Diagnostic Classification of Autism Using Resting-State Fmri Data and Conditional Random Forest

Conference Paper · July 2018

CITATION
1

READS
318

4 authors:

Jack Fredo
San Diego State University
20 PUBLICATIONS **43** CITATIONS
SEE PROFILE

Afrooz Jahedi
San Diego State University
14 PUBLICATIONS **209** CITATIONS
SEE PROFILE

Maya Anne Reiter
University of Washington Seattle
11 PUBLICATIONS **115** CITATIONS
SEE PROFILE

Ralph-Axel Müller
San Diego State University
181 PUBLICATIONS **7,699** CITATIONS
SEE PROFILE

Diagnostic Classification of Autism using Resting-State fMRI Data and Conditional Random Forest

A.R. Jac Fredo[1], Afrooz Jahedi[2], Maya Reiter[3] and Ralph-Axel Müller[4]

[1]SERB-IUSSTF Post-Doctoral Fellow with the Brain Development Imaging Lab at San Diego State University in San Diego, CA, USA.
phone: (+1)7604988740; fax: (619) 594-0176; e-mail: jack247029@sdsu.edu
[2]Computational Science PhD student in San Diego State University/Claremont Graduate University's Joint Doctoral Program, in San Diego, CA, USA.
e-mail: Afrooz.Jahedi@gmail.com
[3]Clinical Psychology PhD student in San Diego State University/UC San Diego's Joint Doctoral Program, in San Diego, CA, USA.
e-mail: maya.a.reiter@gmail.com
[4]Director of the Brain Development Imaging Laboratory, and professor of Psychology, San Diego State University, San Diego, CA, USA.
e-mail: rmueller@mail.sdsu.edu

***Abstract*—Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that is associated with atypical connectivity within and between brain regions. In this study, we attempted to classify functional Magnetic Resonance Images (fMRI) of Typically Developing (TD) and ASD participants using conditional random forest and random forest. Resting-state fMRI images of TD and ASD participants (N=320 for training and N=80 for validation) were obtained from the Autism Imaging Data Exchange; ABIDE-I, ABIDE-II. Images were preprocessed using a standard pipeline. A Functional Connectivity (FC) matrix was calculated using 237 cortical, subcortical, and cerebellar Regions of Interest (ROIs). The dimensionality of the FC matrix was reduced using conditional random forests and at each dimension classification accuracy was tested using random forests. Results suggest that in the current dataset, the random forest is able to classify the TD and ASD with a peak accuracy of 65% using 143 features. Remarkably, the Cingulo-Opercular Task Control (COTC) region contributed the highest number of features linked to more accurate classification, and connectivity between COTC and the dorsal attention network distinguished ASD and TD participants.**

Keyword: *Autism Spectrum Disorder; fMRI; Functional Connectivity; Conditional Random Forest; Random Forest.*

I. Introduction

Autism Spectrum Disorder (ASD) is a neuro developmental disorder of brain characterized by social, behavioral and communication defects [1]. For many, ASD is a lifelong condition and the underlying neural mechanisms remain poorly understood [2]. ASD is primarily diagnosed based on behavioral measures, which is a subjective and time-consuming process [3]. Neuroimaging data evaluated with Functional Connectivity (FC) and Machine Learning (ML) may help to identify brain biomarkers corresponding to ASD [4].

Functional Magnetic Resonance Imaging (fMRI) is a powerful tool for examining the abnormal neurobiological functions in ASD. It visualizes the regions of brain with excellent contrast, spatial and temporal resolution. fMRI depicts localized hemodynamic correlates of neural activity linked to relevant neurocognitive processes [5]. Modeling approaches have revealed that patterns of brain FC calculated from fMRI may serve as a biomarker to classify ASD [6].

Functional connectivity provides the temporal correlation between spatially remote neurophysiological events. The potential neural signatures of ASD are cortical under-connectivity and local over-connectivity [7]. Reports show that decreased [6] and increased [8] functional connectivity are found throughout the brain in ASD. Hypothesis driven approaches examining only few ROIs have not yielded overarching biomarkers that differentiate individuals with ASD from TD individuals. Supervised ML methods can identify significant brain FC patterns and build a model to classify the ASD [4].

