Classification and Biomarker Identification In Autism Using Conjunctive Clause Evolutionary Algorithm

Yu Han *[1], Patricia A. Prelock[1], John P. Hanley[2], Emily L. Coderre[1], and Donna M. Rizzo[3]

[1]Department of Communication Sciences and Disorders, University of Vermont, Burlington, VT, USA
[2]Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, VT, USA
[3]Department of Civil and Environmental Engineering, University of Vermont, Burlington, VT, USA

Abstract

Autism spectrum disorder (ASD) is a developmental disability that can cause significant social, communication and behavioral challenges. Many challenges remain for ASD diagnosis and treatments. Current diagnostic criteria are based on behavioral symptoms alone. Wait times for appointments at diagnostic centers range from 9 to 13 months, and a single diagnostic appointment can last several hours. There is an urgent need to identify ASD associated biomarkers and features to help automate diagnostics and develop predictive ASD models. The present study adopts a novel evolutionary algorithm, the conjunctive clause evolutionary algorithm (CCEA), to select features most significant for distinguishing individuals with and without ASD using a unique dataset having a small number of samples with a very large number of feature measurements. The dataset comprises both behavioral and neuroimaging measurements from a total of 21 children from 7 to 14 years old. Potential biomarker candidates including volume, area, cortical thickness and mean curvature in specific regions in the cingulate cortex, frontal cortex and temporal-parietal junction areas were identified. Behavioral features associated with theory of mind were selected. Study findings demonstrate how machine learning tools can advance ASD research in the genre of big data to benefit this special population in the future.

Keywords: Autism spectrum disorder, evolutionary algorithm, neuroimaging

1 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by impairments in social interaction, and restricted and repetitive behaviors [1–4]. According to the Center for Disease Control and Prevention (CDC) 2018 report, the number of U.S. children diagnosed with ASD has increased from 1 in every 150 in year 2000 to about 1 in every 54 in year 2016 [5]. The economic burden of pediatric ASD is due substantially to the costs associated with an increased use of health services, school support, ASD-related therapy, family services, and caregiver time. Total societal costs in the United States for children with ASD were estimated at $11.5 billion in year 2011 [6, 7]. While genetic and environmental factors have been linked to the development of ASD; at present, there is no identified cause or cure for ASD.

Some symptoms of ASD are not evident until age two or later. In fact, a child may appear to be developing normally until the age of two when they stop learning new skills and may even forget old skills [8, 9]. Currently, the diagnosis of autism is based on behavioral symptoms alone. There are two common behavioral assessment tools guiding the diagnostic process, *The Autism Diagnostic Observation Schedule-second edition (ADOS-2)* and *The Autism Diagnostic Interview-revised (ADI-R)* [10, 11]. However, a typical diagnostic appointment consists of evaluations lasting several hours at a designated clinical office. Due to the rigorous and time-consuming nature of ASD diagnostic examinations, the demand exceeds the capacity to see patients. As a result, many diagnostic centers have expanding wait lists for appointments. This bottleneck translates to delays in diagnosis of 13 months and longer for minority and lower socio-economic status groups

*yu.han@uvm.edu

[2, 8, 12–18]. It is also believed that a substantial number of individuals on the spectrum remain undetected [19]. With growing awareness of ASD, there is a high demand for a faster and automated ASD diagnostic approach that might allow for more efficient diagnosis and early identification of high-risk populations [20].

Building an automated diagnostic and predictive model of ASD is timely as many studies have adopted machine learning approaches to identify significant biomarkers that include both behavioral and biological features. Duda and colleagues (2016) applied machine learning to distinguish ASD from attention deficit hyperactivity disorder (ADHD) using a 65-item *Social Responsiveness Scale* [21]. Bone et al. (2015) trained their models to diagnose autism against healthy controls using the same *Social Responsiveness Scale* and the *Autism Diagnostic Interview-Revised score* [22]. Other studies aggregated items from the ADOS and scores from the *Autism Quotient (AQ)* to accurately classify an ASD group. However, behavioral measures may be interpreted as subjective, and there can be a wide range of select features depending on which tests are used. Consequently, it becomes important to identify consistent markers associated with ASD.

As a result of the wide range in ASD behavioral measures and their subjective nature, many studies have searched for brain-based biological markers to identify a common etiology across individuals with ASD. These brain-based biological markers are less subjective than behavioral measures and may represent potential targets for treatments. Currently, markers that are measurable via magnetic resonance imaging (MRI) are highly desirable because they may represent potential targets for both treatments and diagnostic tools [23]. Independent structural MRI studies have found differences in whole brain volume and the developmental trajectories between individuals with ASD and those without ASD [24–33]. Other structural brain abnormalities associated with ASD include cortical folding signatures appearing in the following regions of the brain: temporal-parietal junction, anterior insula, posterior cingulate, lateral and medial prefrontal, corpus callosum, intra-parietal sulcus, and the occipital cortex [24–33]. Evidence also shows that an accelerated expansion of cortical surface area, but not cortical thickness, causes an early overgrowth of the brain in children with ASD [34], while other studies suggest that individuals with ASD tend to have thinner cortices and reduced surface area as an effect of aging [35].

Machine learning (ML) has been introduced to the neuroimaging field to identify the abnormal brain regions in individuals with ASD. The support vector machine (SVM) is an algorithm that avoids over-fitting and is capable of high classification accuracy without requiring large sample sizes. The SVM algorithm is able to classify ASD from corresponding controls using extracted features from functional connections and grey matter volume [36–40]. Other ML classifications of ASD include deep neural networks [41] and the random forests (RF) algorithm; the latter uses random ensembles of independently grown decision trees [42]. Although these methods have demonstrated high accuracy for classifying ASD, they have not identified precise neuroimaging-based biomarkers and features associated with ASD. The majority of studies have adopted data from the *Autism Brain Imaging Data Exchange (ABIDE)* dataset that includes 1112 existing resting-state functional magnetic resonance (rs-fMRI) imaging datasets with corresponding structural MRI and phenotypic information from 539 individuals with ASD and 573 age-matched typical controls collected from 24 international brain imaging laboratories [43].

Classification across a heterogeneous population is challenging [44, 45], particularly when neuroimaging data are pooled from multiple acquisition sites (e.g., the ABIDE dataset, which has considerable variation in demographic and phenotypic profiles). Variances introduced in the data due to scanner hardware, imaging protocols, operator characteristics, regional demographics, and other factors that are acquisition site-specific, can affect the classification performance. This problem is especially relevant for ASD given the inherent heterogeneity of the population. It is often difficult to collect neuroimaging data from individuals

with autism given the loudness of the scanner and difficulties remaining still. In fact, most individual site datasets have small sample sizes that can lead to over-fitting and classification inaccuracies. Moreover, while many traditional ML algorithms were designed to classify large amounts of data (e.g., ABIDE) rather than optimize the selection of features, the ultimate goal for ML-based diagnostic classification in neuroimaging is to identify discriminative features to provide insight into abnormal structure and dysfunctional connectivity patterns in the affected population [46].

