Diagnosing Autism using individual brain regions from rs-fMRIs

Jason Bard, Pedro Maia

Department of Mathematics, The University of Texas at Arlington, Arlington, TX 76019

Abstract

Autism spectrum disorder (ASD) can be characterized by impairment of social ability and by repetitive interests and/or behaviors. As of 2022, about 1 in 36 children have been identified as having the disorder. The Autism Brain Imaging Data Exchange (ABIDE) is a collection of brain scans of more than 1000 individuals, about half of whom have been diagnosed with ASD. This dataset contains both phenotypic data of individual patients as well as functional MRIs collected at resting state (rs-fMRIs), which reflects changes in blood flow and brain activity of individual patients. In this study, the Python package Nilearn was used to transform these rs-fMRIs into time series, using the Harvard-Oxford Atlas via a masker to determine potential regions of interest, such as the amygdala or the orbitofrontal cortex (OFC). The time series from these regions were then used to train a multilayer perceptron (MLP) model to produce a binary classification result of whether an individual patient has ASD. Results from training a model on the OFC time series indicate an accuracy of up to 81%. Future results may be improved using CNNs on all regions and by further optimizing hyperparameters.

Background

ASD is typically diagnosed in patients between 18-36 months of age, and the severity of symptoms can vary widely, with some individuals having relatively mild symptoms while others have severe impairments. While there is currently no known cure, early intervention and treatments, such as behavioral therapy, speech therapy, and other specifically tailored activities, can help affected individuals develop skills and abilities that allow them to better navigate social situations and ideally to function independently. Research has shown that there are differences in brain structure and function in individuals with ASD, and there is ongoing research aimed at identifying the specific neural mechanisms underlying the disorder.[1] The ABIDE dataset is a collection of publicly available brain imaging data provided by dozens of universities and medical imaging sites. It consists of over a thousand patients near-evenly split into two groups: those in a healthy control group (HC), and those who have been diagnosed with ASD. The dataset as a whole includes both resting-state and task-based fMRI data in addition to clinical and demographic information about the participants, such as age, gender, IQ, handedness, and diagnostic information relating to treat scores and researchers' and interviewers' observations.[2]

Methods

Data was loaded directly from Nilearn, including 80 patients' data from NYU and the Harvard-Oxford atlas as a mask. The patients' functional brain scans were taken to be the full scale of the input from which other smaller inputs could be drawn. These scans were then smoothed by removing background noise and then were run through a mask, which is an object Nilearn uses to identify the overall brain activity in each region of interest using the smoothed image. This process produced the complete time series that were utilized. Specifically, in these trials, the time series for the orbitofrontal cortex (OFC) were used, with an overall length of 176 features, representing an actual time of 354.64 seconds. The phenotypic data containing the diagnosis group of the corresponded patients was also extracted and used as the target variable. The data was then split into a training group of 80% of the total data and a testing group of 20%. Finally, the input and target variables were used to train an MLP classifier using the Python package scikit-learn (sklearn). Various models of a similar structure were trained and tested, using accuracy as the metric by which each model could be compared. Precision and recall were also measured.

