesne, "Inverse correlation between frontal lobe and cerebellum sizes in children with autism," Brain **123**, 836–844 (2000).

[64] D. R. Hampson and G. J. Blatt, "Autism spectrum disorders and neuropathology of the cerebellum," Frontiers in neuroscience **9**, 420 (2015).

[65] C. J. Stoodley, "Distinct regions of the cerebellum show gray matter decreases in autism, adhd, and developmental dyslexia," Frontiers in systems neuroscience **8**, 92 (2014).

[66] A. Chaddad, C. Desrosiers, L. Hassan, and C. Tanougast, "Hippocampus and amygdala radiomic biomarkers for the study of autism spectrum disorder," BMC neuroscience **18**, 52 (2017).

[67] D. G. Amaral and B. A. Corbett, "The amygdala, autism and anxiety," in "Autism: Neural Basis and Treatment Possibilities," , G. Bock and J. Goode, eds. (Wiley, Chichester, 2003), pp. 177–197.

[68] C. M. Schumann, C. C. Barnes, C. Lord, and E. Courchesne, "Amygdala enlargement in toddlers with autism related to severity of social and communication impairments," Biological psychiatry **66**, 942–949 (2009).

[69] S. R. Dager, L. Wang, S. Friedman, D. Shaw, J. Constantino, A. Artru, G. Dawson, and J. Csernansky, "Shape mapping of the hippocampus in young children with autism spectrum disorder," American journal of neuroradiology **28**, 672–677 (2007).

[70] J. Piven, J. Bailey, B. J. Ranson, and S. Arndt, "No difference in hippocampus volume detected on magnetic resonance imaging in autistic individuals," Journal of autism and developmental disorders **28**, 105–110 (1998).

[71] L. K. Paul, C. Corsello, D. P. Kennedy, and R. Adolphs, "Agenesis of the corpus callosum and autism: a comprehensive comparison," Brain **137**, 1813–1829 (2014).

[72] H. Sharif and R. A. Khan, "A novel framework for automatic detection of autism: A study on corpus callosum and intracranial brain volume," arXiv preprint arXiv:1903.11323 (2019).

[73] N. Boddaert, N. Chabane, H. Gervais, C. Good, M. Bourgeois, M. Plumet, C. Barthelemy, M. Mouren, E. Artiges, Y. Samson *et al.*, "Superior temporal sulcus anatomical abnormalities in childhood autism: a voxel-based morphometry mri study," Neuroimage **23**, 364–369 (2004).

[74] C. W. Nordahl, D. Dierker, I. Mostafavi, C. M. Schumann, S. M. Rivera, D. G. Amaral, and D. C. Van Essen, "Cortical folding abnormalities in autism revealed by surface-based morphometry," Journal of Neuroscience **27**, 11725–11735 (2007).

[75] A. S. Nunes, V. A. Vakorin, N. Kozhemiako, N. Peatfield, U. Ribary, and S. M. Doesburg, "Atypical age-related changes in cortical thickness in autism spectrum disorder," Scientific Reports **10**, 1–15 (2020).

[76] P. R. Huttenlocher, "Morphometric study of human cerebral cortex development," Neuropsychologia **28**, 517–527 (1990).

[77] P. Rakic and D. Swaab, "Defects of neuronal migration and the pathogenesis of cortical malformations," in "Progress in brain research," , vol. 73 (Elsevier, 1988), pp. 15–37.

[78] D. C. Van Essen, "A tension-based theory of morphogenesis and compact wiring in the central nervous system," Nature **385**, 313–318 (1997).

[79] C. Ecker, A. Marquand, J. Mourão-Miranda, P. Johnston, E. M. Daly, M. J. Brammer, S. Maltezos, C. M. Murphy, D. Robertson, S. C. Williams *et al.*, "Describing the brain in autism in five dimensions—magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach," Journal of Neuroscience **30**, 10612–10623 (2010).

[80] M. Schuetze, M. T. M. Park, I. Y. Cho, F. P. MacMaster, M. M. Chakravarty, and S. L. Bray, "Morphological alterations in the thalamus, striatum, and pallidum in autism spectrum disorder," Neuropsychopharmacology **41**, 2627–2637 (2016).