The present study employed an evolutionary algorithm, the conjunctive clause evolutionary algorithm (CCEA), to select features most correlated with classifying individuals with and without ASD. The dataset has a relatively large number of features (both behavioral and neuroimaging measurements) from a total of 26 children, which comprises a training set (7 children with ASD and 14 neurotypical (NT) children), and a testing set (1 child with ASD and 4 NT). Children in the testing set were enrolled as a cohort at a later time after the CCEA was trained.

The neuroimaging measurements included brain volume, brain surface area, cortical thickness, and cortical curvature extracted from MRI whole brain T1 weighted scans. Behavioral measurements included scores from the *Comprehensive Assessment of Spoken Language (CASL)* [47], the *Universal Nonverbal Intelligence Test-2 (UNIT-2)* [48, 49], the *Theory of Mind Task Battery (ToMTB)* [50] and the *Theory of Mind Inventory-2 (ToMI-2)* [50]. The present study examined the validity of using the CCEA algorithm for feature selection in ASD, particularly to address challenges associated with traditional ML algorithms when working with small datasets. It aims to identify discriminative biomarkers and behavioral features to help develop an automatic diagnostic and predictive system for ASD.

Although some ML-based methods have been applied to ASD, the suitability of machine learning and the choice of algorithms with regard to the specific behavior examined, as well as the quality and quantity of the data obtained from individual studies, needs further investigation [51]. We believe this study is the first study to:

- Classify ASD and select discriminative biomarkers among children from 7 to 14 years old.
- Include both behavioral and biological measurements in the feature selection model.
- Identify models (sets of features) that most strongly correlate to children with ASD given a dataset with a relatively small sample size (i.e., N=26) and large number of features (i.e., 247 neuroimaging features and 14 behavioral features).

2 Methods

2.1 Participants

A total of 8 children with ASD (*1 female, mean age = 11*) and 18 NT children (*7 female, mean age = 10.28*) were enrolled in the study. All children participated in 2-3 hours of baseline behavioral assessments in which both groups completed the CASL, UNIT-2, ToMTB, and ToMI-2,while the ASD group also completed the ADOS-2 and the Social Communication Questionnaire-Lifetime version (SCQ) [52] to confirm their ASD diagnosis. All ASD children demonstrated understanding of the instructions given in the behavioral and functional magnetic resonance imaging (fMRI) tasks.

2.2 Behavioral Measurements

The CASL is an orally administered research-based assessment consisting of 15 subtests measuring language for individuals ranging from 3 to 21 years of age. For the present study, only those basic subsets that establish the CASL language core are used: Antonyms, Sentence Completion, Syntax Construction, Paragraph Comprehension, and Pragmatic Judgment. The UNIT-2 is a multidimensional assessment of intelligence for individuals with speech, language, or hearing impairments. It consists of nonverbal tasks that test symbolic memory, non-symbolic quantity, analogic reasoning, spatial memory, numerical series, and cube design.

The ToMTB and ToMI-2 are two norm-referenced tools and behavioral tasks used as outcome measures to assess theory of mind (ToM) [53, 54]. ToM is the ability to reason about the thoughts and feelings of self and others, including the ability to predict what others will do or how they will feel in a given situation on the basis of their inferred beliefs. ToM is a core social deficit in ASD. Scores from both ToMTB and ToMI-2 provide valid representations of a child's social cognition level. The ToMI-2 is a parent-informant measure of a child's functional level of ToM. Each of the 60 items assesses a particular ToM dimension using items that range from simple content to those that evaluate more complex skills. Each item is rated on a 20-unit continuous scale anchored by "Definitely Not" and "Definitely." Respondents indicate their response with a vertical hash mark at the point on the scale that best reflects their attitude. Item, subscale, and composite scores range from 0-20 with higher values reflecting greater parental confidence that the child possesses a particular ToM skill. The ToMI-2 is designed to be a socially and ecologically valid index of ToM as it occurs in everyday social interactions. It has demonstrated excellent test-retest reliability, internal consistency, and criterion-related validity for both NT and ASD children as well as contrasting-group validity and statistical evidence of construct validity (i.e., factor analysis). The ToMTB directly assesses a child's understanding of a series of scenarios tapping theory of mind. It consists of 15 test questions within nine tasks, arranged in ascending difficulty. Tasks are presented as short vignettes that appear in a story-book format. Each page has color illustrations and accompanying text. For all tasks, children are presented with one correct response option and three plausible distracters. Memory control questions are included that must be passed for credit on the test questions. The ToMTB has strong test-retest reliability [55–57].

We included the total score of the CASL, full scale score of the UNIT-2, abbreviated score of the UNIT-2, total score of the ToMTB, total composite mean of the ToMI-2 (i.e., assessing overall ToM ability), early subscale mean of the ToMI-2 (i.e., assessing early developing ToMI ability such as regulating desire-based emotion and recognition of happy and sad), basic subscale mean of the ToMI-2 (i.e., assessing basic ToM ability such as recognition of surprise), advanced subscale mean of the ToMI-2 (assessing advanced ToM ability such as recognition of embarrassment) in the CCEA algorithm for feature selection. We also included scores from single items assessing recognition of simple emotions such as happy and sad in the model, as well as more complex emotions such as surprise and embarrassment, which ASD children often find difficult to recognize and process [58–62]. There were 13 behavioral features in total. Table 1 provides demographic information for all participants (i.e., N represents NT subject and A represents ASD subject) including their age, gender and scores on the 13 behavioral tests. Results from T-tests found significant differences (i.e. $p<0.05$) between the NT group and the ASD group on scores of CASL, UNIT-2 full scale, ToMTB total, ToMI-2 total, ToMI-2 early subscale, ToMI-2 basic subscale, ToMI-2 advanced subscale, ToMI-2 embarrassment and ToMI-2 desire based, where NT subjects scored higher.