Supervised ML methods build classifier models from high dimensional FC matrices. Most studies that combine brain imaging and ML have applied methods such as support vector machine or Gaussian naïve Bayes classifiers [9]. Recently, Random Forest (RF) and Conditional Random Forest (CRF) have been used to classify the large cohort of ASD and Typical Developing (TD) participants with higher accuracy [10].

The goal of this study was to build a classifier model using conditional random forest to classify TD and ASD participants, and to determine whether sample characteristics, influence accuracy of classification. A subset of participants was set aside for independent validation of the classifier. Another goal of this study was to examine the features most influential in achieving higher classification accuracy.

II. Methodology

A. Database

The data analyzed in this study was obtained from Autism Imaging Data Exchange-I (ABIDE-I), Autism Imaging Data Exchange-II (ABIDE-II) publically available datasets [11]. Only male individuals whose age ranged between 7-18 years, and whose resting-state fMRI data was acquired with eyes open in the scanner were included. Additionally, only participants with low motion were included in the current study. In both the training and validation samples, ASD and

TD groups were well matched on motion and age, however, Intelligence Quotient scores were significantly lower in the ASD group. Performance Intelligent Quotient (PIQ) replaced Full-scale Intelligent Quotient (FIQ) when FIQ was unavailable. The datasets for the training model and validation were selected randomly from each site in a proportion of 80:20. ABIDE scanning sites/groups with fewer than 10 participants were excluded from the analysis. After matching, the training model included a sample of 160 TD and 160 ASD participants, and the validation sample included 40 TD and 40 ASD participants. Table 1 presents the participant demographic information including age, motion and PIQ/FIQ.

TABLE I. PARTICIPANT INFORMATION OF SAMPLE SET

	ASD, Mean±SD (range)	TD, Mean±SD (range)	*p*-value (2 sample t-test)
Training sample			
N (Handedness)	160 (R-108, L-11, M-8, ND-33)	160 (R-121, L-9, M-5, ND-25)	-
Age (years)	12.16±2.76 (6.41-17.94)	12.03±2.9 (6.36-18.8)	0.7
Motion (mm)	0.094±0.041 (0.021-0.191)	0.09±0.037 (0.026-0.191)	0.32
PIQ/FIQ	106.8±16.67 (69-146)	111.9±12.78 (83-147)	0.002
Validation sample			
N (Handedness)	40 (R-24, L-1, M-1, ND-14)	40 (R-23, L-2, M-3, ND-12)	-
Age (years)	12.06±2.67 (7.15-17.5)	12.82±2.57 (8.06-18.21)	0.19
Motion (mm)	0.088±0.41 (0.035-0.172)	0.08±0.045 (0.027-0.177)	0.43
PIQ/FIQ	105.5±17.19 (77-145)	111.8±16.17 (62-147)	0.095

SD-Standard Deviation, N-Number of samples, R-Right, L-Left, M-Mixed, ND-No data

B. Pre-processing

A standard pipeline is followed to pre-process the fMRI data, which included motion correction, censoring of volumes with high head motion (0.5mm, after censoring, time series segments of less than 10 contiguous volumes were also removed.), special realignment of functional images, to a standard image (MNI152, 3 mm isotropic resolution), band-pass filtering (Butterworth $0.008 < f < 0.08$ Hz) and global signal regression [12-15]. All pre-processed images were visually inspected for artifacts and suitable brain coverage. Datasets with Root Mean Square Deviation (RMSD) higher than 0.2 mm and less than 80% of time points after censoring were also excluded from the analysis.

