Transfer Learning: Leveraging Big Data for Prediction in Clinical (Stuttering) Datasets

Poster No:

2733

Submission Type:

Abstract Submission

Authors:

Saige Rutherford[1], Mike Angstadt[1], Chandra Sripada[1], Chelsea Johnson[2], Soo-Eun Chang[1]

Institutions:

[1]University of Michigan, Ann Arbor, MI, [2]Michigan State University, East Lansing, MI

Introduction:

Large datasets, consisting of hundreds or thousands of subjects, are becoming the new data standard within the neuroimaging community. Big data creates numerous benefits, most importantly, allowing for more ambitious statistical analyses compared to smaller studies. Many of these big datasets have focused on healthy populations. The heterogeneity of clinical populations makes creating datasets of equal size and quality more challenging. There is a need for methods that can connect these powerful large datasets with the carefully curated clinical datasets collected over the past decades.

Methods:

We applied a multivariate predictive modeling method, brain basis set (Sripada, 2019) in the Human Connectome Project resting-state fMRI data (N=810). Brain basis set takes advantage of the fact that, though functional connectomes are large and complex, there is massive redundancy in the set of connections that differ across people. This allows a small set of components to capture the most meaningful inter-individual variations and it provide generalizability allowing us to use the basis set developed in one group (HCP) for prediction in a separate clinical dataset. Our clinical dataset consists of resting-state fMRI data collected from children who stutter (N=58) and healthy controls (N=57) in the 3-10 age range. The data was collected as part of an on-going longitudinal study in stuttering (Chow, 2017, Garrett, 2018). Preprocessing and connectome generation methods were consistent across datasets. The prediction task is classification of children who stutter versus controls. Our predictive model uses 10-fold logistic regression and controls for nuisance variables (age, sex, linear and quadratic effects of motion). We selected HCP basis set components that were significantly correlated ($p<0.01$) with a phenotype of interest, sustained attention, and used these as the features in our model. The choice of sustained attention among the phenotypic measures collected in HCP was based on an earlier connectomics investigation (Chang, 2018) conducted on a subset of the stuttering dataset reported here, where significant differences involving attention networks (DAN, VAN) and their connectivity with fronto-parietal and somatomotor networks were found to differentiate CWS from controls. Clinically, ADHD and subclinical attention deficits are commonly reported in stuttering (Donaher and Richels, 2012),

but to date there have been few studies investigating how neural networks supporting attention are affected in stuttering. All code used is available at https://https://github.com/saigerutherford/OHBM2020-stuttering