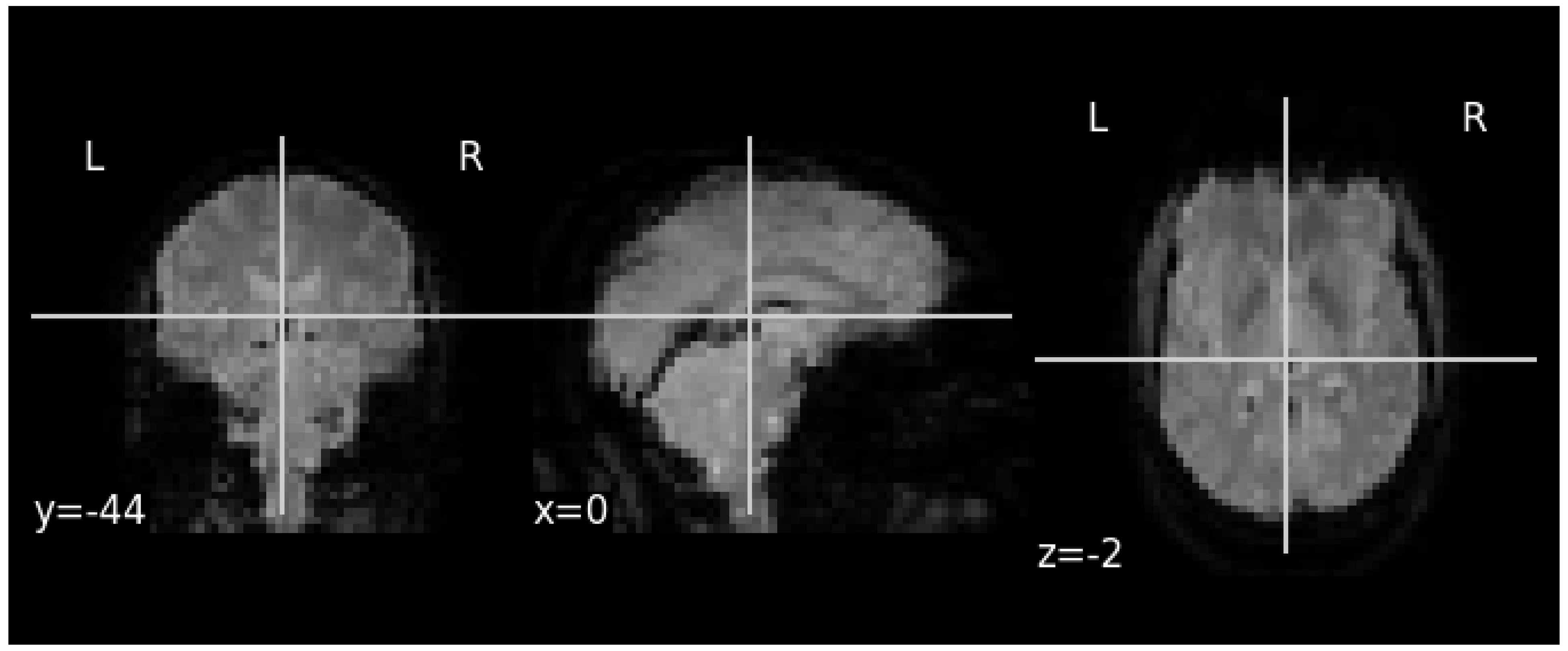

Figure 1. Three cross-sections of one slice of an rs-fMRI. Views for y = -44, x = 0, and z = -2.