[81] D. Y.-J. Yang, D. Beam, K. A. Pelphrey, S. Abdullahi, and R. J. Jou, "Cortical morphological markers in children with autism: a structural magnetic resonance imaging study of thickness, area, volume, and gyrification," Molecular autism **7**, 11 (2016).

[82] A. M. Pereira, B. M. Campos, A. C. Coan, L. F. Pegoraro, T. J. de Rezende, I. Obeso, P. Dalgalarrondo, J. C. da Costa, J.-C. Dreher, and F. Cendes, "Differences in cortical structure and functional mri connectivity in high functioning autism," Frontiers in neurology **9**, 539 (2018).

[83] M. C. Postema, D. Van Rooij, E. Anagnostou, C. Arango, G. Auzias, M. Behrmann, G. Busatto Filho, S. Calderoni, R. Calvo, E. Daly *et al.*, "Altered structural brain asymmetry in autism spectrum disorder in a study of 54 datasets," Nature communications **10**, 1–12 (2019).

[84] O. Dekhil, M. Ali, Y. El-Nakieb, A. Shalaby, A. Soliman, A. Switala, A. Mahmoud, M. Ghazal, H. Hajjdiab, M. F. Casanova *et al.*, "A personalized autism diagnosis cad system using a fusion of structural mri and resting-state functional mri data," Frontiers in Psychiatry **10** (2019).

[85] W. Yassin, H. Nakatani, Y. Zhu, M. Kojima, K. Owada, H. Kuwabara, W. Gonoi, Y. Aoki, H. Takao, T. Natsubori *et al.*, "Machine-learning classification using neuroimaging data in schizophrenia, autism, ultra-high risk and first-episode psychosis," Translational psychiatry **10**, 1–11 (2020).

[86] L. Kanner *et al.*, "Autistic disturbances of affective contact," Nervous child **2**, 217–250 (1943).

[87] C. Tye, A. K. Runicles, A. J. Whitehouse, and G. A. Alvares, "Characterizing the interplay between autism spectrum disorder and comorbid medical conditions: an integrative review," Frontiers in Psychiatry **9**, 751 (2019).

[88] J. Y. Kim, M. J. Son, C. Y. Son, J. Radua, M. Eisenhut, F. Gressier, A. Koyanagi, A. F. Carvalho, B. Stubbs, M. Solmi *et al.*, "Environmental risk factors and biomarkers for autism spectrum disorder: an umbrella review of the evidence," The Lancet Psychiatry **6**, 590–600 (2019).

[89] A. Kraneveld, K. Szklany, C. de Theije, and J. Garssen, "Gut-to-brain axis in autism spectrum disorders: Central role for the microbiome," International review of neurobiology **131**, 263–287 (2016).

[90] L. Shen, X. Liu, H. Zhang, J. Lin, C. Feng, and J. Iqbal, "Biomarkers in autism spectrum disorders: Current progress," Clinica Chimica Acta **502**, 41–54 (2020).

[91] E. Moradi, B. Khundrakpam, J. D. Lewis, A. C. Evans, and J. Tohka, "Predicting symptom severity in autism spectrum disorder based on cortical thickness measures in agglomerative data," Neuroimage **144**, 128–141 (2017).

[92] O. Dekhil, M. Ali, Y. El-Nakieb, A. Shalaby, A. Soliman, A. Switala, A. Mahmoud, M. Ghazal, H. Hajjdiab, M. F. Casanova *et al.*, "A personalized autism diagnosis cad system using a fusion of structural mri and resting-state functional mri data," Frontiers in psychiatry **10**, 392 (2021).

[93] M. T. Ali, Y. A. Elnakieb, A. Shalaby, A. Mahmoud, A. Switala, M. Ghazal, A. Khelifi, L. Fraiwan, G. Barnes, and A. El-Baz, "Autism classification using smri: A recursive features selection based on sampling from multi-level high dimensional spaces," in "2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)," (IEEE, 2021), pp. 267–270.