Table 1: Demographic Summary Table: NT vs. ASD

Subject	Gender	CASL	UNIT-2 Full Scale	UNIT-2 Abbreviation	ToMTB Total	ToMI-2 Total	ToMI-2 Early Subscale	ToMI-2 Basic Subscale	ToMI-2 Advanced Subscale	ToMI-2 Happy	ToMI-2 Sad	ToMI-2 Surprise	ToMI-2 Embarrassment	ToMI-2 Desire-Based
N1	M	112	130	121	11	15.6	17.4	16	14	17.1	19.3	19.4	16.2	17.4
N2	M	89	110	109	14	15.1	15.3	16.7	13.4	15.8	12.8	14.4	6.5	16.4
N3	M	112	105	91	13	18.6	18.4	18.8	18.4	20	20	20	20	20
N4	M	126	123	121	13	18.8	18.8	18.8	18.7	20	19.9	20	20	20
N5	M	117	122	118	14	14.3	15.9	15.6	12.1	17.4	15.5	16	12.6	16.3
N6	M	114	101	91	15	13.6	14.1	16.8	10	15	7.1	12.2	7	17.3
N7	F	115	110	118	14	18.9	19.4	19.1	18.3	18.4	19	19	16.7	19.5
N8	M	106	113	124	15	18.5	18.6	19.7	17.1	20	13.6	18.4	15	20
N9	F	132	102	91	11	18.9	19.2	19	18.6	19.4	20	19.3	18.3	19
N10	F	121	103	94	15	17.8	17.1	18.9	17.2	18.5	18.8	17.2	17	19
N11	F	110	106	121	13	14.7	16.2	16.2	12.4	18.4	18.9	19.4	14.9	19.4
N12	M	97	100	106	11	18.4	19.6	19.1	16.9	20	20	20	20	20
N13	M	110	107	115	14	18.1	17.2	19.2	17.5	20	17.5	20	10	20
N14	F	110	117	124	14	18.1	19	18.8	17	20	20	19.9	17.4	19.9
N15	M	115	110	109	12	19.3	20	19.6	18.6	20	20	15	17.8	20
N16	F	102	105	100	13	18.2	18.7	18.4	17.6	19.3	17.9	17.4	18.3	20
N19	F	112	113	124	12	18.7	18	19.3	18.5	17.5	17	15.3	17.4	19.9
N20	M	122	135	124	15	11.5	13.8	13.2	8.3	14.1	15.5	14.4	5.8	15
A1	M	97	121	121	12	11.2	11.6	13	9.1	14.9	13.1	12.7	11	15.3
A3	M	40	64	73	5	9.1	15	8.5	6.3	20	20	15	9.9	14.8
A5	F	91	94	97	9	9.4	12.5	10.9	6	18.6	14.8	11	4.8	14.8
A6	M	121	127	118	13	17.6	16.9	19.2	16.4	19.9	20	20	8.4	20
A7	M	105	110	118	13	15.1	14.3	17.2	13.6	15.1	14.8	10.3	10	15.1
A8	M	55	93	106	11	12.7	16.2	14.8	8.6	20	20	19.9	10.2	20
A9	M	63	96	112	13	13.7	17.9	17	7.8	20	20	20	20	20
A11	M	50	56	67	10	10.4	13.6	12.8	6	15.3	14.7	14.8	5	14.6

2.3 MRI Acquisition and Preprocessing

All data were acquired using the MRI Center for Biomedical Imaging 3T Philips Achieva dStream scanner and 32-channel head coil at the University of Vermont (UVM). Parameters for T1 acquisition are TR 800ms, TE 30ms, flip angle 52 degree, 2.4mm isotropic imaging resolution with a $216 \times 216 \times 144mm^3$ field of view using a multiband acceleration factor of 6 (60 slices, no gap). Participants watched three videos at home before coming to the MRI center. The first was a cartoon video explaining what an MRI is, and what one might experience while laying in an MRI scanner [63]. The second video, recorded at the UVM MRI mock scanner room, helped visualize the real setting and procedures a child would experience. The third video explained the procedures of wearing earplugs. All participants practiced laying still and became familiar with the scanner noise in the mock scanner room. The T1 structural scan was preprocessed using the Human Connectome Project (HCP) minimal preprocessing pipelines, including spatial artifact/distortion removal, surface generation, cross-modal registration, and alignment to standard space. These pipelines are specially designed to capitalize on the high quality data offered by the HCP. The final standard space makes use of a recently introduced CIFTI file format and the associated grayordinates spatial coordinate system. This allows for combined cortical surface and subcortical volume analyses while reducing the storage and processing requirements for high spatial and temporal resolution data [64]. Brain anatomical features are extracted using FreeSurfer aparcstats2tabl script [65], including volume, cortical thickness, mean curvature, and area of all ROIs for each subject. These ROIs are defined using the automatic segmentation procedures that assign one of 37 labels to each brain voxel, including left and right caudate, putamen, pallidum, thalamus, lateral ventricles, hippocampus, and amygdala [66]. There are 276 brain features included in total.

2.4 Conjunctive Clause Evolutionary Algorithm

We used a novel evolutionary algorithm to identify the features associated with ASD. The CCEA is a machine learning tool that searches for both the combinations of features associated with a given category (e.g., ASD) as well as their corresponding range of values [67]. The CCEA is capable of finding feature interactions even

in the absence of main-effects, and can, therefore, find feature combinations that would be difficult to discover using traditional statistics. The CCEA selects for the best conjunctive clauses (CC) of the form:

$$CC_k = F_i \in a_i \wedge F_j \in a_j ...,$$

where F_i represents a risk factor i whose value lies in the range a_i; and the symbol $\wedge$ represents a conjunction (i.e., logical AND). The benefit of the CCEA is that it produces parsimonious models that are correlated with a select category (e.g., ASD). The model parsimony is measured using the order of the conjunctive clause, which is the total number of features in the conjunctive clause. One example of a parsimonious 2nd-order conjunctive clause is: a person with a right hemisphere isthmus cingulate volume of 3,300 – 4,100 AND a right hemisphere posterior cingluate volume of 4,100 – 6,200 is more likely to have ASD than someone who does not meet these criteria.

The fitness of each conjunctive clause (CC) is evaluated using the hypergeometric probability mass function (PMF) and only the most-fit conjunctive clauses are saved. The hypergeometric PMF is not a p-value and thus, is not constrained by issues associated with what threshold is "significant" [68–70]. To prevent overfitting, the CCEA performs feature sensitivity on each conjunctive clause to ensure each feature contributes to the overall fitness. For each feature in a conjunctive clause, the sensitivity is calculated by taking the difference between the conjunctive clause fitness and the fitness when that feature is removed. Thus, a feature's sensitivity may be viewed as the amount of fitness that it contributes to the conjunctive clause. To visualize the fitness landscape, both positive predictive value and coverage are calculated. Positive predictive value (PPV) is the number of true positives divided by the sum of true and false positives; and class coverage is the number of true positives divided by the sum of true positives and false negatives (i.e., the percent of ASD individuals that match the conjunctive clause). In this work, the CCEA was run five times using the training set to ensure a more thorough search of the fitness landscape.

