RESEARCH ARTICLE

Identifying neuroanatomical and behavioral features for autism spectrum disorder diagnosis in children using machine learning

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

1 Department of Communication Sciences and Disorders, University of Vermont, Burlington, VT, United States of America, 2 Department of Civil and Environmental Engineering, University of Vermont, Burlington, VT, United States of America, 3 Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, VT, United States of America

* yhan@broadinstitute.org

Abstract

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that can cause significant social, communication, and behavioral challenges. Diagnosis of ASD is complicated and 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, and is able to accommodate datasets having a small number of samples with a large number of feature measurements. The dataset is unique and comprises both behavioral and neuroimaging measurements from a total of 28 children from 7 to 14 years old. Potential biomarker candidates identified include brain volume, area, cortical thickness, and mean curvature in specific regions around the cingulate cortex, frontal cortex, and temporal-parietal junction, as well as behavioral features associated with theory of mind. A separate machine learning classifier (i.e., k-nearest neighbors algorithm) was used to validate the CCEA feature selection and for ASD prediction. Study findings demonstrate how machine learning tools might help move the needle on improving diagnostic and predictive models of ASD.

Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that can cause significant social, communication, and behavioral challenges. According to the most recent report from the Center for Disease Control and Prevention (CDC), the number of children in the U.S. diagnosed with ASD are about 1 in every 54 [1]. While genetic and environmental factors have been linked to the development of ASD, at present there is no identified cause or cure of ASD.

ASD is characterized by impairments in social communication and social interaction and the presence of restricted and repetitive behaviors, interests, or activities [2–5]. At the core of social impairment is often a difference in theory of mind (ToM), the ability to recognize and understand the thoughts, feelings and perspectives of others. Some symptoms of ASD are not

evident until age two or later. In fact, a child may appear to be developing according to typical milestones until the age of two and then stop learning new skills and may even lose skills [6, 7]. Currently, the diagnosis of autism is based on behavioral symptoms alone, often assessed through the Autism Diagnostic Observation Schedule-Second edition (ADOS-2) and the Autism Diagnostic Interview-revised (ADI-R) [8, 9]. A typical diagnostic appointment consists of evaluations lasting several hours at a designated clinical office. The rigorous and time-consuming nature of ASD diagnostic examinations often leads to a demand that exceeds the capacity to see patients. As a result, many diagnostic centers have expanding wait lists for appointments. This bottleneck can translate to delays in diagnosis of 13 months and longer [6, 10–16]. It is also believed that a substantial number of individuals on the spectrum remain undetected [17]. With growing awareness of ASD, there is a high demand for a faster and more automated ASD diagnostic approach that might allow for more efficient diagnosis and early identification of high-risk populations [18].

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. For instance, Duda and colleagues (2016) applied machine learning to distinguish ASD from attention deficit hyperactivity disorder (ADHD) using the Social Responsiveness Scale for children between 5 to 13 years old [19]. Bone et al. (2015) trained their models to diagnose children with ASD versus neurotypical (NT) children also using the Social Responsiveness Scale and the ADI-R score for children between 5 to 17 years old [20]. Other studies aggregated items from the ADOS and scores from the Autism Quotient (AQ) to accurately classify an ASD group [21]. Although behavioral outcome measures remain the gold standard for ASD diagnosis and assessment with solid reliability and validity, there are some limitations when using behavioral measures to classify participants. Specifically, these measures are often subjected to clinicians' and care providers' opinions built upon their training and experiences. In addition, they are typically prone to performance demands, as well as environmental and personal factors that could influence responses on any given day. In contrast, it is important that neuroimaging techniques and measures supplement behavioral measures to provide objective metrics that allow us to identify more consistent ASD markers. As a result of the wide range and subjective nature of behavioral measures used in diagnosing ASD, many studies are exploring brain-based biological markers (e.g., measurable via magnetic resonance imaging (MRI)) to identify a common etiology across individuals with ASD. Currently, these less subjective markers are attractive not only for diagnostic purposes, but as possible targets for interventions [22]. Independent structural MRI studies have found differences in whole brain volume and the developmental trajectories between individuals with and without ASD [23–31]. Other structural brain abnormalities associated with ASD include cortical folding signatures that appear in the following regions of the brain: temporal-parietal junction, anterior insula, posterior cingulate, lateral and medial prefrontal cortices, corpus callosum, intraparietal sulcus, and occipital cortex [23–31]. 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 [32], while other studies suggest that individuals with ASD tend to have thinner cortices and reduced surface area as an effect of aging [33]. With these clear brain differences among those with and without ASD, it is informative and critical to look for brain-based ASD biomarkers.

Machine learning (ML) has been introduced to the neuroimaging field to identify the atypical brain regions in individuals with ASD. The support vector machine (SVM) is an algorithm that avoids overfitting and is known for high classification accuracy without requiring large sample sizes. It has been used to classify ASD from NT participants using extracted features from functional connectivity metrics and grey matter volume [34–37]. Other ASD applications

of ML classifiers include deep neural networks [38] and the random forest (RF) algorithm; the latter uses random ensembles of independently grown decision trees [39]. Although these methods have demonstrated high accuracy for classifying ASD, to our knowledge most of them have not been used to identify input variables most closely associated with ASD (i.e., feature selection). A recent review paper indeed highlighted some machine learning methods involving feature selection for ASD [40]; however, it is worth noting that the majority of ASD studies still focus on classification. In addition, many studies use data from the Autism Brain Imaging Data Exchange (ABIDE) dataset, which includes 1112 existing resting-state functional MRI (rs-fMRI) datasets with corresponding structural MRI and phenotypic information from 539 individuals with ASD and 573 age-matched NT controls between the ages of 7 and 64 collected from 24 international brain imaging laboratories [41]. For example, Guo et al. (2017) selected features associated with ASD using a deep neural network from brain resting-state functional connectivity patterns using the ABIDE dataset [42]. However, findings using datasets from various sites have several limitations.

Specifically, classification across a heterogeneous population as is true in autism is challenging [43–45], particularly when neuroimaging data are pooled from multiple acquisition sites such as the ABIDE dataset, which has considerable variation in demographic and phenotypic profiles. Data variance introduced via scanner hardware, imaging protocols, operator characteristics, regional demographics, and other site-specific acquisition factors can also affect the classification performance. It is often difficult to collect neuroimaging data from individuals with ASD given the scanner noises and high expectations for individuals to remain still. In fact, most individual site datasets have small sample sizes that can lead to overfitting and classification inaccuracies using traditional ML algorithms. Moreover, while the ultimate goal for ML-based diagnostic classification in neuroimaging is to identify discriminative features that provide insight into atypical structure and connectivity patterns in the affected population [46], many of the ML algorithms applied to ASD were designed to classify large amounts of data (e.g., ABIDE) rather than optimize the selection of input features.

ASD drivers or markers are likely the result of a complex interaction of factors with no single factor (i.e., main effect or univariate model) driving the system. As such, traditional statistical tools (e.g., logistic regression) that search for univariate drivers of ASD are unlikely to find consistent patterns. Thus, ML techniques that explore large search spaces for multivariate interactions are both needed and becoming popular in helping to elucidate the complex interactions in systems such as ASD. Our study employs one such ML tool: an evolutionary algorithm [47] called the conjunctive clause evolutionary algorithm (CCEA) [48]. The CCEA was specifically designed to efficiently explore large search spaces for complex interactions between features and some associated nominal outcome (e.g., ASD or NT). In addition, the CCEA has built-in tools to prevent overfitting to produce easily interpretable parsimonious models.