[94] J. Gao, M. Chen, Y. Li, Y. Gao, Y. Li, S. Cai, and J. Wang, "Multisite autism spectrum disorder classification using convolutional neural network classifier and individual morphological brain networks," Frontiers in Neuroscience **14**, 1473 (2021).

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

[96] C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B. S. Khundrakpam, J. D. Lewis, Q. Li, M. Milham *et al.*, "The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives," Neuroinformatics **4** (2013).

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

[98] F. Segonne, J. Pacheco, and B. Fischl, "Geometrically accurate topology-correction of cortical surfaces using nonseparating loops," IEEE Trans Med Imaging **26**, 518–529 (2007).

[99] M. Reuter, H. D. Rosas, and B. Fischl, "Highly accurate inverse consistent registration: A robust approach," NeuroImage **53**, 1181–1196 (2010).

[100] M. Reuter, N. J. Schmansky, H. D. Rosas, and B. Fischl, "Within-subject template estimation for unbiased longitudinal image analysis," NeuroImage **61**, 1402–1418 (2012).

[101] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).

[102] T. pandas development team, "pandas-dev/pandas: Pandas," (2020).

[103] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del R'ıo, M. Wiebe, P. Peterson, P. G'erard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," Nature **585**, 357–362 (2020).

[104] W. McKinney, "Data structures for statistical computing in python," pp. 56 – 61 (2010).

[105] M. Brett, C. J. Markiewicz, M. Hanke, M.-A. Côté, B. Cipollini, P. McCarthy, D. Jarecka, C. P. Cheng, Y. O. Halchenko, M. Cottaar, E. Larson, S. Ghosh, D. Wassermann, S. Gerhard, G. R. Lee, H.-T. Wang, E. Kastman, J. Kaczmarzyk, R. Guidotti, O. Duek, J. Daniel, A. Rokem, C. Madison, B. Moloney, F. C. Morency, M. Goncalves, R. Markello, C. Riddell, C. Burns, J. Millman, A. Gramfort, J. Leppäkangas, A. Sólon, J. J. van den Bosch, R. D. Vincent, H. Braun, K. Subramaniam, K. J. Gorgolewski, P. R. Raamana, J. Klug, B. N. Nichols, E. M. Baker, S. Hayashi, B. Pinsard, C. Haselgrove, M. Hymers, O. Esteban, S. Koudoro, F. Pérez-García, N. N. Oosterhof, B. Amirbekian, I. Nimmo-Smith, L. Nguyen, S. Reddigari, S. St-Jean, E. Panfilov, E. Garyfallidis, G. Varoquaux, J. H. Legarreta, K. S. Hahn, O. P. Hinds, B. Fauber, J.-B. Poline, J. Stutters, K. Jordan, M. Cieslak, M. E. Moreno, V. Haenel, Y. Schwartz, Z. Baratz, B. C. Darwin, B. Thirion, C. Gauthier, D. Papadopoulos Orfanos, I. Solovey, I. Gonzalez, J. Palasubramaniam, J. Lecher, K. Leinweber, K. Raktivan, M. Calábková, P. Fischer, P. Gervais, S. Gadde, T. Ballinger, T. Roos, V. R. Reddam, and freec84, "nipy/nibabel: 3.2.1," (2020).

[106] J. P. Lerch, A. J. van der Kouwe, A. Raznahan, T. Paus, H. Johansen-Berg, K. L. Miller, S. M. Smith, B. Fischl, and S. N. Sotiropoulos, "Studying neuroanatomy using mri," Nature neuroscience **20**, 314–326 (2017).

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

[108] B. Fischl, M. I. Sereno, and A. M. Dale, "Cortical surface-based analysis: Ii: inflation, flattening, and a surface-based coordinate system," Neuroimage **9**, 195–207 (1999).

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

[110] F. Ségonne, J. Pacheco, and B. Fischl, "Geometrically accurate topology-correction of cortical surfaces using nonseparating loops," IEEE transactions on medical imaging **26**, 518–529 (2007).