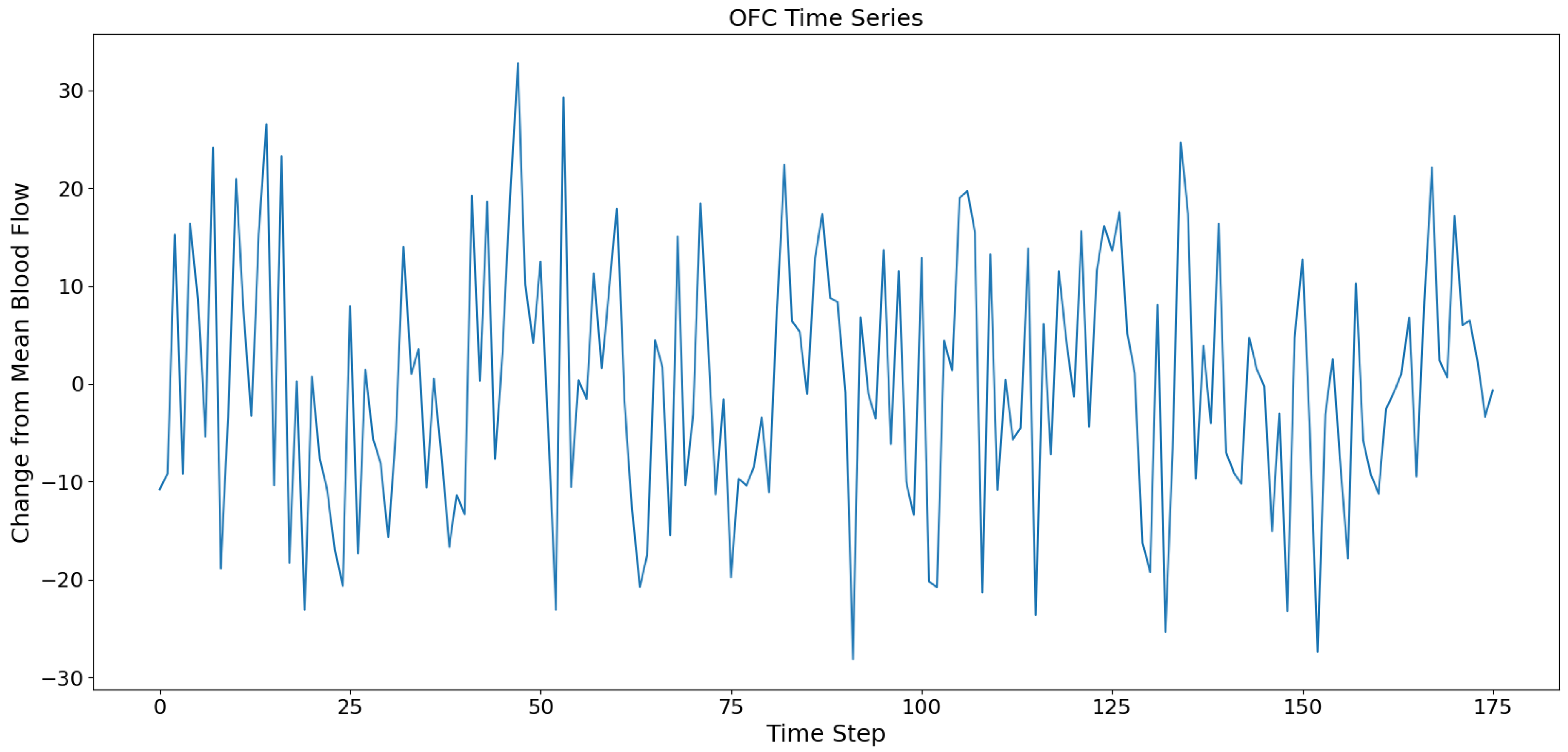

Figure 2. A graph plotting individual time steps versus change from mean blood flow in the OFC of one patient.

Results

After 1000 simulations of running the classifier model, results consistently showed a normal distribution around 50% accuracy, with a maximum accuracy of 81.25%. Precision and recall each showed similar results, except both categories showed a maximum of 100% during many runs. The study was limited by the small sample size due to hardware limitations; any sample size larger than 100 needs at least 32 GB of RAM to run, especially during the smoothing process. The entirety of the dataset takes up over 100 GB of disk space storage. Using 80 patients took up only 8 GB that was able to be loaded into Google Colaboratory, the working environment in which this study was done. In addition, patients in this study were from only one center; using patients from multiple brains scan centers as provided in the dataset may also contribute to a higher accuracy. Furthermore, the utilization of the subcortical atlas could explore regions further embedded in the brain, such as the amygdala, which may also contribute to a more successful model.

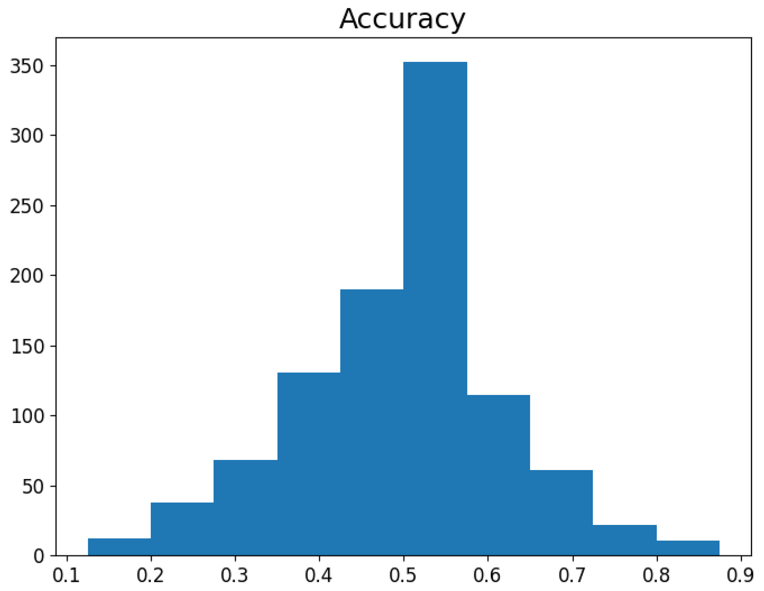

Figure 3. The histogram of accuracies obtained after 1000 applications of the same MLP model.

Acknowledgements

Thank you to Dr. Pedro Maia and Dr. Amir Farbin for guidance during this project. In addition, a big thanks to Khitam Aqel, Rasha Alsaeed, Bavithra Lakshmanasamy, and Humza Humayun for assistance in coding and talking through this project during our weekly meetings.

References

[1] Ha, Sungji, et al. "Characteristics of Brains in Autism Spectrum Disorder: Structure, Function and Connectivity across the Lifespan." *Experimental Neurobiology*, U.S. National Library of Medicine, Dec. 2015, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4688328.

[2] "ABIDE II." *ABIDE*, NITRC, 27 Mar. 2017, http://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html.