This study examines the utility of using the CCEA for feature selection in ASD, particularly to address traditional statistical challenges associated with datasets having small sample sizes and a large number of feature measurements. Additionally, the selected features are validated and used for diagnostic classification by applying a separate and more traditional ML classifier (i.e., the k-nearest neighbors (KNN) algorithm). The dataset in the present study has a relatively large number of features, consisting of both behavioral and neuroimaging measurements. Although some behavioral features may seem less objective compared to an assessment of neuroanatomical features, they are often easier to obtain and more pragmatic for children with ASD especially if they have strong psychometric properties. Clinicians seldom place a child with ASD in a scanner to obtain neuroanatomical information before conducting behavioral assessment for the purpose of diagnosis and treatment. The combination of behavioral measures with neuroanatomical information supports the value of making brain-behavior

connections that will advance our understanding of ASD. In the present study, the behavioral measurements include scores of language ability, intellectual ability, and ToM. The neuroimaging measurements include brain volume, brain surface area, cortical thickness, and cortical curvature extracted from MRI whole-brain T1-weighted scans. These features were collected from a total of 28 children ages 7 to 14, of which 9 children had been diagnosed with ASD [49]. Only a subset of these 28 children were used for feature selection in the CCEA (7 children with ASD and 14 NT children), as another subset (2 children with ASD and 5 NT children) were enrolled at a later time (i.e., after the CCEA was trained). While this later cohort was not included in the CCEA feature selection analysis, it was included in the subsequent validation and predictive k-nearest neighbors (KNN) modeling.

Using the CCEA, we aim to identify discriminative biomarkers and behavioral features to help develop an automated diagnostic and predictive system for ASD. We believe this is a pioneering study for:

- Selecting discriminative biomarkers among children from 7 to 14 years old and classifying ASD.

- Including both behavioral and neuroimaging measurements in the feature selection model for better prediction and understanding of ASD.

- Identifying models (sets of features) that most strongly correlate to children with ASD given a dataset with a relatively small sample size (i.e., N = 28) and a large number of features (i.e., 247 neuroimaging features and 13 behavioral features).

- Developing a predictive ML model using input features selected by the CCEA.

Materials and methods

Participants

A total of 9 children with ASD (1 female) and 19 NT children (7 female), ages 7–14, were enrolled in the study. In addition to the behavioral assessments described below, the ASD group also completed the ADOS-2 and the Social Communication Questionnaire-Lifetime version (SCQ) [50] to confirm their ASD diagnosis. Although diagnosis of ASD is typically done at an early age, the characteristics of ASD are long-term, and classification with additional neurobiological information at any age recognizes the potential for brain-behavior comparisons with NT populations. Potential changes behaviorally and neurobiologically at any age may also inform types, duration, and intensity of intervention that may influence these changes. Therefore, although we tested older children, this study informs an increased understanding of likely biomarkers of ASD that can be applied to younger populations.

Ethical statement

The present study was approved by the University of Vermont (UVM) Research Protection Office and IRB committee. The specific staff member who approved the study protocol, 19–0005, was Karen Crain. Consent forms were obtained from each participant's legal guardian and assent forms were obtained from each participant.

Behavioral measurements

All children participated in 2–3 hours of baseline behavioral assessments that included the Comprehensive Assessment of Spoken Language (CASL) [51], the Universal Nonverbal Intelligence Test-2 (UNIT-2) [52, 53], the Theory of Mind Task Battery (ToMTB) [54] and the

Theory of Mind Inventory-2 (ToMI-2) [55, 56]. Measures of language and cognition are typical in the assessment of ASD, as the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) requires an assessment of language and intellectual functioning beyond the diagnosis of ASD.

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. Reliability for the core subtests ranges from 0.80 to 0.90 with composite scores ranging from 0.92 to 0.96. Test-retest reliability for individual subtests range from 0.65 to 0.95 with core composite coefficients from 0.92 to 0.93. 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. Reliability studies include internal consistency, test-retest, and scorer reliability. Subtest coefficients range from 0.88 to 0.96 and 0.93 to 0.98 for the composites. Content, criterion and construct validity were largely correlated with other cognitive measures and demonstrated the value of the UNIT-2 for the assessment of diverse groups of children.

ToM is a core deficit in ASD that is often used to explain the social impairments of the disorder. 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 [57, 58]. Children with ASD often achieve early and basic ToM skills later developmentally than their neurological peers and frequently fail to achieve competence in advanced ToM. Thus, ToM is a good marker for social-cognitive performance differences that can be aligned with neurological differences. Scores from both the ToMTB and ToMI-2 were included to provide representative measures of a child's social cognition level. The ToMTB and ToMI-2 are two norm-referenced tools and behavioral tasks used as outcome measures to assess ToM [57, 58]. Scores from both the 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 ToM. 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 distractors. Memory control questions are included that must be passed for credit on the test questions. The ToMTB demonstrates excellent internal consistency and inter task agreement (α = 0.91 at T1 and α = 0.94 at T2) and strong construct validity (r = 0.66, p < .01) [54, 56, 59].

In selecting potential features for the CCEA, we included 13 behavioral features. These included the total score of the CASL, full scale score of the UNIT-2, abbreviated score of the

Table 1. Participant behavioral assessments scores: NT vs. ASD.

		NT Group (n = 19)	ASD Group (n = 9)	Group Difference
	Age	10.2 (7–14)	11 (8–13)	$p = 0.36$
	CASL	113 (89–132)	78.8 (40–121)	$p < 0.001$**
	UNIT-2 Full Scale	111.9 (100–135)	96.2 (56–127)	$p < 0.05$*
	Unit-2 Abbreviation	110.9 (91–124)	102 (67–121)	$p = 0.16$
	ToMTB	13.3 (11–15)	11.1 (5–14)	$p < 0.05$*
ToMI-2	Total	17.2 (11.5–19.3)	12.4 (9.1–17.6)	$p < 0.001$**
	Early Subscale	17.7 (13.8–20)	14.6 (11.6–17.9)	$p < 0.05$*
	Basic Subscale	18 (13.2–19.7)	14 (8.5–19.2)	$p < 0.001$**
	Advanced Subscale	16 (8.3–18.9)	9.4 (6–16.4)	$p < 0.001$**
	Happy	18.5 (14.1–20)	17.3 (12.5–20)	$p = 0.24$
	Sad	17.5 (7.1–20)	16.9 (13.1–20)	$p = 0.62$
	Surprise	17.8 (12.2–20)	15.3 (10.3–20)	$p < 0.05$*
	Embarrassment	15.3 (5.8–20)	10.3 (4.8–20)	$p < 0.05$*
	Desire-Based	18.9 (15–20)	16.7 (14.6–20)	$p < 0.05$*

https://doi.org/10.1371/journal.pone.0269773.t001

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 ToM abilities such as regulating desire-based emotion and recognition of happiness and sadness), 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 (i.e., assessing advanced ToM ability such as recognition of embarrassment). Drawn from a larger study examining the neural mechanisms underlying ToM and emotion recognition [49], we also included scores from single ToMI-2 items assessing recognition of simple emotions such as happiness and sadness, as well as more complex emotions such as surprise and embarrassment, which ASD children often find difficult to recognize and process [60–63]. Table 1 provides an overview of scores on the 13 behavioral measures. Results from independent sample *t*-tests found that NT participants scored significantly higher ($p < 0.05$) than ASD participants on the CASL, UNIT-2 full scale, ToMTB, ToMI-2 total, ToMI-2 early subscale, ToMI-2 basic subscale, ToMI-2 advanced subscale, as well as ToMI-2 single items of surprise, embarrassment and desire-based emotion. In its psychometric development, the ToMI and ToMTB were tested against children with and without autism with scores on both measures, discriminating those children with a diagnosis of ASD from those without such a diagnosis. These measures of ToM also showed a delay in development of early developing and basic ToM skills and a failure to achieve more advanced ToM skills [54–56].