[111] B. Fischl, M. I. Sereno, R. B. Tootell, and A. M. Dale, "High-resolution intersubject averaging and a coordinate system for the cortical surface," Human brain mapping **8**, 272–284 (1999).

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

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

[114] B. Fischl and A. M. Dale, "Measuring the thickness of the human cerebral cortex from magnetic resonance images," Proceedings of the National Academy of Sciences of the United States of America **97**, 11050–11055 (2000).

[115] W. Zhang, W. Groen, M. Mennes, C. Greven, J. Buitelaar, and N. Rommelse, "Revisiting subcortical brain volume correlates of autism in the abide dataset: effects of age and sex," Psychological medicine **48**, 654–668 (2018).

[116] P. Coupé, G. Catheline, E. Lanuza, J. V. Manjón, and A. D. N. Initiative, "Towards a unified analysis of brain maturation and aging across the entire lifespan: A mri analysis," Human brain mapping **38**, 5501–5518 (2017).

[117] J.-M. Jo, "Effectiveness of normalization pre-processing of big data to the machine learning performance," The Journal of the Korea institute of electronic communication sciences **14**, 547–552 (2019).

[118] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," Pattern recognition **38**, 2270–2285 (2005).

[119] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," Machine learning **46**, 389–422 (2002).

[120] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "sklearn.feature_selection.rfecv¶," .

[121] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in "2010 20th International Conference on Pattern Recognition," (IEEE, 2010), pp. 3121–3124.

[122] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should I trust you?": Explaining the predictions of any classifier," in "Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016," (2016), pp. 1135–1144.

[123] G. J. Katuwal, N. D. Cahill, S. A. Baum, and A. M. Michael, "The predictive power of structural mri in autism diagnosis," in "2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)," (IEEE, 2015), pp. 4270–4273.

[124] M. R. Sabuncu, E. Konukoglu, A. D. N. Initiative *et al.*, "Clinical prediction from structural brain mri scans: a large-scale empirical study," Neuroinformatics **13**, 31–46 (2015).

[125] J. Cai, X. Hu, K. Guo, P. Yang, M. Situ, and Y. Huang, "Increased left inferior temporal gyrus was found in both low function autism and high function autism," Frontiers in psychiatry **9**, 542 (2018).

[126] F. Cauda, E. Geda, K. Sacco, F. D'Agata, S. Duca, G. Geminiani, and R. Keller, "Grey matter abnormality in autism spectrum disorder: an activation likelihood estimation meta-analysis study," Journal of Neurology, Neurosurgery & Psychiatry **82**, 1304–1313 (2011).

[127] J. Xu, C. Wang, Z. Xu, T. Li, F. Chen, K. Chen, J. Gao, J. Wang, and Q. Hu, "Specific functional connectivity patterns of middle temporal gyrus subregions in children and adults with autism spectrum disorder," Autism Research **13**, 410–422 (2020).

[128] D. H. Han, H. J. Yoo, B. N. Kim, W. McMahon, and P. F. Renshaw, "Brain activity of adolescents with high functioning autism in response to emotional words and facial emoticons," PLoS One **9**, e91214 (2014).

[129] Y. Yoshimura, M. Kikuchi, N. Hayashi, H. Hiraishi, C. Hasegawa, T. Takahashi, M. Oi, G. B. Remijn, T. Ikeda, D. N. Saito *et al.*, "Altered human voice processing in the frontal cortex and a developmental language delay in 3-to 5-year-old children with autism spectrum disorder," Scientific reports **7**, 1–11 (2017).

[130] A. Kobayashi, S. Yokota, H. Takeuchi, K. Asano, M. Asano, Y. Sassa, Y. Taki, and R. Kawashima, "Increased grey matter volume of the right superior temporal gyrus in healthy children with autistic cognitive style: A vbm study," Brain and Cognition **139**, 105514 (2020).

[131] T. C. Ramos, J. B. Balardin, J. R. Sato, and A. Fujita, "Abnormal cortico-cerebellar functional connectivity in autism spectrum disorder," Frontiers in systems neuroscience **12**, 74 (2019).

