Addressing the Replicability and Generalizability of Clinical Prediction Models

Naim Rashid, PhD
Associate Professor
Department of Biostatistics
University of North Carolina at Chapel Hill

12:00 Noon Eastern time, Tuesday, September 7, 2021

ABSTRACT
The replicability of statistical algorithms for clinical decision-making has been of significant concern in biomedical research, where multiple factors may limit the generalizability of models trained on individual studies. In this talk we illustrate the impact of such factors on existing strategies for training prediction models, motivated by recent problems in the cancer genomics. To improve replicability, we introduce a novel high dimensional data integration and multi-study learning approach, pGLMM, for supervised learning problems. We show that our method improves prediction accuracy in new studies, while also selecting features whose effects are consistently non-zero across training studies. We introduce an R package implementation of our method and present some current work enhancing the scalability of this framework, while also accounting for arbitrary non-linear interactions between features. Applications to subtype prediction in pancreatic cancer will be discussed.