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 were TR 6.4s, TE 2.9s, flip angle 8 degree, 1mm isotropic imaging resolution with a 256 x 240 *mm2* field of view and 225 slices. 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 lying in an MRI scanner [64]. 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 [65]. Brain anatomical features were extracted using FreeSurfer aparcstats2tabl script [66], which reads brain parcellation statistics directly from T1 scan and exports them to excel sheets. These extracted anatomical features include volume, cortical thickness, mean curvature, and area of all ROIs for each subject. These ROIs were 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 [67]. There were 276 brain features included in total.

Conjunctive clause evolutionary algorithm

We used an 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 feature values [48]. The CCEA can find feature interactions even in the absence of main-effects, and can, therefore, identify feature combinations that would be difficult to discover using traditional statistics. The CCEA selects for the best conjunctive clauses (CCs) of the form:

$$CC_k = F_i \in a_i \land F_j \in a_j \ldots, \tag{1}$$

where F_i represents a risk factor i whose value lies in the range a_i; and the symbol $\land$ represents a conjunction (i.e., logical AND). One benefit of the CCEA is that it produces parsimonious models that are correlated with a select category (e.g., ASD). The models generated by the evolutionary algorithm can be described by their order or total number of features in the conjunctive clause. One example of a parsimonious second order conjunctive clause is: a person with a right hemisphere isthmus cingulate volume of 3,300–4,100 mm^3 AND a right hemisphere posterior cingulate volume of 4,100–6,200 mm^3 is more likely to have ASD than someone who does not meet these criteria. The CCEA was able to generate models that ranged from 1st to 5th-order in the present study. Since we are only interested in the most parsimonious (i.e., lowest order models) to draw meaningful conclusions and to avoid the overfitting that often occurs with higher-order models, we focused on second-order models (i.e., those having only two features) with the highest fitness (PMF) in which all of the features identified were brain anatomical features. Because of our desire to explore the predictive capability of the behavioral features, we also expanded our analysis to include third-order models (i.e., model combinations with three features) in which behavioral features were included.

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. The sensitivity of each feature 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. 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

Table 2. Subject inclusion and distribution.

Models		Subjects	Include Subjects Used In CCEA? N = 21 (14 NT & 7 ASD)	Include Subjects from the Later Cohort That Was Not Used In CCEA? N = 7 (5 NT & 2 ASD)
CCEA		NT = 14 ASD = 7	All	None
KNN Validation Model		NT = 19 ASD = 9	All	All
KNN Prediction Model	Training	NT = 6	NT = 6	NT = None
		ASD = 6	ASD = 4	ASD = 2
	Testing	NT = 13	NT = 8	NT = 5
		ASD = 3	ASD = 3	ASD = None

(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.

K-nearest neighbors algorithm and leave-one-out cross validation

In order to further validate the CCEA's selection of features capable of discriminating between children with and without ASD, we built a separate KNN classification model and used leave-one-out cross validation on all 28 subjects. The KNN is a classification algorithm that assumes that things that exist in close proximity (i.e., nearer to each other) are more similar. In this study, each subject was classified into one of two output classes (i.e., ASD or NT) based on a plurality vote of its neighbors, with the subject being assigned to the class most similar to its k-nearest neighbors. When k = 1, then the subject is assigned to the class of a single nearest neighbor [71]. Leave-one-out cross validation is a special case of cross validation where the number of folds equals the number of subjects in the data set. Thus, the KNN algorithm is applied once for each subject, using all other instances as the training set and using the selected subject as a single-item test set [72]. After model validation, we trained three separate KNN classifiers using a balanced dataset (6 NT and 6 ASD subjects) and feature sets identified by the CCEA to classify the remaining 16 subjects. See Table 2.

Results

CCEA feature selection: 14 NT and 7 ASD

When using the CCEA for feature selection, 2438 CCs (i.e., models or sets of features) were generated ranging from first-order to fifth-order. The PPV of the 2438 models ranged from 46.47% to 100% and their class coverage ranged from 42.86% to 100%. Among these models, we looked for the most parsimonious (i.e., lowest order models) to draw meaningful conclusions and to avoid the overfitting that often occurs with higher-order models. As a result, we selected 8 second-order models (i.e., those having only two features) with the highest fitness (PMF) among the total 520 second-order models. These 8 "best performing" models each have 100% PPV and 100% class coverage, see Table 3. All of the features identified were brain anatomical features.

Using CC 113 (Table 3) as an example, this second-order model can be interpreted as: any subject 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

Table 3. Second-order CC model features and range of values.

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]

https://doi.org/10.1371/journal.pone.0269773.t003

features to appear most frequently (i.e., four times) across all second-order models, suggesting that the volume of cingulate gyrus is a potentially important biomarker for ASD. Fig 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.

Because of our desire to explore the predictive capability of 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; while some consisted only of 310 anatomical brain features, others included two behavioral features plus one brain anatomical feature. We selected the 6 best-performing third-order models with the highest fitness (PMF); each had 100% PPV and 100% class coverage, see Table 4. Each of these third-order models contained two behavioral features and one brain anatomical feature.

Using CC 46 (Table 4) as an example, any subject 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 be classified as having ASD. 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 the ToMTB and ToMI-2 might be effective for ASD testing and diagnosis. See Fig 2.

KNN leave-one-out cross validation

As mentioned earlier, a cohort of new subjects comprising 2 ASD and 5 NT children were enrolled at a later time. This later cohort was combined with the 21 subjects used in CCEA to cross-validate the KNN classifiers. Using the 8 unique features of the second-order model, the KNN (k = 3) achieved 89.29% classification accuracy (ACC), where 17 of the 19 NT subjects and 8 of the 9 ASD were classified accurately; see Table 5, diagonals of the second-order confusion matrix.

Using the 9 unique features of the third-order model, the KNN (k = 7) validation accuracy fell to 78.57% compared to when using the 8 unique features from the second-order model, where 15 of the 19 NT and 7 of the 9 ASD subjects were classified accurately. See Table 5, diagonals of third-order confusion matrix.