3 Results

3.1 Training Set: 7 ASD and 14 NT

In the training set, 2438 CCs (i.e., models) were generated ranging from first-order to fifth-order. The PPV of the 2438 models range anywhere from 46.47% to 100% and their class coverage ranges from 42.86% to 100%. Among these models, we looked for the most parsimonious (i.e., lower order models) to draw meaningful conclusions and to avoid overfitting, which exist with higher-order models. As a result, we selected 8 second-order models (i.e., those having only two features) and the highest fitness (PMF) among the total 520 second-order models. These 8 best performing models have 100% PPV and 100% class coverage. All of the features identified have only brain anatomical features (Table 2). Because of our desire to examine the behavioral features, we expanded our analysis to include third-order models (i.e., model combinations with three features). There were 651 third-order models in total; some consisted only of anatomical brain features, while others had two behavioral features plus one brain anatomical feature. We selected the six best performing third-order models with the highest fitness (PMF); each had 100% PPV and 100% class coverage. Each of these third-order models (Table 3) contained two behavioral features and one brain anatomical feature.

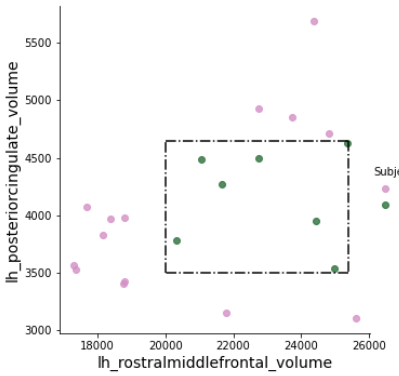
Table 2: Second-Order CCs

CC#	Feature One	Value	Feature Two	Value
113	lh posteriorcingulate volume	[3500, 4600]	lh rostralmiddlefrontal volume	[20000, 25000]
679	lh posteriorcingulate volume	[3500, 4600]	rh rostralmiddlefrontal volume	[21000, 26000]
1449	lh posteriorcingulate volume	[3500, 4600]	lh medialorbitofrontal thickness	[2.6, 2.8]
2199	lh posteriorcingulate volume	[3500, 4600]	rh rostralmiddlefrontal area	[6400, 8500]
887	rh isthmuscingulate volume	[3300, 4100]	rh posteriorcingulate volume	[4100, 6200]
1488	rh isthmuscingulate volume	[3300, 4100]	lh medialorbitofrontal thickness	[2.6, 2.8]
2200	rh isthmuscingulate volume	[3300, 4100]	rh rostralmiddlefrontal area	[6400, 8500]
2269	rh isthmuscingulate volume	[3300, 4100]	rh posteriorcingulate area	[1400, 2900]

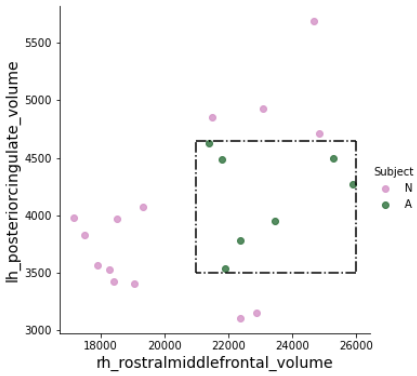
Table 3: Third-Order CCs

CC#	Feature One	Value	Feature Two	Value	Feature Three	Value
46	ToMTB total	[5, 13]	ToMI early subscale mean	[12, 18]	lh parsorbitalis meancurv	[0.17, 0.2]
191	ToMTB total	[5, 13]	ToMI early subscale mean	[12, 18]	rh superiorparietal thickness	[2.4, 2.7]
264	ToMTB total	[5, 13]	ToMI early subscale mean	[12, 18]	rh parsorbitalis meancurv	[0.18, 0.2]
317	ToMTB total	[5, 13]	ToMI early subscale mean	[12, 18]	rh inferiortemporal meancurv	[0.15, 0.18]
1163	ToMTB total	[5, 13]	ToMI total composite mean	[9, 18]	lh postcentral thickness	[2.2, 2.5]
1446	ToMTB total	[5, 13]	ToMI early subscale mean	[12, 18]	lh medialorbitofrontal thickness	[2.6, 2.8]

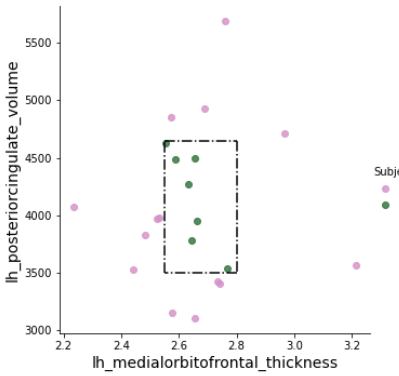
Selected second-order models have 100% PPV and 100% class coverage. These models, using CC 113 (Table 2) as an example, would be interpreted as - any subjects whose posterior cingulate gyrus volume was within the range of 3500 to 4600 mm^3 AND left rostral middle frontal gyrus volume was within the range of 20,000 to 25,000 mm^3 would be classified as having ASD. The volume of the left hemisphere posterior cingulate gyrus and the volume of the right hemisphere isthmus of the cingulate gyrus were the two features to appear most frequently (i.e., four times) across all models, suggesting that the volume of cingulate gyrus is a potentially important biomarker for ASD. Figure 1 provides a 2D visualization for the range of feature values (numerical boundaries) associated with these models and the placement of each subject within this range. Note: Green dots represent ASD subjects and group together within the rectangle defining the range of values in Table 2.



