

# Multimodal Neuroimaging Data Integration and Pathway Analysis

Yi Zhao<sup>1</sup>, Lexin Li<sup>2</sup>, and Brian Caffo<sup>3</sup>

<sup>1</sup>Department of Biostatistics, Indiana University School of Medicine

<sup>2</sup>Department of Biostatistics and Epidemiology, University of California, Berkeley

<sup>3</sup>Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

## Abstract

With fast advancements in technologies, the collection of multiple types of measurements on a common set of subjects is becoming routine in science. Some notable examples include multimodal neuroimaging studies for the simultaneous investigation of brain structure and function, and multi-omics studies for combining genetic and genomic information. Integrative analysis of multimodal data allows scientists to interrogate new mechanistic questions. However, the data collection and generation of integrative hypotheses is outpacing available methodology for joint analysis of multimodal measurements. In this article, we study high-dimensional multimodal data integration in the context of mediation analysis. We aim to understand the roles different data modalities play as possible mediators in the pathway between an exposure variable and an outcome. We propose a mediation model framework with two data types serving as separate sets of mediators, and develop a penalized optimization approach for parameter estimation. We study both the theoretical properties of the estimator through an asymptotic analysis, and its finite-sample performance through simulations. We illustrate our method with a multimodal brain pathway analysis having both structural and functional connectivities as mediators in the association between sex and language processing.