Given the better ASD classification performance of the second-order neuroanatomical features, and that the behavioral measurements/features are relatively easier to collect among children with ASD, it was important to explore whether the ASD prediction results might be improved when the behavioral features were combined with the second-order features. As a result, we added the three behavioral features (i.e.,ToMTB total score, ToMI-2 total composite
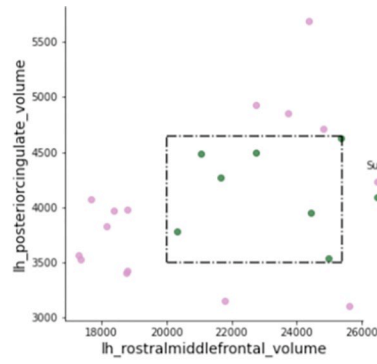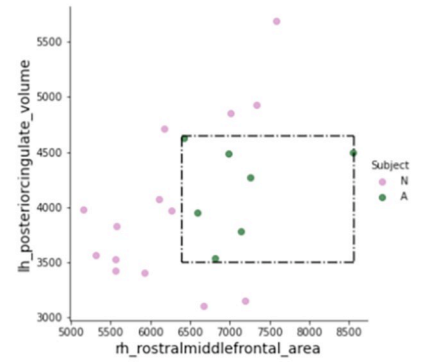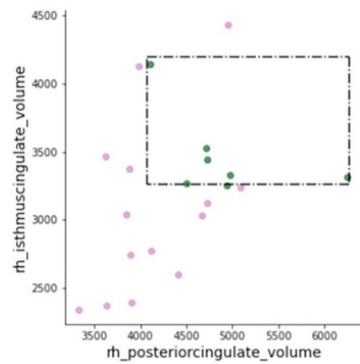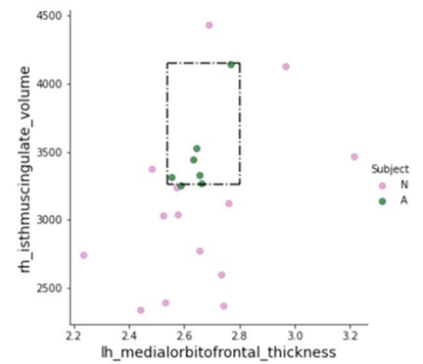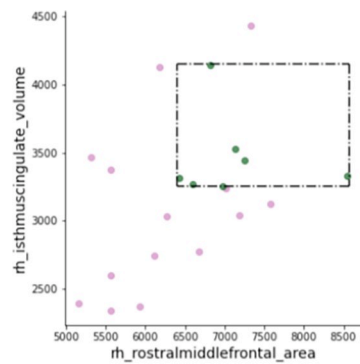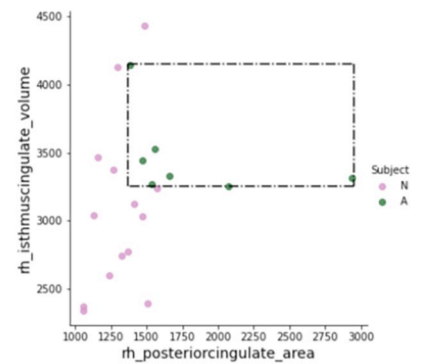
(a) CC113
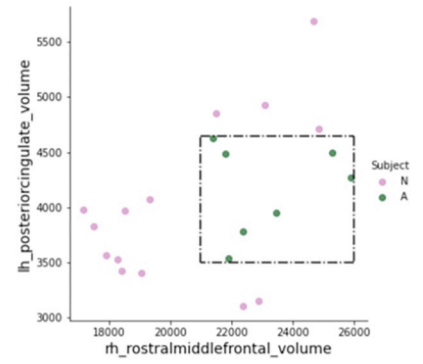
(b) CC679
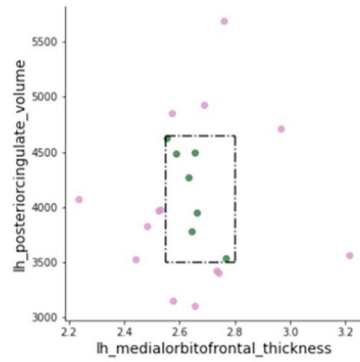
(c) CC1449

(d) CC2199

(e) CC887

(f) CC1488

(g) CC2200

(h) CC2269

8

Fig 1. 2D visualization of second-order CC models. Green dots represent ASD subjects and group together within the rectangle defining the range of values in Table 3.

mean and ToMI-2 early subscale mean) from the third-order models to the 8 second-order brain anatomical features and cross-validated a new KNN (k = 2) classifier. With the total of 11 unique features, a validation accuracy of 85.71% was achieved with 16 of the 19 NT subjects and 8 of the 9 ASD being classified accurately. See the confusion matrix of Table 5.

Classification of ASD and NT subjects using the KNN model

To further examine whether the KNN classifier could discriminate subjects with ASD and NT, we developed three classification models–one using the 8 unique features from the second-order models, one using the 9 features from the third-order models, and one using the 11-feature model (i.e., eight second-order neuroanatomical features and three behavioral features). The best classification accuracy was achieved using a balanced training set that consisted of 6 NT and 6 ASD subjects, among which only 2 of the 6 ASD subjects were not part of the original CCEA feature selection. The remaining 16 subjects were used for testing (13 NT and 3 ASD).

The KNN (k = 1) results for the second-order model features are shown in Table 6, columns 2 and 3; a classification accuracy of 87.5% was achieved, with all 3 of the ASD subjects and 11 of the 13 NT being classified accurately. Both of the misclassified NT subjects were part of the original CCEA feature selection.

The KNN (k = 3) classification accuracy for the third-order model features was 81.25%, with all 3 of the ASD subjects and 10 of the 13 NT subjects correctly classified. Of the 3 misclassified NT subjects, 2 were included in the original CCEA feature selection. See Table 6, columns 4 and 5.

Lastly, the KNN (k = 3) predictions for the combined 11-feature model (8 neuroanatomical features and three behavioral features) are shown in Table 6, columns 6 and 7; classification accuracy is 93.75%, with all 3 of the ASD subjects and 12 out of 13 NT subjects classified accurately. The one misclassified NT subject was not part of the original CCEA feature selection analysis.

Discussion

This study used a new ML feature selection tool, the CCEA, to identify biomarkers and behavioral features capable of successfully discriminating between children (7 to 14 year of age) with and without ASD given a small dataset collected from a single research site. ML tools have long been applied to ASD research, but it remains a far-reaching goal to build a diagnostic system for ASD that incorporates both feature selection and prediction. Previous studies face the challenge of using datasets across different research sites for classification purposes, rather

Table 4. Third-order CC model features and range of values.