[132] Y. Lee, B.-y. Park, O. James, S.-G. Kim, and H. Park, "Autism spectrum disorder related functional connectivity changes in the language network in children, adolescents and adults," Frontiers in human neuroscience **11**, 418 (2017).

[133] X. Di and B. B. Biswal, "Similarly expanded bilateral temporal lobe volumes in female and male children with autism spectrum disorder," Biological Psychiatry: Cognitive Neuroscience and Neuroimaging **1**, 178–185 (2016).

[134] T. A. Knaus, A. M. Silver, K. C. Dominick, M. D. Schuring, N. Shaffer, K. A. Lindgren, R. M. Joseph, and H. Tager-Flusberg, "Age-related changes in the anatomy of language regions in autism spectrum disorder," Brain Imaging and Behavior **3**, 51–63 (2009).

[135] H. C. Hazlett, M. D. Poe, G. Gerig, R. G. Smith, and J. Piven, "Cortical gray and white brain tissue volume in adolescents and adults with autism," Biological psychiatry **59**, 1–6 (2006).

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

[137] D. Sarovic, N. Hadjikhani, J. Schneiderman, S. Lundström, and C. Gillberg, "Autism classified by magnetic resonance imaging: A pilot study of a potential diagnostic tool," International journal of methods in psychiatric research **29**, 1–18 (2020).

[138] Y. Jiao, R. Chen, X. Ke, K. Chu, Z. Lu, and E. H. Herskovits, "Predictive models of autism spectrum disorder based on brain regional cortical thickness," Neuroimage **50**, 589–599 (2010).

[139] K. N. Thakkar, F. E. Polli, R. M. Joseph, D. S. Tuch, N. Hadjikhani, J. J. Barton, and D. S. Manoach, "Response monitoring, repetitive behaviour and anterior cingulate abnormalities in autism spectrum disorders (asd)," Brain **131**, 2464–2478 (2008).

[140] G. J. Katuwal, S. A. Baum, and A. M. Michael, "Early brain imaging can predict autism: Application of machine learning to a clinical imaging archive," BioRxiv p. 471169 (2018).

[141] M. Andersson, Ä. Tangen, L. Farde, S. Bölte, C. Halldin, J. Borg, and J. Lundberg, "Serotonin transporter availability in adults with autism—a positron emission tomography study," Molecular Psychiatry pp. 1–12 (2020).

[142] J. S. Kohli, M. K. Kinnear, C. H. Fong, I. Fishman, R. A. Carper, and R.-A. Müller, "Local cortical gyrification is increased in children with autism spectrum disorders, but decreases rapidly in adolescents," Cerebral Cortex **29**, 2412–2423 (2019).

[143] Y. Huang, B. Zhang, J. Cao, S. Yu, G. Wilson, J. Park, and J. Kong, "Potential locations for noninvasive brain stimulation in treating autism spectrum disorders—a functional connectivity study," Frontiers in Psychiatry **11**, 388 (2020).

[144] A. L. Oblak, D. L. Rosene, T. L. Kemper, M. L. Bauman, and G. J. Blatt, "Altered posterior cingulate cortical cyctoarchitecture, but normal density of neurons and interneurons in the posterior cingulate cortex and fusiform gyrus in autism," Autism Research **4**, 200–211 (2011).

[145] C. S. Monk, S. J. Peltier, J. L. Wiggins, S.-J. Weng, M. Carrasco, S. Risi, and C. Lord, "Abnormalities of intrinsic functional connectivity in autism spectrum disorders," Neuroimage **47**, 764–772 (2009).

[146] A. Di Martino, D. O'connor, B. Chen, K. Alaerts, J. S. Anderson, M. Assaf, J. H. Balsters, L. Baxter, A. Beggiato, S. Bernaerts *et al.*, "Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii," Scientific data **4**, 1–15 (2017).

[147] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," arXiv preprint arXiv:2010.16061 (2020).

[148] A. Alksas, M. Shehata, G. A. Saleh, A. Shaffie, A. Soliman, M. Ghazal, A. Khelifi, H. A. Khalifeh, A. A. Razek, G. A. Giridharan *et al.*, "A novel computer-aided diagnostic system for accurate detection and grading of liver tumors," Scientific Reports **11**, 1–18 (2021).