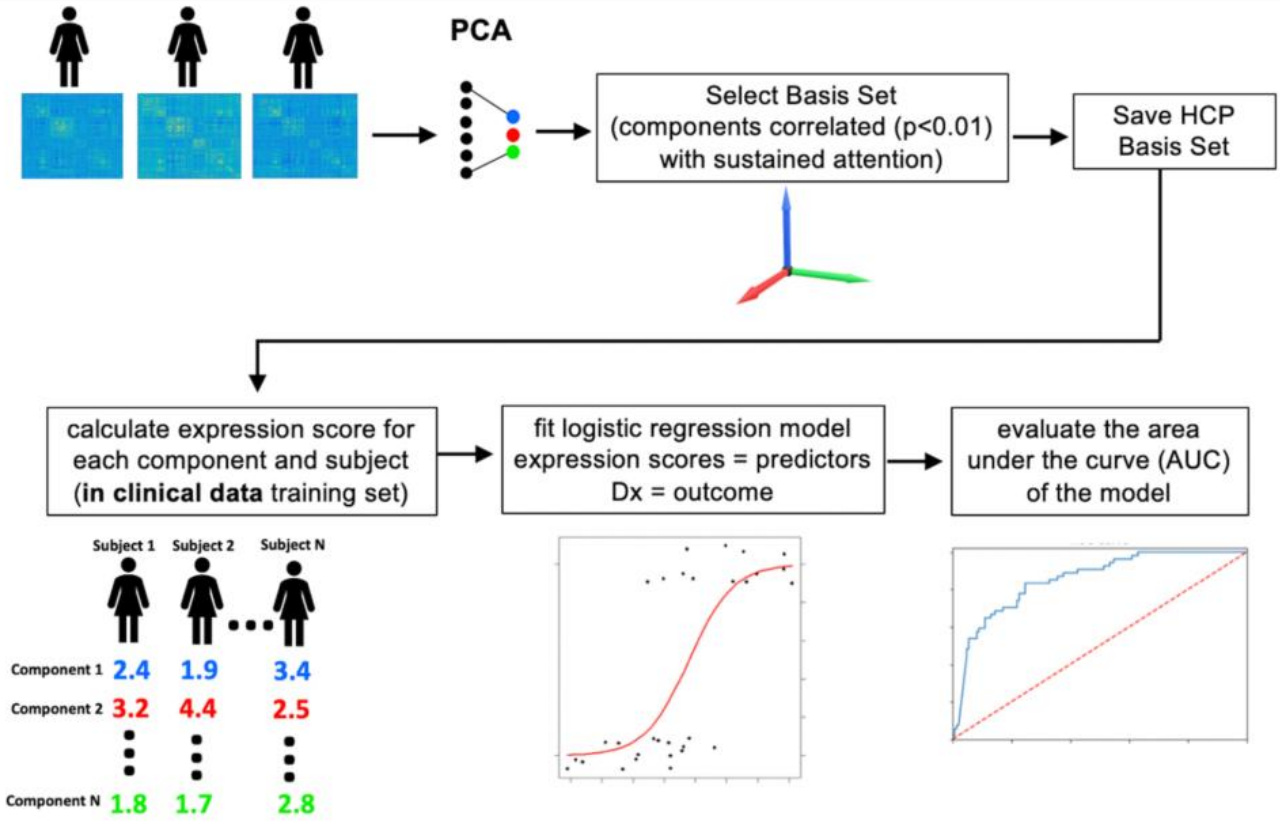

Figure 1: *Overview of Brain Basis Set (BBS) Modeling. BBS utilizes dimensionality reduction with principal components analysis (PCA) to construct a high quality feature set in a large data set (i.e. HCP), and then uses this basis set in out-of-sample data to make predictions.*

Results:

There were 12 components in the HCP dataset that were significantly correlated (p<0.01) with sustained attention, measured by the Short Penn Continuous Performance Test. As shown in Figure 2, the components were densely populated in cells representing within-attention-network connectivity, frontoparietal-attention network connectivity, and default-attention network connectivity. We were able to classify children who stutter based on the brain basis set features with 66% accuracy (AUC=0.6612). Permutation testing was run using 10,000 iterations to test the significance of our observed AUC and the resulting p-value was 0.0013.

