Abstract #14 | Poster

Not one model fits all: unfairness in RSFC-based prediction of behavioral data in African American

J. Li[1, 3], D. Bzdok[2], A. Holmes[4], T. Yeo[3], S. Genon[1]

[1] *Forschungszentrum Julich, Institute of Neuroscience and Medicine, Jülich, Germany*
[2] *McGill University, Department of Biomedical Imaging, Montreal, Canada*
[3] *National University of Singapore, ECE, CSC, CIRC, N.1 & MNP, Singapore, Singapore*
[4] *Yale University, New Haven, United States of America*

While predictive models are expected to play a major role in personalized medicine approaches in the future, biases towards specific population groups have been evidenced, hence raising concerns about the risks of unfairness of machine learning algorithms. As great hopes and intense work have been invested recently in the prediction of behavioral phenotypes based on brain resting-state functional connectivity (RSFC), we here examined potential differences in RSFC-based predictive models of behavioral data between African American (AA) and White American (WA) samples matched for the main demographic, anthropometric, behavioral and in-scanner motion variables.

We used resting-fMRI data with 58 behavioral measures of 953 subjects comprising 130 African American (AA) and 724 White American (WA). For each subject, a 419 x 419 matrix summarizing connectivity of 419 brain regions was computed.

Matching between AA and WA was performed at the subject level by creating 102 pairs of AA and WA subjects, matched for 6 types of variables (age, sex, intracranial volume, education, in-scanner motion and behavioral scores). We performed 10-fold nested cross-validation by randomly splitting the 102 pairs across 10 sets. The remaining 749 subjects were also divided across the 10 sets. A predictive model was built for each behavioral variable by using kernel ridge regression.

All analyses focused on the 102 matched AA and WA groups. After FDR correction ($q < 0.05$), no significant difference was found between the matched AA and WA groups for the matching variables.

Out of 58 behavioral variables, 38 showed significantly above chance prediction accuracies (based on permutation test, FDR corrected). Overall, average prediction performance for these variables was higher in the WA group than in the AA group. Furthermore, significant differences in prediction performance between the two groups were found in 35 behavioral variables (FDR corrected; $q < 0.05$).

Our results suggest that RSFC-based prediction models of behavioral phenotype trained on the entire HCP population show different prediction performance in different subsets of the population. This suggest that one model might not fit all that, in some cases, RSFC-based predictive models might have poorer prediction accuracies for African Americans compared to matched White Americans. Future work should evaluate the factors contributing to these discrepancies and the potential consequences, as well as possible recommendations.

Keywords: Brain,MRI,Behavior,Prediction,Fairness