A data-driven method to gain pluralistic insights into resting state data

K.R. Patil¹, S. Arnold-Anteraper², G. Joshi³

¹Neuroeconomics Lab, MIT Sloan School, Cambridge, MA, USA, ²McGovern Institute for Brain Research, MIT, Cambridge, MA, USA, ³Massachussetts General Hospital, Boston, MA, USA

Background: The clinical presentation of Autism Spectrum Disorder (ASD) is highly heterogeneous. Machine learning methods, e.g. support vector machine (SVM), are used to identify discriminatory patterns. Such commonly used approaches identify solitary profiles that is limiting, as there can exist multiple combinations of ROIs that can have discriminatory power. This plurality is ignored by the traditional methods. We explored pattern recognition of atypical functional architecture of Default Mode Network (DMN) and the 84 Brodmann Area Regions of Interest (BA ROIs) using machine learning approaches to unravel diagnostic neural markers of ASD. We propose a method that can separate the two groups and provide insights into the potential plurality of the solution space.

Methods: 6 minutes of resting-state fMRI data were acquired from 15 young adult males with high-functioning ASD and from 16 age-, sex-, and IQ-matched HC. Noise corrected times-series (DMN seeds and 84 BA ROIs) were generated using the CONN Toolbox using SPM8 for spatial pre-processing. The method can be divided into the following steps: (1) Identify combinations of the ROIs that lead to high generalization accuracy. We use binary genetic algorithm (GA) that encodes the ROIs as an indicator (1=selected and 0=not selected). The fitness of a ROI combination is assessed by building a SVM model on the selected ROI and performing nested cross-validation. (2) Run step (1) multiple times (e.g. 100) and record which ROIs were selected in each run. Record the corresponding weight assigned by the SVM, remove runs that gave similar results. (3) From the results in step (2) remove ROIs that flip the weight sign, these ROI provide inconsistent evident for the conditions. (4) Rank the ROIs by how many runs they appear.

Results: Taking the small number of subjects in our data set into consideration, we used 2-fold outer with 2-fold inner nested cross-validation to avoid overfitting. We ran the GA 100 times, cleaned results contained 91 runs and 59 ROIs. Primary Auditory Cortex (PAC, BA 42 R) appeared in 39 runs with positive weight (higher relation with DMN provides more evidence for ASD) followed by the Ventral Anterior Cingulate Cortex and Superior Temporal Gyrus (BA 24, 22 both R) appearing in 21 and 11 runs. All other ROIs appeared less than 10 times. The accuracy for the runs with PAC selected was significantly higher than other runs (mean of 83% vs 77%, t-test: t=7.75, P<1e-10).

Conclusions: Results suggest a strengthened connection between PAC and DMN in the ASD group. Presence of multiple solutions with high accuracy suggests possible existence of multiple independent systems that can explain difference between ASD and HC. Overall, the proposed method is promising to analyze complex data where there can be plurality in the solutions.