C. Image Processing

Mean time series were extracted from 237 cortical, subcortical, and cerebellar Regions of Interest (ROIs), and used to create a 237x237 Fisher-transformed Pearson correlation connectivity matrix for each participant. The final set of ROIs included 214 of 333 ROIs from Gordon's cortical atlas [16], and all 14 subcortical, as well as 9 of 26 cerebellar ROIs from the Harvard Oxford subcortical [17] and cerebellar [18] atlases. The following procedure was used to determine inclusion of each ROI: First, we identified all voxels of the brain which included BOLD signal in 95% of participants. Next, we calculated the percentage of voxels in each ROI that included true signal for 95% or more of participants. Only ROIs for which $\geq$95% of voxels contained true BOLD signal for $\geq$95% of participants were included in the final analysis; 136 ROIs (119 cortical, 17 cerebellar) were excluded because they did not meet these criteria. A group-level connectivity matrix was then calculated separately for the participants in the training and validation models, with matrix sizes of 320×27966 and 80×27966 respectively.

D. Classification

The high dimensional FC matrix (320×27966) was reduced using CRF in R package [19]. CRF with 2001 trees calculates the Conditional Permutation Importance (CPI) and is sorted at the end of each iteration. The features corresponding to the positive values of CPI that represent the most significant features are fed as the input to the tree in the next iteration. The process is repeated until the number of features is reduced to two. It has been suggested that CRF performs better in feature reduction, while RF is preferable for classification [20]. The classification model built using CRF was validated in an independent sample using an RF classifier with a tree number of 20001 [10].

III. RESULTS

A. Diagnostic prediction accuracies

The RF Out of Bag (OOB) error rate, sensitivity, and specificity rates that were obtained for the validation sample of 80 participants (for the models below 1100 features) are shown in Fig 1. The models that consisted of a higher number of features are not shown due to high OOB error rates. The minimum RF OOB error rate for training was achieved with 473 and 236 features. The classifier model obtained in the training model, which consisted of 143 features, was able to distinguish ASD from TD participants in the independent validation model with a peak accuracy of 65% (sensitivity and specificity were also 65%).

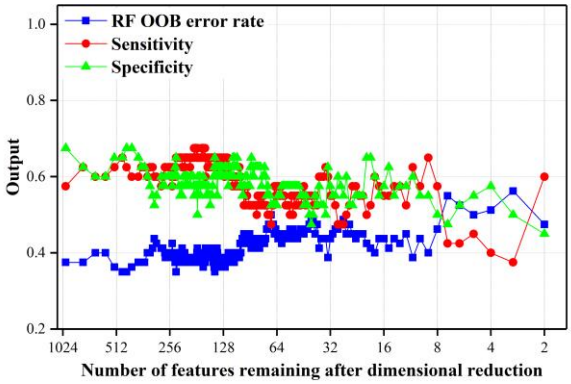


Figure 1. RF OOB error rate, sensitivity and specificity after each dimension reduction for the validation sample.

A. Informative networks

The most informative networks were identified based on the percentage of ROI participation in achieving peak

accuracy shown in Fig. 2. The most prominent networks (which together account for over 50% of classification importance) were the Cingulo-Opercular Task Control (COTC), Dorsal Attention (DA), Visual, Fronto-Parietal Task Control (FPTC), and Subcortical networks. The strongest network to network connection was between the COTC and DA networks, and is shown in Fig. 3, and contributed 6 informative features to the final classifier.