[149] D. J. Hand, P. Christen, and N. Kirielle, "F*: an interpretable transformation of the f-measure," Machine Learning **110**, 451–456 (2021).

[150] C. J. Van Rijsbergen, *The geometry of information retrieval* (Cambridge University Press, 2004).

[151] D. A. Berry, "Biomarker studies and other difficult inferential problems: Statistical caveats," in "Seminars in oncology," , vol. 34 (Elsevier, 2007), vol. 34, pp. S17–S22.

[152] P. M. Altham, "Two generalizations of the binomial distribution," Journal of the Royal Statistical Society: Series C (Applied Statistics) **27**, 162–167 (1978).

[153] M. T. Ali, "Personalized Classification of Behavioral Severity of ASD," .

[154] A. D. M. M. P. M. Tanmay Nath, Bosi Chen, "Abide," (2016).

[155] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in "Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining," (2016), pp. 785–794.

[156] B. Deen, K. Koldewyn, N. Kanwisher, and R. Saxe, "Functional organization of social perception and cognition in the superior temporal sulcus," Cerebral cortex **25**, 4596–4609 (2015).

[157] V. Gallese, C. Keysers, and G. Rizzolatti, "A unifying view of the basis of social cognition," Trends in cognitive sciences **8**, 396–403 (2004).

[158] G. Rizzolatti, L. Craighero *et al.*, "The mirror-neuron system." (2004).

[159] Y. Zhou, L. Shi, X. Cui, S. Wang, and X. Luo, "Functional connectivity of the caudal anterior cingulate cortex is decreased in autism," PLoS One **11**, e0151879 (2016).

[160] M. S. Koyama, D. O'Connor, Z. Shehzad, and M. P. Milham, "Differential contributions of the middle frontal gyrus functional connectivity to literacy and numeracy," Scientific reports **7**, 1–13 (2017).

[161] S. Vanni, T. Tanskanen, M. Seppä, K. Uutela, and R. Hari, "Coinciding early activation of the human primary visual cortex and anteromedial cuneus," Proceedings of the National Academy of Sciences **98**, 2776–2780 (2001).

[162] M. G. Baxter, "Involvement of medial temporal lobe structures in memory and perception," Neuron **61**, 667–677 (2009).

[163] G. J. Blatt, "Inhibitory and excitatory systems in autism spectrum disorders," The Neuroscience of Autism Spectrum Disorders p. 335 (2012).

[164] K. Denys, W. Vanduffel, D. Fize, K. Nelissen, H. Peuskens, D. Van Essen, and G. A. Orban, "The processing of visual shape in the cerebral cortex of human and nonhuman primates: a functional magnetic resonance imaging study," Journal of Neuroscience **24**, 2551–2565 (2004).

[165] D. M. Clower, R. A. West, J. C. Lynch, and P. L. Strick, "The inferior parietal lobule is the target of output from the superior colliculus, hippocampus, and cerebellum," Journal of Neuroscience **21**, 6283–6291 (2001).

[166] H. Namkung, S.-H. Kim, and A. Sawa, "The insula: an underestimated brain area in clinical neuroscience, psychiatry, and neurology," Trends in neurosciences **40**, 200–207 (2017).

[167] F. Å. Nielsen, D. Balslev, and L. K. Hansen, "Mining the posterior cingulate: segregation between memory and pain components," Neuroimage **27**, 520–532 (2005).

[168] K. Grill-Spector, Z. Kourtzi, and N. Kanwisher, "The lateral occipital complex and its role in object recognition," Vision research **41**, 1409–1422 (2001).

[169] E. T. Rolls, W. Cheng, and J. Feng, "The orbitofrontal cortex: reward, emotion and depression," Brain Communications **2**, fcaa196 (2020).

[170] S. A. Kozlovskiy, M. M. Pyasik, A. V. Korotkova, A. V. Vartanov, J. M. Glozman, and A. A. Kiselnikov, "Activation of left lingual gyrus related to working memory for schematic faces," International Journal of Psychophysiology **2**, 241 (2014).

