MALIS Project Proposal

.

Machine Learning for Autism Spectrum Disorder Diagnosis

Giulia Lorini
Gaganjot Shan
Dario Ferrero

October 28, 2021

1 Motivation

Autism Spectrum Disorder (ASD) is a range of neurodevelopmental disorders including classical autism, Asperger syndrome, and pervasive developmental disorder not otherwise specified (PDD-NOS), characterized by impaired social skills, repetitive behaviors, sensory issues, and language delay. More than 1% of the population falls into this spectrum, with a high imbalance between the sexes, males being 4 to 5 times more likely to be affected than females[1].

The high heterogeneity of ASD makes it hard to define diagnostic criteria that can be applied to identify affected children as soon as possible to select optimal treatments. Currently, ASD diagnosis involves long processes and multiple specialists' evaluations, using behavioural assessment instruments. Application of Machine Learning methods to identify the underlying brain mechanisms could significantly speed up the diagnostic process.[1][2]

If the data allows it, it could also be interesting to identify the different manifestations of the disorder in females with respect to males, since females are known to have a completely different neuropathology, which could be a reason why they are less affected, or possibly just fail to be diagnosed using the current male-based criteria [3].

2 Method

We have identified the **ABIDE (Autism Brain Imaging Data Exchange)** project[2] which offers 2 different databases: the older ABIDE I containing 1112 datasets, including 539 from individuals with ASD and 573 from typical controls, for which the Preprocessed Connectomes Project created a pre-processing script [4][5], and the newer and better-characterized ABIDE II database [6], collecting 1114 datasets from 521 individuals with ASD and 593 controls.

We haven't yet decided which one to use, since at this stage we don't know what pre-processing entails and if it would be a significant (and allowed) advantage to start from there, and if it's worth it despite the fact that the ABIDE II database might be of better quality. For instance, a previous work [7] using the ABIDE I data concluded that there weren't enough female samples there to perform a comparative analysis, and actually proceeded to discard those minority samples from their training set.

Previous Machine Learning research on this topic[7] has pointed to some of the most effective learning algorithms used when working on image data for these purposes: we could explore Support Vector Machines and Gaussian Naïve Bayes.

3 Intended Experiments

We plan on testing the performance of our built model by applying k-fold cross-validation. We repute this step necessary in order to check for possible overfitting over the data, which is common given its natural high number of features per sample. Our goal therefore remains to maximize the final testing accuracy.

In order to explore the differences between male and female subjects we could train separate models on the two partitions of our dataset to compare performance and overall accuracy when testing each model with samples of the same versus the opposite gender.

References

[1] Gene Blatt, "autism", Encyclopedia Britannica, 9 Sep. 2021, https://www.britannica.com/science/autism, Accessed 27 October 2021.

[2] Child Mind Institute, "Autism Brain Imaging Data Exchange", ABIDE, Child Mind Institute, http://fcon_1000.projects.nitrc.org/indi/abide/, Accessed 27 October 2021.

[3] Kaat Alaerts, Stephan P. Swinnen, and Nicole Wenderoth., "Sex Differences in Autism: A Resting-State Fmri Investigation of Functional Brain Connectivity in Males and Females.", Social cognitive and affective neuroscience., U.S. National Library of Medicine., Accessed October 27, 2021., https://pubmed.ncbi.nlm.nih.gov/26989195/.

[4] Cameron Craddock, Pierre Bellec., "ABIDE Preprocessed.", Preprocessed Connectomes Project., Accessed October 27, 2021., http://preprocessed-connectomes-project.org/abide/.

[5] Cameron Craddock, Yassine Benhajali, Carlton Chu, Francois Chouinard, Alan Evans, András Jakab, Budhachandra Singh Khundrakpam, John David Lewis, Qingyang Li, Michael Milham, Chaogan Yan, Pierre Bellec, "The Neuro Bureau Preprocessing Initiative: open sharing of preprocessed neuroimaging data and derivatives.", In Neuroinformatics 2013, Stockholm, Sweden.

[6] Child Mind Institute., "Autism Brain Imaging Data Exchange II", ABIDE. Child Mind Institute., http://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html., Accessed 27 October 2021.

[7] Pegah Kassraian-Fard, Caroline Matthis, Joshua H. Balsters, Marloes H. Maathuis, Nicole Wenderoth, "Promises, Pitfalls, and Basic Guidelines for Applying Machine Learning Classifiers to Psychiatric Imaging Data, with Autism as an Example", Frontiers in Psychiatry, vol. 7, 2016, ISSN-1664-0640, doi:10.3389/fpsyt.2016.00177