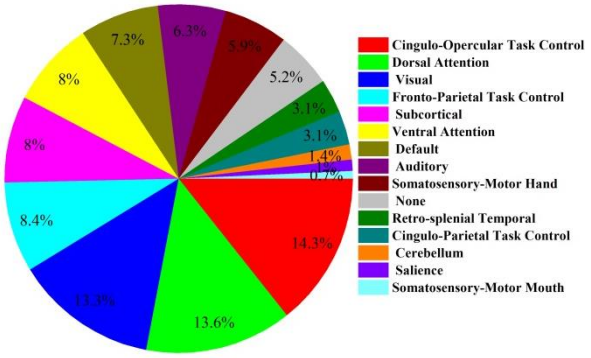


Figure 2. ROI participation in most informative features, separated by network

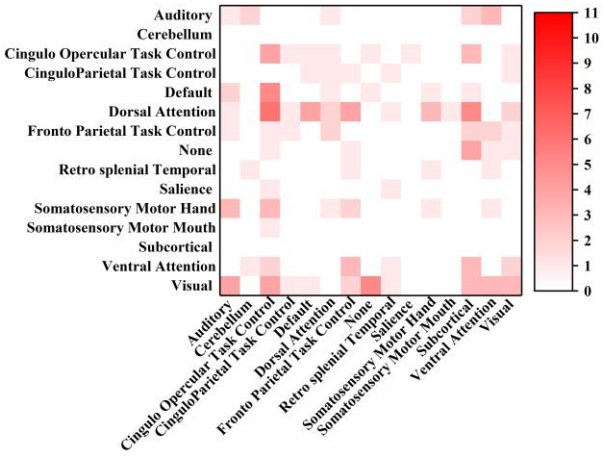


Figure 3. Heat map showing the number of features (ROI to ROI pairings from different networks) which contibuted to the diagnostic classifier.

IV. Discussions

Peak accuracy (65%) was achieved with the model that consisted of 143 features using CRF feature reduction and the RF classifier. Classification accuracy decreased further when the number of features included in the classifier was reduced. This suggests that important diagnostic information is lost when models become oversimplified [21].

Using the described methods the classification accuracy obtained in the present study was higher than that previously reported by other studies using the ABIDE data base [22, 23]. As suggested by [24], [25], classification accuracy may increase as sample size increases. Accuracy rates may also increase by improving the matching between training and validation sample [10], creating specific feature sets for different age and gender groups [26], including high severity ASD participants in the dataset [27], including behavioral measures [21], and improving matching demographic information within each site [9].

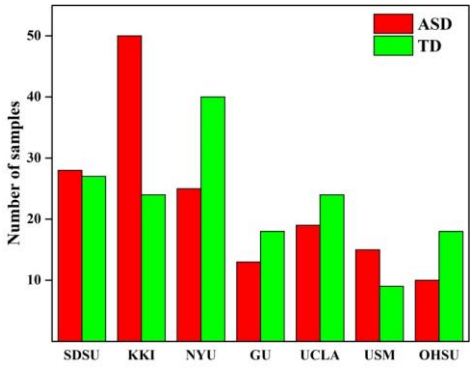
The COTC network contributed the highest number of features in the model that achieved peak accuracy. This appears consistent with a report by [2], also found the COTC network to be highly informative in diagnostic prediction of ASD. The most informative network to network connectivity was detected between the COTC and the DA networks. This finding is in agreement with a recent study by [28]. Due to the highly complex pattern and number of connections (143 features) included in the diagnostic classifier, it is difficult to pinpoint mechanisms that that provide a "simple" or "overarching" biomarker that could reliably distinguish individuals with ASDs. However, improving the accuracy of data-driven classifiers may bring us one step closer to achieving classification accuracies which are more comparable with current gold-standard clinical behavioral diagnostic tools using neurobiological measures such as fMRI.

V. Conclusion

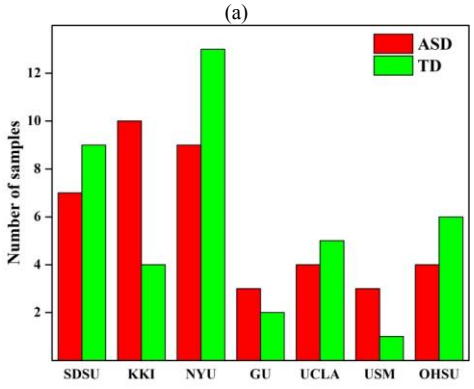
In this study, an RF based classifier model was built to distinguish ASD from TD participants. The performance of the model was tested using an independent validation sample set. The results showed that the model was able to classify the ASD and TD participants with accuracy, sensitivity, and specificity rates of 65%. The peak accuracy was achieved for a model consisting of 143 features. The COTC network, compared to other networks, contributed the highest number of features to achieve peak accuracy. The strongest network to network connections involved in classification were found between the COTC and the DA networks.

Appendix

The Number of participants included from each ABIDE site is shown in Supplementary Fig. S1 (below).



(a)



(b)

S1. Number of samples chosen from each ABIDE scanning site (a) Training model and (b) Validation models. Most of the participants were obtained from the SDSU, KKI and NYU groups.