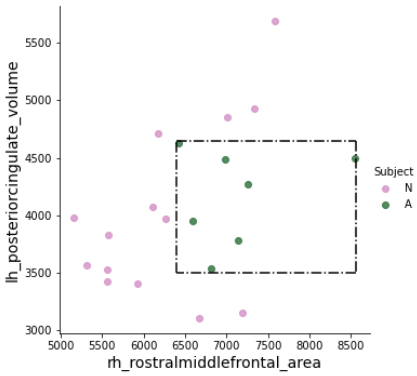
(a) CC113



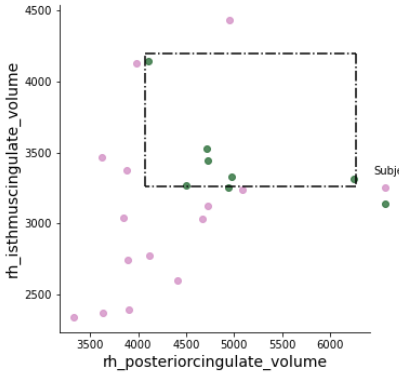
(b) CC679



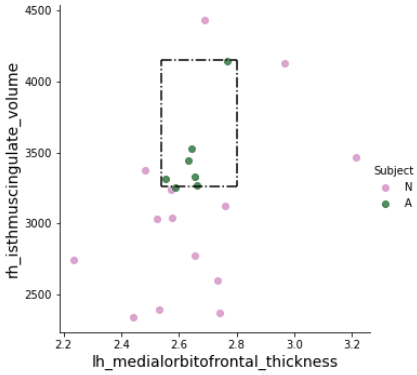
(c) CC1449



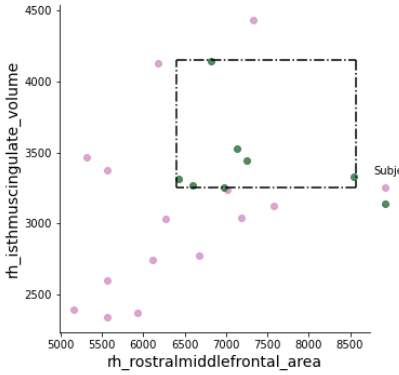
(d) CC2199



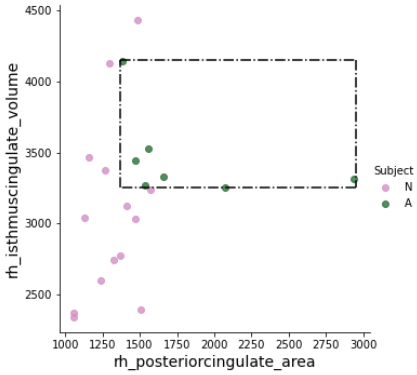
(e) CC887



(f) CC1488



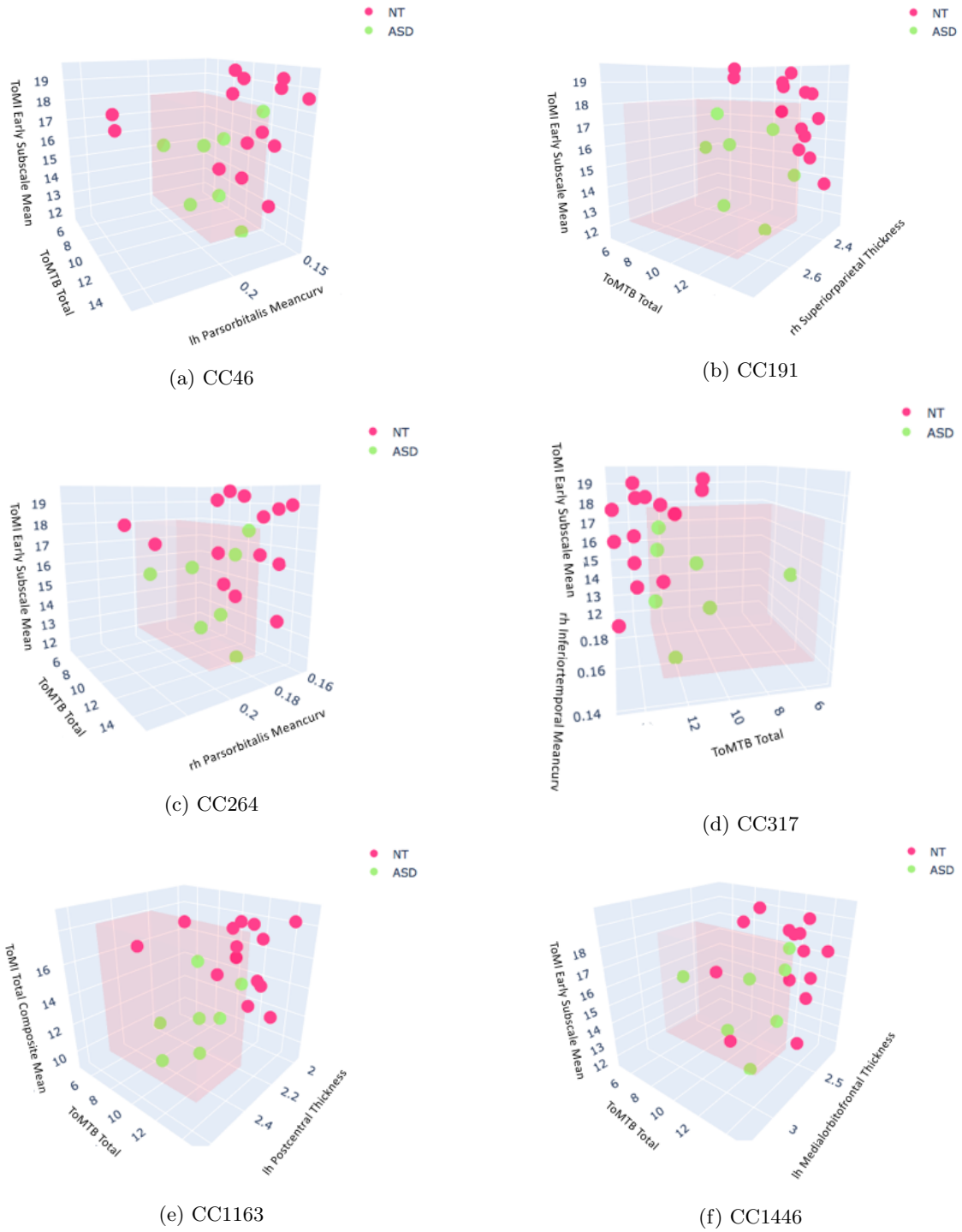
(g) CC2200



(h) CC2269

Figure 1: Second-Order CC 2D Visualization

Selected third-order models have 100% PPV and 100% class coverage. Using CC 46 (Table 3) as an example, any subjects who had a total score on ToMTB within the range of 5 to 13 AND an early subscale mean score on ToMI-2 within the range of 12 to 18 AND a mean curvature value of the left hemisphere pars orbitalis within the range of 0.17 to 2 would fall in the ASD class. The ToMTB total score feature occurred in all of our best fit third-order models; and the ToMI-2 early subscale mean score occurred in all but one (CC 1163) of the models, where the ToMI-2 total composite mean played a role. Such a finding further suggests that ToMTB and ToMI-2 might be effective for ASD testing and diagnosis tools. Figure 2 is a 3D visualization of the CC feature value boundaries and classification placement of each subject, where green dots represent ASD subjects and group together within the pink cube defined by feature values in Table 3.



(a) CC46

(b) CC191

(c) CC264

(d) CC317

(e) CC1163

(f) CC1446

Figure 2: Third-Order CC 3D Visualization

3.2 Testing Set: 1 ASD and 4 NT

A cohort of new subjects comprising 1 ASD child and 4 NT children were recruited separately at a later time. This later cohort served as a testing set to validate the original 2438 models generated in the training set. Three third-order models (Table 4) and four fourth-order models (Table 5) were selected from the total 2438 models generated using the training set, as these 7 models were the only models that remained 100% PPV and 100% class coverage on the testing set. Among the important features selected for the testing set, cortical thickness of the left hemisphere pericalcarine cortex and mean curvature of the right hemisphere pars orbitalis were the two features to appear most frequently across all models, indicating the important roles of these areas in ASD. The inclusion of this testing provided an additional observations to examine the robustness of the best-fit models identified using the original training set (i.e., parsimonious models with 100% PPV and 100% class coverage).

