

Reliability and predictability of phenotype information from functional connectivity in large imaging datasets

Jessica Dafflon^{1,2}, Dustin Moraczewski¹, Eric Earl¹, Dylan M. Nielson², Gabriel Loewinger², Patrick McClure³, Adam G. Thomas¹, and Francisco Pereira²

¹ Data Science & Sharing Team, National Institute of Mental Health, Bethesda, MD, USA ² Machine Learning Team, National Institute of Mental Health, Bethesda, MD, USA

³ Naval Postgraduate School, Monterey, CA, USA

Abstract. One of the central objectives of contemporary neuroimaging research is to create predictive models that can disentangle the connection between patterns of functional connectivity across the entire brain and various behavioral traits. Previous studies have shown that models trained to predict behavioral features from the individual's functional connectivity have modest to poor performance. In this study, we trained models that predict observable individual traits (phenotypes) and their corresponding singular value decomposition (SVD) representations – herein referred to as latent phenotypes from resting state functional connectivity. For this task, we predicted phenotypes in two large neuroimaging datasets: the Human Connectome Project (HCP) and the Philadelphia Neurodevelopmental Cohort (PNC). We illustrate the importance of regressing out confounds, which could significantly influence phenotype prediction. Our findings reveal that both phenotypes and their corresponding latent phenotypes yield similar predictive performance. Interestingly, only the first five latent phenotypes were reliably identified, and using just these reliable phenotypes for predicting phenotypes yielded a similar performance to using all latent phenotypes. This suggests that the predictable information is present in the first latent phenotypes, allowing the remainder to be filtered out without any harm in performance. This study sheds light on the intricate relationship between functional connectivity and the predictability and reliability of phenotypic information, with potential implications for enhancing predictive modeling in the realm of neuroimaging research.