ACKNOWLEDGMENTS

The Science Engineering and Research Board (SERB)/Indo-Us Science and Technology Forum (IUSSTF) provided funding for A.R. Jac Fredo, to complete this study. Ms. Reiter is funded by Autism Speaks Weatherstone fellowship #10609. This work was also supported by the National Institutes of Health (grant R01-MH081023, R01-MH1173).

REFERENCES

[1] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders*, American Psychiatric Association, 5th ed., Washington, DC, 2013.

[2] N. Yahata, J. Morimoto, R. Hashimoto, G. Lisi, K. Shibata, Y. Kawakubo, H. Kuwabara, M. Kuroda, T. Yamada, F. Megumi, H. Imamizu, J. E. Náñez Sr, H. Takahashi, Y. Okamoto, K. Kasai, N. Kato, Y. Sasaki, T. Watanabe, and M. Kawato, "A small number of abnormal brain connections predicts adult autism spectrum disorder", *Nat. Commun*., vol. 7, pp. 1-12, 2016.

[3] C. M. Murphy, C. E. Wilson, D. M. Robertson, C. Ecker, E. M. Daly, N. Hammond, A. Galanopoulos, I. Dud, D. G. Murphy, and G. M. McAlonan, "Autism spectrum disorder in adults: diagnosis, management, and health services development", *Neuropsychiatr Dis Treat*., vol. 12, pp. 1669-1686, 2016.

[4] L. M. Hernandez, J. D. Rudie, S. A. Green, S. Bookheimer, and M. Dapretto, "Neural Signatures of Autism Spectrum Disorders: Insights into Brain Network Dynamics", *Neuropsychopharmacology*, vol. 40, no. 1, pp. 171-189, 2015.

[5] G. S. Dichter, "Functional magnetic resonance imaging of autism spectrum disorders", *Dialogues Clin Neurosci*., vol. 14, no. 3, pp. 319-351, 2012.

[6] J. S. Anderson, J. A. Nielsen, A. L. Froehlich, M. B. DuBray, T. J. Druzgal, A. N. Cariello, J. R. Cooperrider, B. A. Zielinski, C. Ravichandran, P. T. Fletcher, A. L. Alexander, E. D. Bigler, N. Lange, J. E. Lainhart, "Functional connectivity magnetic resonance imaging classification of autism", *Brain*, vol. 134, no. 12, pp. 3742-3754, 2011.

[7] J. O. Maximo, E. J. Cadena, and R. K. Kana, "The implications of brain connectivity in the neuropsychology of autism", *Neuropsychol Rev*., vol. 24, no. 1, pp. 16-31, 2014.

[8] R. A. Muller, P. Shih, B. Keehn, J. R. Deyoe, K. M. Leyden, and D. K. Shukla, "Underconnected, but how? A survey of functional connectivity MRI studies in autism spectrum disorders", *Cereb. Cortex*., vol. 21, no. 10, pp. 2233-2243, 2011.

[9] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the ABIDE dataset", *NeuroImage: Clinical*, vol. 17, pp. 16-23, 2018.

[10] A. Jahedi, C. A. Nasamran, B. Faires, J. Fan, and R. A. Muller, "Distributed intrinsic functional connectivity patterns predict diagnostic status in large autism cohort", *Brain Connect*., vol. 7, no. 8, pp. 515-525, 2017.

[11] A. Di Martino, C. G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, B. Deen, S. Delmonte, I. Dinstein, B. Ertl-Wagner, D. A. Fair, L. Gallagher, D. P. Kennedy, C. L. Keown, C. Keysers, J. E. Lainhart, C. Lord, B. Luna, V. Menon, N. J. Minshew, C. S. Monk, S. Mueller, R. A. Müller, M. B. Nebel, J. T. Nigg, K. O'Hearn, K. A. Pelphrey, S. J. Peltier, J. D. Rudie, S. Sunaert, M. Thioux, J. M. Tyszka, L. Q. Uddin, J. S. Verhoeven, N. Wenderoth, J. L. Wiggins, S. H. Mostofsky, and M. P. Milham, "The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism", *Mol Psychiatry*, vol. 19, pp. 659-667, 2014.