Table 4: Third-Order CCs

Feature One	Value	Feature Two	Value	Feature Three	Value
lh superioparietal area	[5800, 6700]	lh superiorfrontal volume	[27000, 32000]	ToMI total composite mean	[9, 18]
lh pericalcarine thickness	[1.7, 1.9]	rh posteriorcingulate volume	[4100, 6200]	rh parsorbitalis meancurv	[0.18, 0.2]
rh bankssts area	[1000, 1200]	lh pericalcarine volume	[2300, 3200]	rh parsorbitalis meancurv	[0.18, 0.2]

Table 5: Fourth-Order CCs

Feature One	Value	Feature Two	Value	Feature Three	Value	Feature Four	Value
lh pericalcarine thickness	[1.7, 1.9]	rh parsorbitalis meancurv	[0.18, 0.2]	rh insula meancurv	[0.14, 0.17]	ToMI early subscale mean	[12, 18]
lh pericalcarine thickness	[1.7, 1.9]	rh lingual volume	[7000, 9000]	rh superiorparietal meancurv	[0.14, 0.18]	lh superiorparietal volume	[16000, 20000]
lh pericalcarine thickness	[1.7, 1.9]	rh lingual volume	[7000, 9000]	rh parsorbitalis meancurv	[0.18, 0.2]	lh superiorparietal volume	[16000, 20000]
lh postcentral thickness	[2.2, 2.5]	rh parsorbitalis meancurv	[0.18, 0.2]	rh insula meancurv	[0.14, 0.17]	rh pericalcarine thickness	[1.6, 1.9]

4 Discussion

The present study successfully classified and selected discriminative biomarkers and behavioral features in ASD children from 7 to 14 years old using a small dataset collected from a single research site. Machine learning (ML) tools have long been introduced to ASD research; but it still remains a far-reaching goal to build an ultimate prediction model and diagnostic system for ASD. With some progress, other studies often face the problem of using datasets across different research sites for classification purpose, rather than identifying discriminative features for diagnostic and treatment development purposes [21, 22, 36–43, 71, 72]. Additionally, traditional ML algorithms do not work well with ASD datasets due to high variances and the heterogeneous nature of the disease [44, 45]. Meanwhile, it requires a tremendous amount of effort to include ASD individuals in a research study given the social and language challenges of such a population. Thus, nearly all ASD datasets have a large number of features with small sample size, which despite being inappropriate for many ML algorithms, often leads to overfitting and poor classification accuracy. The

present study, however, has demonstrated exceptionally good performance (i.e., 100% PPV and 100% class coverage) of the CCEA with a dataset containing a large number of features yet small sample size; and in this case, identifies features significantly associated with ASD.

The select CCEA features included volume, area, cortical thickness and mean curvature in specific regions in the cingulate cortex, frontal cortex and temporal-parietal junction areas as biomarkers for ASD (e.g., the pericalcarine cortex, posterior cingulate cortex, isthmus of the cingulate gyrus, pars orbitalis, etc.). Such findings are consistent with previous literature suggesting that individuals with ASD have abnormalities in these brain regions [24–33, 73–75]. Additionally, third-order models from the training set include measurements from the ToMI-2 and the ToMTB as significant features [55–57], which has further validated the use of these tools in ASD assessments. Such findings can potentially help clinicians and researchers address specific domains in ToM to improve the social skills of children with ASD. Besides the select 8 second-order and 6 third-order models from the training set, it is impressive that the CCEA is also able to validate 7 additional best performing third- and fourth-order models on the testing set. Although it would have been ideal if the 8 second-order and 6 third-order models identified on the training set had perfectly modeled the testing set, the fact that an additional 7 or the original models classify, with 100% PPV and 100% class coverage, on the testing set further demonstrates the robustness of the original models generated using the training set, as well as the exceptional performance of the CCEA algorithm in this study.

5 Limitations and Future Directions

ASD research often struggles with balancing between issues of either having a large sample size but high heterogeneity between subjects, or the other way around. Additionally, most currently available ML algorithms are designed for solving classification problems using big data. Although the present study has limitations given the relatively small sample size, it is impressive that the CCEA algorithm is able to identify features with exceptional classification performance and biomarker identification using a small dataset containing a large number of features. With the strong performance of the CCEA for the current sample size, the generation of more comprehensive results would be possible given a larger sample size.

The present study has established important biomarker candidates of ASD. These biomarker candidates fall into the same brain regions that have been identified to show abnormalities in ASD from studies adopting traditional neuroimaging measurements. Thus, it provides evidence that AI methodologies can perform as well as the traditional approaches in the field of neuroscience and ASD in selecting neuroanatomical biomarkers. Although AI techniques have been adopted to help with diagnosis and treatment development in medicine, ASD is exceptionally challenging given its great heterogeneity nature. It will require a large, diverse, and comprehensive dataset to extract solid biomarkers, which can be very time-consuming and less accurate using traditional approaches. Under such circumstances, ML techniques can help advance the development of an automatic diagnostic and predictive system for ASD. In summary, the present study has provided a new direction in adopting AI techniques in ASD research and medicine in general.

Acknowledgements

This project was supported by a private donor committed to advancing research in autism spectrum disorder. We thank Jay V. Gonyea, Administrative Director, and Scott Hipko, Senior Research Technologist, in the

MRI Research Unit at the University of Vermont, for their support in acquiring the MRI scans. We thank Dr. Richard Watts, Ph.D., Director at the FAS Brain Imaging Center, and Dr. Joeseph Orr, Ph.D., Assistant Professor at the Texas A&M University, for sharing their knowledge in MRI data pre-processing.

References

[1] M. Fitzgerald, "The clinical gestalts of autism: Over 40 years of clinical experience with autism.," Recent Research and Clinical Applications, vol. 2, 2017. DOI: 10.577265906cf0.

[2] S. Levy, M. Duda, N. Haber, and D. P. Wall, "Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism.," Molecular Autism, vol. 8, 2017. DOI: 10.1186s13229-017-0180-6cf0.

[3] X. Bi, Y. Wang, Q. Shu, Q. Sun, and Q. Xu, "Classification of autism spectrum disorder using random support vector machine cluster.," Front Genet., vol. 9, 2018.

[4] A. Diagnostic and S. Manual, 5th ed DSM-5. Washington, DC: American Psychiatric Association; 2013.

[5] C. for Disease Control and Prevention, "2020 community report on autism," vol. 13, 2020.

[6] L. Achenie, A. Scarpa, R. Factor, T. Wang, D. Robins, and D. Mccrickard, "A machine learning strategy for autism screening in toddlers.," Journal of Developmental & Behavioral Pediatrics, vol. 40, 2019.

[7] N. J. Lavelle T.A. Weinstein M.C., "Economic burden of autism spectrum disorders," Pediatrics, vol. 133, 2014.

