Uncovering biologically relevant Autism subtypes using advanced machine learning techniques: ASD biotypes

Chris "Gabby" Vento[1*], Joseph Cubells, Larry Young (posthumous), Elissar Andari

[1]Division, Department of Psychiatry, The University of Toledo, Toledo, OH 43614

*Corresponding author: christopher.vento@rockets.utoledo.edu

Keywords: Autism, Biotypes, Multimodal, Machine Learning, Random Forest, Resting-State Functional Connectivity, Superior Temporal Sulcus, Salience Network

Published: 22 November 2024

Background: Autism spectrum disorder (ASD) is characterized by main deficits in social interaction, and social communication. However, this vague definition of ASD does not encompass the wide heterogeneity of its phenotypical presentation. There is no current treatment available for the core deficits of ASD. It is critical to identify ASD subtypes that are biologically relevant and that respond to personalized treatment.

Objectives: The objective of this study is to use a multimodal approach to create biologically relevant subtypes. We aim to use behavioral and personality data to derive original subtypes using machine learning approaches, and then use biological data (brain function, epigenetics) and clinical data to confirm and validate the putative ASD subtypes.

Methods: A total of 114 adult men (18-45 years old), including 74 neurotypical (NT) and 40 ASD were recruited at Emory University and completed a series of behavioral tasks (such as the personality test (NEO-PI-R), reading the mind in the eyes (RMET), Symptom Checklist 90-revised questionnaire (SCL-90), intelligence quotient (IQ), and Broader Autism Phenotype Questionnaire (BAPQ)); clinical diagnostic tests (such as Autism Diagnosis Interview-Revised (ADI-R)); and biological measures (such as resting-state functional connectivity rs-fMRI).

We used independent component analysis and GIFT software to process rs-fMRI data. For the machine learning analysis, we used random forest tree algorithm to classify ASD and NT. We included NEO-PI-R and RMET in the main classifier. We validated the model with K-Fold Cross validation. The resulting model feature contribution was described using shapely values. ASD subtypes were then derived based on the shapely values of the individual data points utilizing K-means clustering. Subtypes were then validated using the other behavioral, clinical and brain function data, and utilizing t-tests and Cohen's d effect sizes.

Results: The random forest tree model classified ASD and NT with an average accuracy of 80%. Top features included personality domains such as extraversion and neuroticism. The resulting k-means clustering derived 3 different subtypes (Subtypes 1, 2, 3). T-tests indicated significant differences in the following measures: ADI-R repetitive behaviors, BAPQ, IQ, SCL-90, neuroticism, extraversion, RMET), and rs-FC between the superior temporal sulcus (STS, theory of mind network) and anterior cingulate (ACC) and insula (salience network).

Conclusion: Our results suggest that ASD subtype 1 is characterized by high neuroticism, lower warmth, higher scores on RMET, higher IQ, and higher rsFC between STS and salience network. Subtype 2 is characterized by high neuroticism, high repetitive behaviors, and lower rsFC between STS and salience network. Subtype 3 was found to be close to neurotypicals. These results are very promising, and the next step is to examine whether these putative subtypes are biologically relevant and whether specific subtypes respond better to certain pharmacological treatments.