[12] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. M. Brady, and P. M. Matthews, "Advances in functional and structural MR image analysis and implementation as FSL", *Neuroimage*, vol. 23, pp. 2018-219, 2004.

[13] J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Steps toward optimizing motion artifact removal in functional connectivity MRI; a reply to carp", *Neuroimage*, vol. 76, pp. 439-441, 2013.

[14] T. D. Satterthwaite, M. A. Elliott, R. T. Gerraty, K. Ruparel, J. Loughead, M. E. Calkins, S. B. Eickhoff, H. Hakonarson, R. C. Gur, R. E. Gur, and D. H. Wolf, "An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data", *Neuroimage*, vol. 64, 240-256, 2013.

[15] D. Cordes, V. M. Haughton, K. Arfanakis, J. D. Carew, P. A. Turski, C. H. Moritz, M. A. Quigley, and M. E. Meyerand, "Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data", *A.J.N.R. Am. J. Neuroradiol*, vol. 22, no. 7, pp. 1326-1333, 2001.

[16] E. M. Gordon, T. O. Laumann, B. Adeyemo, J. F. Huckins, W. M. Kelley, and S. E. Petersen, "Generation and evaluation of a cortical area parcellation from resting-state correlations". *Cerebral Cortex*, vol. 26, no. 1, pp. 288-303, 2016.

[17] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest", *Neuroimage*, vol. 31, no. 3, pp. 968-80, 2006.

[18] J. Diedrichsen, J. H Balsters, J. Flavell, E. Cussans, and N. Ramnani "A probabilistic MR atlas of the human cerebellum", *Neuroimage*, vol.46, pp. 39-46, 2009.

[19] C. Strobl, A. L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional variable importance for random forests", *BMC Bioinformatics*, vol. 9, pp. 1-11, 2008.

[20] R. H. Baayen, "*Multivariate Statistics*" in Research Methods in Linguistics, R. Podesva, and D. Sharma, Ed, Cambridge, Cambridge University Press, 2013, pp. 337-372.

[21] M. Plitt, K. A. Barnes, and A. Martin, "Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards", *NeuroImage: Clinical*, vol. 7, pp. 359-366, 2015.

[22] J. A. Nielsen, B. A. Zielinski, P. T. Fletcher, A. L. Alexander, N. Lange, E. D. Bigler, J. E. Lainhart, and J. S. Anderson, "Multisite functional connectivity MRI classification of autism: ABIDE results", *Front Hum Neurosci*, vol. 7, 2013.

[23] S. Rane, E. Jolly, A. Park, H. Jang, and C. Craddock, "Developing predictive imaging biomarkers using whole-brain classifiers: Application to the ABIDE I dataset", *Project Report Research Ideas and Outcomes*, vol. 3, pp.1-5, 2017.

[24] T. Iidaka, "Resting state functional magnetic resonance imaging and neural network classified autism and control", *Cortex*, vol. 63, pp. 55-67, 2015.

[25] A. Abraham, M. Milham, A. Di Martino, R. C. Craddock, D. Samaras, B. Thirion, and G. Varoquaux, "Deriving robust biomarkers from multi-site resting-state data: an autism-based example", *NeuroImage*, vol. 147, pp. 736-745, 2017.

[26] V. Subbaraju, M. B. Suresh, S. Sundaram, and S. Narasimhan, "Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging : A spatial filtering approach", *Medical Image Analysis*, vol. 35, vol. 375-389, 2017.

[27] C. L. Keown, P. Shih, A. Nair, N. Peterson, M. E. Mulvey, and R. A. Müller, "Local functional overconnectivity in posterior brain regions is associated with symptom severity in autism spectrum disorders, *Cell Rep*, vol. 5, no. 3, pp. 567-572, 2013.

[28] N. de Lacy, D. Doherty, B. H. King, S. Rachakonda, and V. D. Calhoun, "Disruption to control network function correlates with altered dynamic connectivity in the wider autism spectrum", *Neuroimage Clin*., vol. 15, pp. 513-524, 2017.