[8] P. F. Bolton, J. Golding, A. Emond, and S. CD., "Autism spectrum disorder and autistic traits in the avon longitudinal study of parents and children: Precursors and early signs," J Am Acad Child Adolesc Psychiatry, vol. 51, p. 3, 2012.

[9] J. Kleinman, P. Ventola, J.Pandey, A.D.Verbalis, M.Barton, S.Hodgson, J.Green, T.Dumont-Mathieu, D.L.Robins, and D.Fein, "Diagnostic stability in very young children with autism spectrum disorders," J Autism Dev Disord, vol. 38, p. 4, 2008.

[10] M. C.Lord and A. Couteur, "A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders," J Autism Dev Disord, vol. 24, p. 659, 1994.

[11] C. Lord, S. Risi, and L. Lambrecht, "The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism," J Autism Dev Disord, vol. 30, p. 205, 2000.

[12] J. Baio, "Prevalence of autism spectrum disorders: Autism and developmental disabilities monitoring network, 14 sites, united states," vol. 61, 2012.

[13] R. A. Rhoades, A. Scarpa, and B. Salley, "The importance of physician knowledge of autism spectrum disorders: Results of a parent survey," BMC Pediatr, vol. 7, p. 37, 2007.

[14] D. Mandell, J. Listerud, and S. Levy, "Race differences in the age at diagnosis among medicaid-eligible children with autism," J Am Acad Child Adolesc Psychiatry, vol. 41, p. 1447, 2002.

[15] D. S. Mandell, M. M. Novak, and C. Zubritsky, "Factors associated with age of diagnosis among children with autism spectrum disorders," Pediatrics, vol. 116, p. 1480, 2005.

[16] M. J. Morrier, K. L. Hess, and L. J. Heflin, "Ethnic disproportionality in students with autism spectrum disorders," Multicultural Educ Fall, vol. 16, p. 31, 2008.

[17] L. D. Wiggins, J. Baio, and C. Rice, "Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample," J Dev Behav Pediatr, vol. 27, p. 2, 2006.

[18] R. Bernier, A. Mao, and J. Yen, "Psychopathology, families and culture: Autism," Child Adolesc Psychiatr Clin N Am, vol. 19, p. 4, 2010.

[19] F. Thabtah and D. Peebles, "A new machine learning model based on induction of rules for autism detection.," Health Informatics Journal, vol. 1460, 2019. DOI: 10.1177/1460458218824711 00.

[20] M. N. Parikh, H. Li, and L. He, "Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data.," Frontiers in Computational Neuroscience, vol. 13, 2019.

[21] M. Duda, R. Ma, N. Haber, and D. P. Wall, "Use of machine learning for behavioral distinction of autism and adhd.," Transl Psychiatry, vol. 6, 2016.

[22] D. Bone, M. S. Goodwin, M. P. Black, C. C. Lee, K. Audhkhasi, and S. Narayanan, "Applying machine learning to facilitate autism diagnostics: Pitfalls and promises.," J Autism Dev Disord., vol. 45, p. 1121, 2015. DOI: 10.1007/s10803-014-2268-6.

[23] E. Feczko, N. Balba, O. Miranda-Dominguez, M. Cordova, S. Karalunas, L. Irwin, and D. Fair, "Subtyping cognitive profiles in autism spectrum disorder using a functional random forest algorithm.," NeuroImage, vol. 172, 2018.

[24] R. Chen, Y. Jiao, and H. E.H., "Structural mri in autism spectrum disorder," Pediatr Res, vol. 69, p. 63, 2011.

[25] D. Amaral, C. Schumann, and C. Nordahl, "Neuroanatomy of autism," Trends in Neurosci, 2008.

[26] P. Brambilla, "Brain anatomy and development in autism: Review of structural mri studies.," Brain Res, vol. 61, p. 557, 2003.

[27] N. Lange, "Longitudinal volumetric brain changes in autism spectrum disorder ages 35 years," Autism Res, vol. 8, p. 82, 2015.

[28] D. L. Dierker, "Analysis of cortical shape in children with simplex autism," Cereb. Cortex, vol. 25, p. 1042, 2015.

[29] C. W. Nordahl, "Cortical folding abnormalities in autism revealed by surface-based morphometry.," J Neurosci, vol. 27, p. 11 725, 2007.

[30] S. L. Valk, A. Di Martino, M. P. Milham, and B. BC., "Multicenter mapping of structural network alterations in autism," Hum Brain Mapp, vol. 36, p. 2364, 2015.

[31] G. L. Wallace, "Longitudinal cortical development during adolescence and young adulthood in autism spectrum disorder: Increased cortical thinning but comparable surface area changes," Acad. Child Adolesc. Psychiatry, vol. 54, p. 464, 2015.

[32] R. Kucharsky Hiess, "Corpus callosum area and brain volume in autism spectrum disorder: Quantitative analysis of structural mri from the abide database," J. Autism Dev Disord, vol. 45, p. 3107, 2015.

[33] M. Shokouhi, J. H. Williams, G. D. Waiter, and B. Condon, "Changes in the sulcal size associated with autism spectrum disorder revealed by sulcal morphometry.," Autism Res, vol. 5, p. 245, 2012.

[34] S. Ha, I.-J. Sohn, N. Kim, H. J. Sim, and K.-A. Cheon, "Characteristics of brains in autism spectrum disorder: Structure, function and connectivity across the lifespan," Experimental Neurobiology, vol. 24, no. 4, p. 273, 2015. DOI: 10.5607/en.2015.24.4.273.

[35] C. Ecker, A. Shahidiani, Y. Feng, and E. Daly, "The effect of age, diagnosis, and their interaction on vertex-based measures of cortical thickness and surface area in autism spectrum disorder," J Neural Transm (Vienna), vol. 121, pp. 1157–1170, 2014.

[36] I. Gori, A. Giuliano, F. Muratori, I. Saviozzi, P. Oliva, and R. Tancredi, "Gray matter alterations in young children with autism spectrum disorders: Comparing morphometry at the voxel and regional level," J Neuroimaging, vol. 25, p. 866, 2015.

[37] Y. Jin, C. Wee, F. Shi, K. Thung, D. Ni, and P. Yap, "Identification of infants at high-risk for autism spectrum disorder using multiparameter multiscale white matter connectivity networks," Hum. Brain Mapp, vol. 36, p. 4880,

[38] H. Chen, X. Duan, F. Liu, F. Lu, X. Ma, and Y. Zhang, "Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivity," Prog. Neuropsychopharmacol. Biol. Psychiatry, vol. 64, p. 1,

[39] P. Odriozola, L. Q. Uddin, C. J. Lynch, J. Kochalka, T. Chen, and V. Menon, "Insula response and connectivity during social and non-social attention in children with autism.," Cogn. Affect. Neurosci., vol. 11, p. 433, 2015.