| 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]
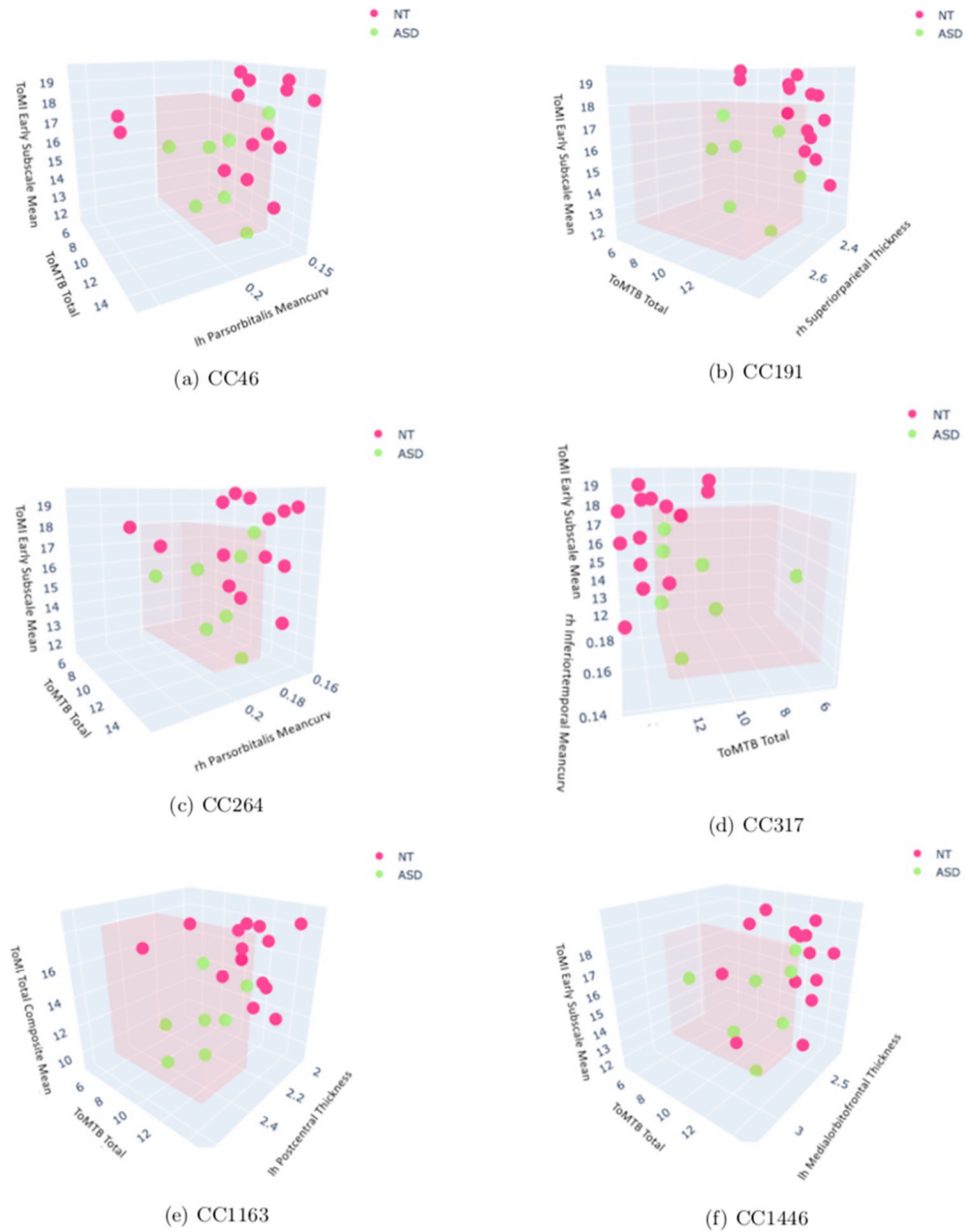
Fig 2. 3D visualization of third-order CC models. Green dots represent ASD subjects and group together within the pink cube defining the range of values in Table 4.

Table 5. Cross validation confusion matrices.

N = 28 (9 ASD & 19 NT)	Second-order Model Features *ACC Score*: 89.29%		Third-order Model Features *ACC Score*: 78.57%		Second-order Model Features Plus Three Behavioral Features *ACC Score*: 85.71%	
	Predicted ASD	Predicted NT	Predicted ASD	Predicted NT	Predicted ASD	Predicted NT
True ASD	8	1	7	2	8	1
True NT	2	17	4	15	3	16

Table 6. Classification confusion matrices.

N = 16 (3 ASD & 13 NT)	Second-order Model Features *ACC Score*: 87.5%		Third-order Model Features *ACC Score*: 81.25%		Second-order Model Features Plus Three Behavioral Features *ACC Score*: 93.75%	
	Predicted ASD	Predicted NT	Predicted ASD	Predicted NT	Predicted ASD	Predicted NT
True ASD	3	0	3	0	3	0
True NT	2	11	3	10	1	12

than identifying input variables most closely associated with ASD (i.e., feature selection) [19, 20, 34–37, 38–41, 73, 74]. Additionally, traditional ML algorithms do not work well with ASD datasets given the large amount of variance and the heterogeneous nature of the disorder [43, 44]. Meanwhile, it requires tremendous effort to include ASD individuals in a research study given the social, cognitive, and language challenges of this population. Thus, nearly all ASD datasets have a large number of features with relatively small sample sizes which, despite being unsuitable for many ML algorithms, often leads to overfitting and poor classification accuracy. However, the CCEA in this work was able to address such issues by efficiently exploring large search spaces for feature interactions associated with some nominal outcome (e.g., ASD or NT). It also adopted built-in tools to prevent overfitting to produce parsimonious models.

The present study demonstrated exceptionally good performance (i.e., 100% ASD PPV and 100% class coverage) of the features identified by the CCEA. The selected CCEA features from the parsimonious second and third order models included volume, area, cortical thickness, and mean curvature in specific regions around 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 [19, 23–31, 75–77], which facilitates our identification of potential ASD biomarkers. Additionally, third-order models from the training set include measurements from the ToMI-2 and the ToMTB as important features [56, 59]. This provides evidence that outcome measures from ToMI-2 and ToMTB are able to distinguish ASD from NT which further validates the use of these tools in ASD assessments.

It is impressive that the KNN classifiers are able to achieve such high classification accuracy given our sample size, and validation of the discriminant features selected by the CCEA models. In particular, the KNN classifiers perform better using the second-order neuroanatomical features than the third-order feature models, which emphasizes the importance of focusing on parsimonious models selected by the CCEA. In addition, the KNN achieved the highest classification accuracy when adding the behavioral features from the third-order models to the neuroanatomical features from the second-order models. In most cases, neuroimaging measurements are conducted along with behavioral assessments. As the third-order models are included to explore the potential role that behavioral features might play along the side of neuroanatomical features, it is convincing to see the highest classification accuracy achieved when combining the behavioral features and the neuroanatomical features. These findings highlight the heterogeneous and multi-facet characteristics of ASD itself. Thus, although it is more difficult to implement MRI among children with ASD, such findings support the idea that neuroanatomical measurements increase confidence in diagnosis. It also suggests that a good ASD prediction model should consider including both behavioral and neuroanatomical features to establish brain-behavior connections to advance our understanding of ASD.

This study further demonstrates the robustness of the CCEA as a feature selection methodology. The accuracy of these features when used as input variables in the KNN classifier suggests their potential to help clinicians and researchers target specific domains in ToM in

treating the social challenges most often seen in children with ASD. The implications of our findings for clinical researchers reinforce earlier findings regarding the brain-behavior connections for children with and without ASD related to ToM understanding [55, 57, 59, 78–80]. Knowing these connections may guide future researchers in the assessment of change following intervention at both a behavioral and neurobiological level. This may also lead to knowledge about which interventions may be most effective for children with specific neurobiological markers.