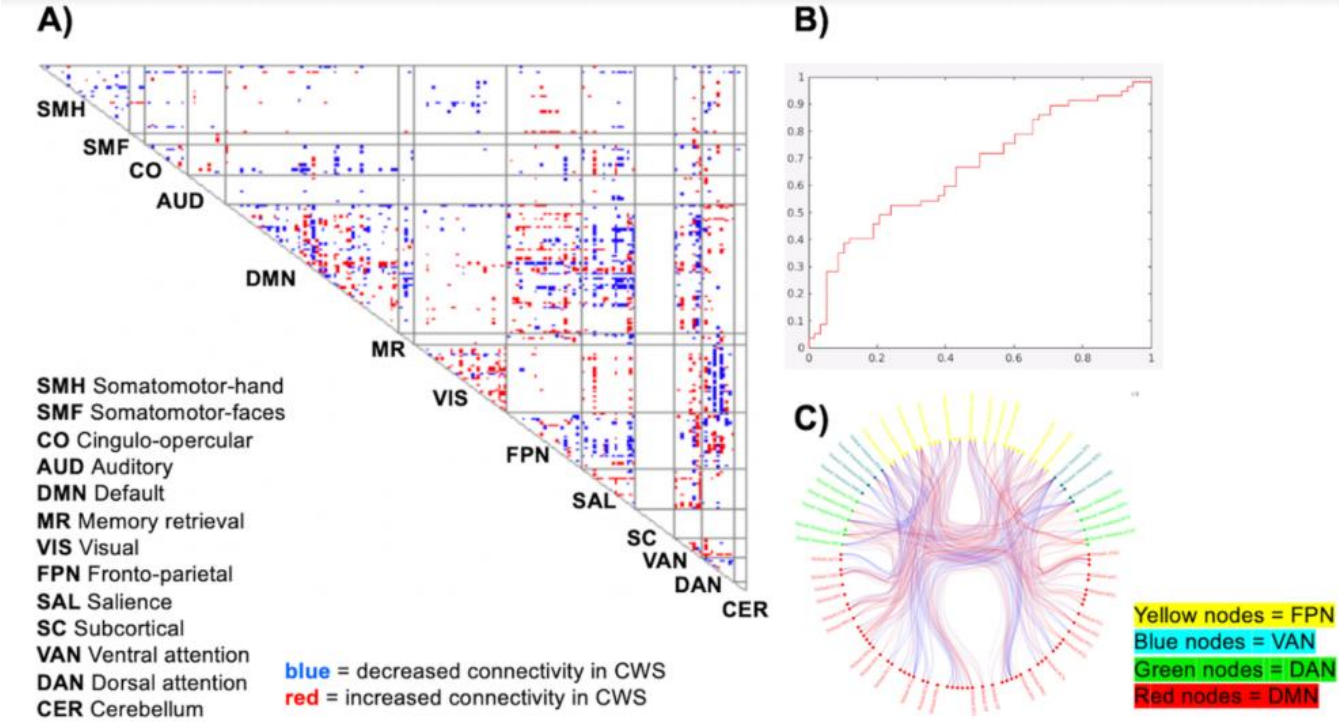

Figure 2: *Results from our classification of children who stutter (CWS) versus controls.* ***A)*** *The most highly correlated component with sustained attention, the expression of this component is increased in children who stutter compared to healthy controls.* ***B)*** *The area under the curve plot of our logistic regression model.* ***C)*** *Highlighting between network pairs that strongly differentiate between CWS and controls.*

Conclusions:

In this preliminary work, we tested whether a brain basis set developed in HCP could help classify stuttering children from fluent peers. Stuttering is a complex neurodevelopmental disorder, where symptom onset occurs in early childhood and persists in 1% of the general population. Identifying the brain basis features predictive of stuttering can offer insight into the neurophysiological basis of stuttering and hopefully provide a mechanism for ultimately predicting the optimal treatment strategy and/or outcomes. Future work is needed to extend this framework to other clinical populations.

Disorders of the Nervous System:

Neurodevelopmental/ Early Life (eg. ADHD, autism)

Language:

Speech Production [2]

Modeling and Analysis Methods:

Classification and Predictive Modeling [1]
Methods Development
Multivariate Approaches

Keywords:

Development
FUNCTIONAL MRI
Machine Learning
Pediatric Disorders
Other - Stuttering

[1|2]Indicates the priority used for review

My abstract is being submitted as a Software Demonstration.

No

Please indicate below if your study was a "resting state" or "task-activation" study.

Resting state

Healthy subjects only or patients (note that patient studies may also involve healthy subjects):

Patients

Are you Internal Review Board (IRB) certified? Please note: Failure to have IRB, if applicable will lead to automatic rejection of abstract.

Yes

Was any human subjects research approved by the relevant Institutional Review Board or ethics panel? NOTE: Any human subjects studies without IRB approval will be automatically rejected.

Yes

Was any animal research approved by the relevant IACUC or other animal research panel? NOTE: Any animal studies without IACUC approval will be automatically rejected.

Not applicable

Please indicate which methods were used in your research:

Functional MRI
Computational modeling

For human MRI, what field strength scanner do you use?

3.0T

Which processing packages did you use for your study?

SPM

Provide references using author date format

Sripada C, Angstadt M, Rutherford S, Kessler D, Kim Y, Yee M, et al. Basic units of inter-individual variation in resting state connectomes. Sci Rep. 2019;9:1900.
Chow, H. M., & Chang, S. E. (2017). White matter developmental trajectories associated with persistence and recovery of childhood stuttering. Human Brain Mapping, 38(7), 3345–3359.
Garnett, E. O., Chow, H. M., Nieto-Castañón, A., Tourville, J. A., Guenther, F. H., & Chang, S.-E. (2018). Anomalous morphology in left hemisphere motor and premotor cortex of children who stutter. Brain, 141, 2670–2684.
Chang SE, Angstadt M, Chow HM, Etchell AC, Garnett EO, Choo AL, and others. 2017. Anomalous network architec- ture of the resting brain in children who stutter. J Fluency Disord 55:46–67.
Donaher, J., & Richels, C. (2012). Traits of attention deficit/hyperactivity disorder in school-age children who stutter. Journal of Fluency Disorders, 37(4), 242–252.