[171] A. D. Friederici, "White-matter pathways for speech and language processing," Handbook of clinical neurology **129**, 177–186 (2015).

[172] J. C. Tamraz, Y. G. Comair, and J. Tamraz, *Atlas of regional anatomy of the brain using MRI* (Springer, 2004).

[173] V. D. Bohbot, J. J. Allen, A. Dagher, S. O. Dumoulin, A. C. Evans, M. Petrides, M. Kalina, K. Stepankova, and L. Nadel, "Role of the parahippocampal cortex in memory for the configuration but not the identity of objects: converging evidence from patients with selective thermal lesions and fmri," Frontiers in human neuroscience **9**, 431 (2015).

[174] F. Binkofski and G. Buccino, "Motor functions of the broca's region," Brain and language **89**, 362–369 (2004).

[175] A. L. Foundas, C. M. Leonard, R. L. Gilmore, E. B. Fennell, and K. M. Heilman, "Pars triangularis asymmetry and language dominance." Proceedings of the National Academy of Sciences **93**, 719–722 (1996).

[176] E. M. Coppen, J. v. d. Grond, A. Hafkemeijer, J. J. Barkey Wolf, and R. A. Roos, "Structural and functional changes of the visual cortex in early huntington's disease," Human brain mapping **39**, 4776–4786 (2018).

[177] J. DiGuiseppi and P. Tadi, "Neuroanatomy, postcentral gyrus," in "StatPearls [Internet]," (StatPearls Publishing, 2021).

[178] A. E. Cavanna and M. R. Trimble, "The precuneus: a review of its functional anatomy and behavioural correlates," Brain **129**, 564–583 (2006).

[179] R. Leech, R. Braga, and D. J. Sharp, "Echoes of the brain within the posterior cingulate cortex," Journal of Neuroscience **32**, 215–222 (2012).

[180] S. Bissière, N. Plachta, D. Hoyer, K. H. McAllister, H.-R. Olpe, A. A. Grace, and J. F. Cryan, "The rostral anterior cingulate cortex modulates the efficiency of amygdala-dependent fear learning," Biological psychiatry **63**, 821–831 (2008).

[181] N. K. Schaal, B. Pollok, and M. J. Banissy, "Hemispheric differences between left and right supramarginal gyrus for pitch and rhythm memory," Scientific Reports **7**, 1–6 (2017).

[182] K. Tsapkini, C. E. Frangakis, and A. E. Hillis, "The function of the left anterior temporal pole: evidence from acute stroke and infarct volume," Brain **134**, 3094–3105 (2011).

[183] L. Fernández, C. Velásquez, J. A. G. Porrero, E. M. de Lucas, and J. Martino, "Heschl's gyrus fiber intersection area: a new insight on the connectivity of the auditory-language hub," Neurosurgical Focus **48**, E7 (2020).

[184] F. d. Boisgueheneuc, R. Levy, E. Volle, M. Seassau, H. Duffau, S. Kinkingnehun, Y. Samson, S. Zhang, and B. Dubois, "Functions of the left superior frontal gyrus in humans: a lesion study," Brain **129**, 3315–3328 (2006).

[185] P. Johns, *Clinical neuroscience* (Elsevier Health Sciences, 2014).

[186] A. P. Leff, T. M. Schofield, J. T. Crinion, M. L. Seghier, A. Grogan, D. W. Green, and C. J. Price, "The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: evidence from 210 patients with stroke," Brain **132**, 3401–3410 (2009).

[187] F. A. Mansouri, M. J. Buckley, M. Mahboubi, and K. Tanaka, "Behavioral consequences of selective damage to frontal pole and posterior cingulate cortices," Proceedings of the National Academy of Sciences **112**, E3940–E3949 (2015).

[188] L. J. Michalski, "Rostral middle frontal gyrus thickness is associated with perceived stress and depressive symptomatology," (2016).