The present study has established important biomarker candidates of ASD. These biomarker candidates support previous research adopting traditional neuroimaging measurements identifying similar brain regions to explain the abnormalities in ASD [19, 23–31]. Importantly, ML methodologies can perform as well as the traditional approaches in the field of neuroscience and specifically in our assessment of ASD in selecting neuroanatomical biomarkers. Although ML techniques have been adopted to help with diagnosis and treatment development in medicine [81–84], the heterogeneity in ASD creates challenges. Typically, large, diverse, and comprehensive datasets are required to extract solid biomarkers, which can be time-consuming and may be less accurate with traditional approaches. Under such circumstances, ML techniques as described in this study can help advance the development of an automatic diagnostic and predictive system for ASD. Although one can argue that the sample size of this study may introduce few limitations, it is important to note that the goal was not to map out the complete "real-world" process. Instead, it aims to find patterns, (i.e., feature selection) and identify important parameters and highlight interesting feature interactions to help with the downstream process of understanding the system behavior involved in ASD biomarkers. The present study has not identified definitive answers on whether the selected features will continue to be correlated in a much larger population, however, unearthing these features enables researchers to hone in on a much smaller search space (i.e., thousands of features reduced to a handful of targeted features) in order to refine hypotheses and design more exacting experiments, such as simulation studies to test whether the correlation is robust with the potential for causal connections. In addition, disseminating these feature interactions will help the community of domain experts spur more refined hypotheses. In summary, the present study provides a new direction for adopting ML techniques in ASD research and other areas of medicine with similar heterogeneity in disease conditions.

Supporting information

S1 Data.
(CSV)

S2 Data.
(CSV)

Acknowledgments

Author Contributions

Conceptualization: Yu Han, Donna M. Rizzo, Patricia A. Prelock.

References

1. Centers for Disease Control and Prevention. 2020 Community Report on Autism.

2. Bi Xia-An, Wang Yang, Shu Qing, Sun Qi, Xu Qian. Classification of Autism Spectrum Disorder Using Random Support Vector Machine Cluster. Front Genet. 2018 Feb 6; 9:18. https://doi.org/10.3389/fgene.2018.00018 PMID: 29467790

3. American Psychiatric Association. Diagnostic and Statistical Manual. 5th ed DSM-5. Washington, DC. 2013.

4. Fitzgerald Michael. The Clinical Gestalts of Autism: Over 40 years of Clinical Experience with Autism. Open access peer-reviewed chapter. 2017. https://doi.org/10.577265906cf0

5. Levy Sebastien, Duda Marlena, Haber Nick, Wall Dennis P. Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. Mol Autism. 2017; 8: 65. https://doi.org/10.1186/s13229-017-0180-6 PMID: 29270283

6. Patrick F Bolton Jean Golding, Emond Alan, Colin D Steer. 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. 2012 Mar; 51(3):249–260.e25. https://doi.org/10.1016/j.jaac.2011.12.009 PMID: 22365461

7. Kleinman Jamie M., Ventola Pamela E., Pandey Juhi et al. Diagnostic Stability in Very Young Children with Autism Spectrum Disorders. J Autism Dev Disord. 2008 Apr; 38(4): 606–615. https://doi.org/10.1007/s10803-007-0427-8 PMID: 17924183

8. Lord Catherine, Rutter Michael, Ann 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 volume 24. 1994. Pages 659–685. https://doi.org/10.1007/BF02172145 PMID: 7814313

9. Lord C, Risi S, Lambrecht L, Cook E H Jr, Leventhal B L, DiLavore P C, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord. 2000 Jun; 30(3):205–23. PMID: 11055457

10. Centers for Disease Control and Prevention. Prevalence of autism spectrum disorders–Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. MMWR Surveill Summ. 2012 Mar 30; 61(3):1–19. PMID: 22456193

11. Bernier Raphael, Mao Alice, Yen Jennifer. Psychopathology, families, and culture: autism. Child Adolesc Psychiatr Clin N Am. 2010 Oct; 19(4):855–67. https://doi.org/10.1016/j.chc.2010.07.005 PMID: 21056350

12. David S Mandell, Maytali M Novak, Cynthia D Zubritsky. Factors associated with age of diagnosis among children with autism spectrum disorders. Pediatrics. 2005 Dec; 116(6):1480–6. https://doi.org/10.1542/peds.2005-0185 September 20, 2021. PMID: 16322174

13. Mandell David S, John Listerud, Susan E Levy, Pinto-Martin Jennifer A. Race differences in the age at diagnosis among medicaid-eligible children with autism. J Am Acad Child Adolesc Psychiatry. 2002 Dec; 41(12):1447–53. https://doi.org/10.1097/00004583-200212000-00016 PMID: 12447031

14. Morrier Michael J, Hess Kristen L, Heflin L. Juane. Ethnic Disproportionality in Students with Autism Spectrum Disorders. Multicultural Education, p31–38 Fall 2008.

15. Rhoades Rachel A, Scarpa Angela, Salley Brenda. The importance of physician knowledge of autism spectrum disorder: results of a parent survey. BMC Pediatr. 2007 Nov 20; 7:37. https://doi.org/10.1186/1471-2431-7-37 PMID: 18021459

16. Wiggins Lisa D, Jon Baio, Rice Catherine. Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. J Dev Behav Pediatr. 2006 Apr; 27(2 Suppl): S79–87. https://doi.org/10.1097/00004703-200604002-00005 PMID: 16685189

17. Thabtah F. and Peebles D. A new machine learning model based on induction of rules for autism detection. Health Informatics Journal. 2019. https://doi.org/10.1177/1460458218824711 PMID: 30693818

18. Parikh Milan N., Li Hailong, He Lili. Enhancing Diagnosis of Autism With Optimized Machine Learning Models and Personal Characteristic Data. Front Comput Neurosci. 2019; 13: 9. https://doi.org/10.3389/fncom.2019.00009 PMID: 30828295

19. Duda M, Ma R, Haber N, Wall D P. Use of machine learning for behavioral distinction of autism and ADHD. Transl Transl Psychiatry. 2016 Feb 9; 6(2):e732. https://doi.org/10.1038/tp.2015.221 PMID: 26859815

20. Bone Daniel, Matthew S Goodwin, Matthew P Black, Chi-Chun Lee, Kartik Audhkhasi, Shrikanth Narayanan. Applying machine learning to facilitate autism diagnostics: pitfalls and promises. J Autism Dev Disord. 2015 May; 45(5):1121–36. https://doi.org/10.1007/s10803-014-2268-6 PMID: 25294649

21. Ashwood K. L., Gillan N. et al. Predicting the diagnosis of autism in adults using the Autism-Spectrum Quotient (AQ) questionnaire. Psychological medicine. 2016. 46(12), 2595–2604. https://doi.org/10.1017/S0033291716001082 PMID: 27353452

22. Feczko E, Balba NM et al. Subtyping cognitive profiles in Autism Spectrum Disorder using a Functional Random Forest algorithm. Neuroimage. 2018 May 15; 172:674–688. https://doi.org/10.1016/j.neuroimage.2017.12.044 PMID: 29274502

23. David G.Amaral, Cynthia Mills Schumann, Christine WuNordahl. Neuroanatomy of autism. Trends in Neurosciences. 2008. Volume 31, Issue 3, Pages 137–145. https://doi.org/10.1016/j.tins.2007.12.005 PMID: 18258309