[40] G. Chanel, S. Pichon, L. Conty, S. Berthoz, C. Chevallier, and G. J., "Classification of autistic individuals and controls using cross- task characterization of fmri activity.," NeuroImage Clin.,

[41] G. J. Katuwal, S. A. Baum, N. D. Cahill, and M. AM., "Divide and conquer: Sub-grouping of asd improves asd detection based on brain morphometry," PLoS One, vol. 11, p. 1, 2016.

[42] C. P. Chen, "Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism.," Clin NeuroImage, vol. 8, p. 238, 2015.

[43] C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, and A. S. Jakab, "The neuro bureau preprocessing initiative: Open sharing of preprocessed neuroimaging data and derivatives," Front. Neuroinform., vol. 7, p. 41, 2013.

[44] C. Kelly, B. B. Biswal, R. C. Craddock, F. Castellanos, and M. Milham, "Characterizing variation in the functional connectome: Promise and pitfalls," Trends in Cognitive Sciences, vol. 16, p. 3, 2012.

[45] W. Huf, K. Kalcher, R. N. Boubela, G. Rath, A. Vecsei, P. Filzmoser, and E. Moser, "On the generalizability of resting-state fmri machine learning classifiers," Frontiers in Human Neuroscience, vol. 8, p. 502, 2014.

[46] P. Lanka, D. Rangaprakash, M. N. Dretsch, J. S. Katz, T. S. Denney, and G. Deshpande, "Supervised machine learning for diagnostic classification from large-scale neuroimaging datasets," Brain Imaging and Behavior, vol. 720, 2019. DOI: 10.1007/s11682-019-00191-8cf0.

[47] E. Carrow-Woolfolk, "Comprehensive assessment of spoken language second edition," 2017.

[48] B. A. Bracken and R. S. McCallum, "Universal nonverbal intelligence test (2nd ed.)," 2016.

[49] F. Moore, M. A., R. S., and B. A. Bracken, Handbook for Nonverbal Assessment. Springer, 2017.

[50] T. L. Hutchins and P. A. Prelock, "Technical manual for the theory of mind inventory-2. copyrighted manuscript," Theoryofmindinventory.com, 2016.

[51] B. Li, A. Sharma, J. Meng, S. Purushwalkam, and E. Gowen, "Applying machine learning to identify autistic adults using imitation: An exploratory study.," Plos One, vol. 12, 2017.

[52] M. Rutter, A. Bailey, and C. S. C. Q. Lord, The Social Communication Questionnaire Manual. Western Psychological Services, 2003.

[53] M. Baron-Cohen, An essay on autism and theory of mind. MIT Press, 1995.

[54] S. Baron-Cohen, A. M. Leslie, and U. Frith, "Does the autistic-child have a theory of mind," Cognition, vol. 277, p. 85, 1985.

[55] T. L. Hutchins, P. Prelock, and W. Chase, "Test-retest reliability of a theory of mind task battery for children with autism spectrum disorders," Focus Autism Other Dev Disabil, vol. 23, p. 195, 2008.

[56] T. L. Hutchins, P. A. Prelock, and L. Bonazinga, "Psychometric evaluation of the theory of mind inventory (tomi): A study of typically developing children and children with autism spectrum disorder," J Autism Dev Disord, vol. 42, p. 327, 2012.

[57] T. L. H. M. Lerner and P. A. Prelock, "Brief report: Preliminary evaluation of the theory of mind inventory and its relationship to measures of social skills," J Autism Dev Disord, vol. 41, p. 4, Apr. 2011.

[58] J. Hadwin and J. Perner, "Pleased and surprised: Childrens cognitive theory of emotion," British Journal of Developmental Psychology, vol. 9, p. 215, 1991.

[59] T. Ruffman and T. R. Keenan, "The belief-based emotion of surprise: The case for a lag in understanding relative to false belief," Developmental Psychology, vol. 32, p. 1, 1996.

[60] T. Russel and T. R. Keenan, "The belief- based emotion of surprise: The case for a lag in understanding relative to false belief," Developmental Psychology, vol. 32, 1996.

[61] B. Seider and S. L. D., "A developmental analysis of elementary school-aged childrens concepts of pride and embarrassment," Child Development, vol. 59, p. 367, 1988.

[62] A. Hillier and L. Allinson, "Understanding embarrassment among those with autism: Breaking down the complex emotion of embarrassment among those with autism," Journal of Autism and Developmental Disorders, vol. 32, p. 6, 2002.

[63] Youtube, [Online]. Available: `https://www.youtube.com/watch?v=Q_Pa6KFL1Nw%5C&t=139s%5Ccf2`.

[64] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, and J. L. Andersson, "The minimal preprocessing pipelines for the human connectome project," NeuroImage, vol. 80, p. 105, 2013.

[65] FreeSurfer, [Online]. Available: `https://surfer.nmr.mgh.harvard.edu/fswiki/aparcstats2table`.

[66] F. B, "Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain.," Neuron, vol. 33, pp. 341–355, 2002.

[67] J. P. Hanley, D. M. Rizzo, J. S. Buzas, and M. J. Eppstein, "A tandem evolutionary algorithm for identifying causal rules for complex data," Evol Comput, vol. 28, no. 1, pp. 87–114, 2020.

[68] R. L. Wasserstein, A. L. Schirm, and N. A. Lazar, "Moving to a world beyond $p < 0.05$.," The American Statistician, vol. 93, pp. 1–19, 2019.

[69] R. Wasserstein and N. Lazar, "The asa 92s statement on p-values: Context, process, and purpose," Amer Statist, vol. 70, pp. 129–133, 2016.

[70] R. Nuzzo, "Scientific method: Statistical errors," Nature, vol. 506, pp. 150–152, 2014.

[71] Y. Zhang and L. Wu, "Classification of fruits using computer vision and a multiclass support vector machine.," Sensors, vol. 1248, 2012.

[72] D. Li, W. Yang, and S. Wang, "Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine.," Electron. Agric., vol. 74, p. 274, 2010.

[73] K. Nickel, L. T. Elst, J. Manko, J. Unterrainer, R. Rauh, and C. Klein, "Volume loss distinguishes between autism and (comorbid) attention-deficit/hyperactivity disorder freesurfer analysis in children," Frontiers in Psychiatry, vol. 2, p. 9, 2018.

[74] B. A. Zielinski, M. B. Prigge, J. A. Nielsen, A. L. Froehlich, T. J. Abildskov, and J. S. Anderson, "Longitudinal changes in cortical thickness in autism and typical development," Brain, vol. 2, no. 137, pp. 1799–1812, 2014.

[75] M. C. Postema, D. van Rooij, and E. Anagnostou, "Altered structural brain asymmetry in autism spectrum disorder in a study of 54 datasets," Nat Commun, vol. 10, p. 4958, 2019. [Online]. Available: https://doi.org/10.1038/s41467-019-13005-8.