APPENDIX A: PERMISSIONS

- Article Title: The Role of Structure MRI in Diagnosing Autism
  Journal: Diagnostics
  Volume: 21
  Pages: 165
  Year: 2022
  Publisher: MDPI

CURRICULUM VITAE

Mohamed T. Ali

Personal Information

E-mail: bioeng.m.tarek@gmail.com
LinkedIn: https://www.linkedin.com/in/mohamed-ali-9a35a656/
GoogleScholar: https://scholar.google.com/citations?user=LXdEtBQAAAAJ&hl=en

Education

Ph.D., BioEngineering, University of Louisville, Louisville, KY, May 2023

MSc., Systems & Biomedical Engineering, Cairo University, Giza, Egypt, August 2017

B.Sc., Systems & Biomedical Engineering, Cairo University, Giza, Egypt, May 2013

Professional and Work Experience

Graduate Research Assistant in the University of Louisville, Jan2018-present (Part-Time)

Research Development Software Engineer Intern, Cognex Corporation LLC, Natick, MA, May2022–Aug2022 (full-Time)

Data Science Scholar in The Data Incubator, April2021-May2021

Research and Development Engineer in DilenyTech, March2016-Dec2017 (Full-Time)

Graduate Research Assistant at Cairo University, March2015-March2016 (Full-Time)

Honors and Awards

- Best Graduate Student Peer-Reviewed Journal Paper from the School of Interdisciplinary and Graduate Studies, University of Louisville, 2022.
- Best Doctoral Journal Article Award from the School of Interdisciplinary and Graduate Studies, University of Louisville, 2021.
- Third place Doctoral Engineering Student Award at Research Louisville 2021.
- Student Champion 2020-2021.
- Dissertation Completion Award, 2023
- Exemplary Research Scholarship Award, 2023

Publications and Presentations

1. "*Using resting state functional mri to build a personalized autism diagnosis system.*", Co-author, Plos one 13, no10 (2018): e0206351.
2. "*A personalized autism diagnosis CAD system using a fusion of structural MRI and resting-state functional MRI data*", Co-author, Frontiers in psychiatry 10 (2019): 392.
3. "*The Role of Structure MRI in Diagnosing Autism.*", First author, Diagnostics 12.1 (2022): 165.
4. "*The Role of Diffusion Tensor MR Imaging (DTI) of the Brain in Diagnosing Autism Spectrum Disorder: Promising Results.*", co-author, Sensors 2021, 21, 8171. https://doi.org/10.3390/s21248171.
5. "*Extract image markers of autism using hierarchical feature selection technique.*", Book Chapter, Neural Engineering Techniques for Autism Spectrum Disorder. Academic Press, 2021. 333-343.
6. "*Autism Classification Using SMRI: A Recursive Features Selection Based on Sampling from Multi-Level High Dimensional Spaces.*", Co-author, 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI).
7. "*A Personalized Autism Diagnosis CAD System Using a Fusion of Structural MRI and Resting-State Functional MRI Data.*", Co-author, Frontiers in psychiatry 10 (2021): 392.
8. "*Computer Aided Autism Diagnosis Using Diffusion Tensor Imaging.*", First author, IEEE Access 8 (2020), doi: 10.1109/ACCESS.2020.3032066.
9. "*A Comprehensive Framework for Differentiating Autism Spectrum Disorder from Neurotypicals by Fusing Structural MRI and Resting State Functional MRI.*", Co-author, Seminars in Pediatric Neurology, 2020.

10. "*Autism Spectrum Disorder Diagnosis framework using Diffusion Tensor Imaging.*", co-author, 2019 IEEE International Conference on Imaging Systems and Techniques (IST).

11. "*Towards Accurate Personalized Autism Diagnosis Using different Imaging Modalities: sMRI, fMRI, and DTI.*", Co-author, IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2018).

12. "*Towards Personalized Autism Diagnosis: Promising Results.*", Co-author, IEEE Conference on Computer Vision and Pattern Recognition, (ICPR 2018).

13. *Ten abstracts accepted*, INSAR2022, four at INSAR2020, BMES2021, BMES2020, BMES2019, BMES2018, and NRSC 2016.