24. Brambilla Paolo, Hardan Antonio, Stefania Ucelli di Nemi, Jorge Perez, Jair C Soares, Francesco Barale. Brain anatomy and development in autism: review of structural MRI studies. Brain Res Bull. 2003 Oct 15; 61(6):557–69. https://doi.org/10.1016/j.brainresbull.2003.06.001 PMID: 14519452

25. Dierker Donna L, Eric Feczkoet al. Analysis of cortical shape in children with simplex autism. Cereb Cortex. 2015 Apr; 25(4):1042–51. https://doi.org/10.1093/cercor/bht294 PMID: 24165833

26. R Kucharsky Hiess R Alter et al. Corpus callosum area and brain volume in autism spectrum disorder: quantitative analysis of structural MRI from the ABIDE database. J Autism Dev Disord. 2015 Oct; 45 (10):3107–14. https://doi.org/10.1007/s10803-015-2468-8 PMID: 26043845

27. Lange Nicholas, Travers Brittany G, Bigler Erin Det al. Longitudinal volumetric brain changes in autism spectrum disorder ages 6–35 years. Autism Res. 2015 Feb; 8(1):82–93. https://doi.org/10.1002/aur.1427 Epub 2014 Nov 7. PMID: 25381736

28. Nordahl Christine Wu, Dierker DonnaMostafavi Iman et al. Cortical folding abnormalities in autism revealed by surface-based morphometry. J Neurosci. 2007 Oct 24; 27(43):11725–35. https://doi.org/10.1523/JNEUROSCI.0777-07.2007 PMID: 17959814

29. Shokouhi Mahsa, Justin H G Williams, Gordon D Waiter, Barrie Condon. Changes in the sulcal size associated with autism spectrum disorder revealed by sulcal morphometry. Autism Res. 2012 Aug; 5 (4):245–52. https://doi.org/10.1002/aur.1232 PMID: 22674695

30. Valk Sofie L, Martino Adriana Di, Milham Michael P, Bernhardt Boris C. Multicenter mapping of structural network alterations in autism. Hum Brain Mapp. 2015 Jun; 36(6):2364–73. https://doi.org/10.1002/hbm.22776 PMID: 25727858

31. Gregory L Wallace, Ian W Eisenberg, Briana Robustelli, Nathan Dankner, Lauren Kenworthy, Jay N Giedd, et al. Longitudinal cortical development during adolescence and young adulthood in autism spectrum disorder: increased cortical thinning but comparable surface area changes. J Am Acad Child Adolesc Psychiatry. 2015 Jun; 54(6):464–9. https://doi.org/10.1016/j.jaac.2015.03.007 PMID: 26004661

32. Ha Sungji, Sohn In-Jung, Kim Namwook, Sim Hyeon Jeong, Cheon Keun-Ah. Characteristics of Brains in Autism Spectrum Disorder: Structure, Function and Connectivity across the Lifespan. Exp Neurobiol. 2015 Dec; 24(4): 273–284. https://doi.org/10.5607/en.2015.24.4.273 PMID: 26713076

33. Ecker C, Shahidiani A, Feng Y, Daly Eet al. 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). 2014 Sep; 121(9):1157–70. https://doi.org/10.1007/s00702-014-1207-1 PMID: 24752753

34. Chanel Guillaume, Pichon Swann, Conty Laurence, Berthoze Sylvie, Chevallier Coralie, ezes Julie Gr'. Classification of autistic individuals and controls using cross-task characterization of fMRI activity. NeuroImage: Clinical, Volume 10, 2016, Pages 78–88. https://doi.org/10.1016/j.nicl.2015.11.010 PMID: 26793434

35. Chen Heng, Duan Xujun, Liu Feng, Lu Fengmei, Ma Xujing, Zhang Youxue, et al. Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivity–A multi-center study. Prog Neuropsychopharmacol Biol Psychiatry. 2016 Jan 4; 64:1–9. https://doi.org/10.1016/j.pnpbp.2015.06.014 PMID: 26148789

36. Gori Ilaria, Giuliano Alessia, Muratori Filippo et al. Gray matter alterations in young children with autism spectrum disorders: comparing morphometry at the voxel and regional level. J Neuroimaging. Nov-Dec 2015; 25(6):866–74. https://doi.org/10.1111/jon.12280 PMID: 26214066

37. Jin Yan, Wee Chong-Yaw, Shi Feng, Thung Kim-Han, Ni Dong, Yap Pew-Thian, Shen Dinggang. Identification of infants at high-risk for autism spectrum disorder using multiparameter multiscale white matter connectivity networks. Hum Brain Mapp. 2015 Dec; 36(12):4880–96. https://doi.org/10.1002/hbm.22957 PMID: 26368659

38. Odriozola Paola, Lucina Q Uddin, Charles J Lynch, John Kochalka, Tianwen Chen, Vinod Menon. Insula response and connectivity during social and non-social attention in children with autism. Soc Cogn Affect Neurosci. 2016 Mar; 11(3):433–44. https://doi.org/10.1093/scan/nsv126 PMID: 26454817

39. Colleen P Chen, Christopher L Keown, Afrooz Jahedi, Aarti Nair, Mark E Pflieger, Barbara A Bailey, et al. Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism Neuroimage Clin. 2015 Apr 9; 8:238–45. https://doi.org/10.1016/j.nicl.2015.04.002 PMID: 26106547

40. Rahman M. M., Usman O. L., Muniyandi R. C., Sahran S., Mohamed S., & Razak R. A. A Review of Machine Learning Methods of Feature Selection and Classification for Autism Spectrum Disorder. *Brain sciences*, 10(12), 949. 2020. https://doi.org/10.3390/brainsci10120949 PMID: 33297436

41. Craddock Cameron, Benhajali Yassine, Chu Carlton et al. The Neuro Bureau Preprocessing Initiative: open sharing of preprocessed neuroimaging data and derivatives. In: Front. Neuroinform. 2013.

42. Guo X, Dominick KC, Minai AA, Li H, Erickson CA, Lu LJ. Diagnosing Autism Spectrum Disorder from Brain Resting-State Functional Connectivity Patterns Using a Deep Neural Network with a Novel Feature Selection Method. Front Neurosci. 2017 Aug 21; 11:460. https://doi.org/10.3389/fnins.2017.00460 PMID: 28871217; PMCID: PMC5566619.

43. Huf Wolfgang, Kalcher Klaudius, Roland N Boubela et al. On the generalizability of resting-state fMRI machine learning classifiers. Front Hum Neurosci. 2014 Jul 29; 8:502. https://doi.org/10.3389/fnhum.2014.00502 PMID: 25120443

44. Kelly Clare, Bharat B Biswal, Craddock R Cameron, Castellanos F Xavier, Michael P Milham. Characterizing variation in the functional connectome: promise and pitfalls. Trends Cogn Sci. 2012 Mar; 16 (3):181–8. https://doi.org/10.1016/j.tics.2012.02.001 PMID: 22341211

45. Li Baihua, Sharma Arjun, Meng James, Purushwalkam Senthil, Gowen Emma. Applying machine learning to identify autistic adults using imitation: An exploratory study. PLOS ONE. 2017.

