



Application number/Title: 30805 Method development and analysis of biobank imaging, behavioral and genetic data

Applicant PI: Dr Satrajit Ghosh

Application Institution: Massachusetts Institute of Technology, Cambridge, United States

Keywords provided by the Applicant PI to describe the research project:

genetics, imaging, machine-learning, precision-medicine

Application Lay Summary:

The overarching principle of our proposed research is to consider each individual as a source of very high dimensional data (behavioral, clinical, genetics, and imaging). We intend to understand associations between these datatypes and develop novel algorithms and prediction models. We hope that such an approach will lead to better characterization of individuals and their relations to one other. Such models may help relate life and/or disease trajectories across individuals leading to better precision in treatment. Our methods will open source and made available to the wider community. The proposed research meets the Biobank's stated purpose in three ways. First, the development of biotypes provides an alternative stratification of individuals different from classical diagnoses. Second, the submission of processed derivatives to the Biobank enriches the data and reduces redundant effort. Third, the development and dissemination of novel algorithms enhances the scientific enterprise. The research is primarily computational in nature, and will take place over the coming years. We will be developing novel workflows and scalable compute infrastructures to handle the increasing amount of data in the biobank. We will first identify any processed data to ensure we do not perform redundant computation. Then we will process data using existing methods and make available the analysis tools using technologies that allow others to reproduce the analysis. At this point, we will start investigating the data using novel algorithms and visualization methods. Subject to any local space and computational constraints, the full cohort will be used. Our initial work will focus on a subset of data with neuroimaging and/or genetic information.