46. Pradyumna Lanka Rangaprakash D, Dretsch Michael N, Katz Jeffrey S, Thomas S Denney Jr, Gopikrishna Deshpande. Supervised machine learning for diagnostic classification from large-scale neuroimaging datasets. Brain Imaging Behav. 2020 Dec; 14(6):2378–2416. https://doi.org/10.1007/s11682-019-00191-8 PMID: 31691160

47. Michalewicz Zbigniew, Schoenauer Marc. Evolutionary Algorithms. Encyclopedia of Information Systems. 2003.

48. Hanley John P, Rizzo Donna M, Buzas Jeffrey S, Eppstein Margaret J. A Tandem Evolutionary Algorithm for Identifying Causal Rules from Complex Data. Evol Comput. Spring 2020; 28(1):87–114. https://doi.org/10.1162/evco_a_00252 PMID: 30817200

49. Han Y., Prelock P. A., Coderre E. L., & Orr J. M. A pilot study using two novel fmri tasks: Understanding theory of mind and emotion recognition among children with ASD. 2021. bioRxiv. https://doi.org/10.1101/2021.03.24.436877

50. Rutter M., Bailey A., and Lord C. The Social Communication Questionnaire Manual. Western Psychological Services, 2003.

51. Elizabeth Carrow-Woolfolk. Comprehensive Assessment of Spoken Language-Second Edition. Pearson.2017.

52. Bracken Bruce A., McCallum R. Steve. Universal nonverbal intelligence test (2nd ed.). Western Psychological Services, 2016.

53. McCallum R.S. Handbook for Nonverbal Assessment. Springer, 2017.

54. Hutchins T.L., Prelock P.A., Chase W. Test-retest reliability of a theory of mind task battery for children with autism spectrum disorders. Focus Autism Other Dev Disabil. 2008.

55. T.L. Hutchins and P.A. Prelock. Technical Manual for the Theory of Mind Inventory-2. Copyrighted manuscript. Theoryofmindinventory.com. 2016.

56. Hutchins T.L., Prelock P.A., Bonazinga L. 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. 2012. https://doi.org/10.1007/s10803-011-1244-7 PMID: 21484516

57. Baron-Cohen S. An essay on autism and theory of mind. MIT Press. 1995.

58. Baron-Cohen S., Leslie A. M., Frith U. Does the autistic-child have a theory of mind. Cognition. 1985. https://doi.org/10.1016/0010-0277(85)90022-8 PMID: 2934210

59. Lerner Matthew D, Hutchins Tiffany L, Prelock Patricia A. Brief report: preliminary evaluation of the theory of mind inventory and its relationship to measures of social skills. J Autism Dev Disord. 2011 Apr; 41 (4):512–7. https://doi.org/10.1007/s10803-010-1066-z PMID: 20628800

60. Hadwin Julie, Perner Josef. Pleased and surprised: Children's cognitive theory of emotion. British Journal of Developmental Psychology. 1991.

61. Hillier A., Allinson L. Understanding embarrassment among those with autism: Breaking down the complex emotion of embarrassment among those with autism. Journal of Autism and Developmental Disorders. 2002. https://doi.org/10.1023/a:1021259115185 PMID: 12553594

62. Ruffman T., Keenan T.R. The belief-based emotion of surprise: The case for a lag in under- standing relative to false belief. Developmental Psychology. 1996.

63. Seider B. A developmental analysis of elementary school-aged children's concepts of pride and embarrassment. Child Development. 1988.

64. Getting an MRI: A Cartoon for Kids. https://www.youtube.com/watch?v=QPa6KFL1Nw%5C&t=139s%5Ccf2.

65. Glasser Matthew F, Sotiropoulos Stamatios N, Wilson J Anthonyet al. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage. 2013 Oct 15; 80:105–24. https://doi.org/10.1016/j.neuroimage.2013.04.127 PMID: 23668970

66. FreeSurfer. https://surfer.nmr.mgh.harvard.edu/fswiki/aparcstats2table.

67. Fischl Bruce, David H Salat Evelina Busa et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002 Jan 31; 33(3):341–55. https://doi.org/10.1016/s0896-6273(02)00569-x PMID: 11832223

68. Nuzzo Regina. Scientific method: Statistical errors. Nature. 2014. https://doi.org/10.1038/506150a PMID: 24522584

69. Wasserstein R.L., Schirm A.L., Lazar N.A. Moving to a World Beyond p <0.05. The American Statistician. 2019.

70. Wasserstein R. L., Schirm A. L., and Lazar N. A. The ASA 92s statement on p-values: context, process, and purpose. The American Statistician. 2016.

71. Mucherino A., Papajorgji P.J., Pardalos P.M. k-Nearest Neighbor Classification. In: Data Mining in Agriculture. Springer Optimization and Its Applications. Springer, New York, NY. 2009. https://doi.org/10.1007/978-0-387-88615-2 4.

72. Mucherino A., Papajorgji P.J., Pardalos P.M. Cross-Validation. Encyclopedia of Machine Learning. Springer, Boston, MA. 2011. https://doi.org/10.1007/978-0-387-30164-8 190.

73. Li D., Yang W., Wang S. Classification of foreign fibers in cotton lint using machine vision and multiclass support vector machine. Electron. Agric. 2010

74. Zhang Y., Wu L. Classification of fruits using computer vision and a multiclass support vector machine. Sensors. 2012. https://doi.org/10.3390/s120912489 PMID: 23112727

75. Nickel K. et al. Volume Loss Distinguishes Between Autism and (Comorbid) Attention- Deficit/Hyperactivity Disorder FreeSurfer Analysis in Children. Frontiers in Psychiatry.2018.

76. Postema M. C., van Rooij D., Anagnostou E. Altered structural brain asymmetry in autism spectrum disorder in a study of 54 datasets. Nat Commun. 2019.

77. Zielinski B.A. et al. Longitudinal changes in cortical thickness in autism and typical development. Brain. 2014. https://doi.org/10.1093/brain/awu083 PMID: 24755274

78. Kana R.K., Maximo J.O., Williams D.L. et al. Aberrant functioning of the theory-of-mind network in children and adolescents with autism. Molecular Autism. 2015. https://doi.org/10.1186/s13229-015-0052-x PMID: 26512314

79. Takeuchi Mayumi, Harada Masafumi, Nishitani Hiromu. Deficiency of "theory of mind" in autism estimated by fMRI. International Congress Series. April 2002, Pages 737–740.

80. O'Nions E. et al. Neural bases of theory of mind in children with autism spectrum disorders and children with conduct problems and callous-unemotional traits. Developmental Science. 2014. https://doi.org/10.1111/desc.12167 PMID: 24636205

81. Ahmed Z., Mohamed K., Zeeshan S., Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. The journal of biological databases and curation. 2020 https://doi.org/10.1093/database/baaa010 PMID: 32185396

82. Editorial. Ascent of machine learning in medicine. Nat. Mater. 2019. https://doi.org/10.1038/s41563-019-0360-1 PMID: 31000807

83. Shah P., Kendall F., Khozin S. et al. Artificial intelligence and machine learning in clinical development: a translational perspective. Digit. Med. 2019. https://doi.org/10.1038/s41746-019-0148-3 PMID: 31372505

84. Davenport T., Kalakota R. The potential for artificial intelligence in healthcare. Future healthcare journal. 2019. https://doi.org/10.7861/futurehosp.6-2-94 PMID: 31363513