A New Artificial Intelligence Tool for Substance Use and Dependence Research

Qing Lu

Department of Epidemiology and Biostatistics, College of Human Medicine, Michigan State University, East Lansing, MI

With rapidly evolving high-throughput technologies and ever-decreasing costs, genetic, behavior, and other types of data have been collected in large-scale studies for substance use and dependence (SUD) research. While fully utilizing data from these large-scale cohorts holds great promise for novel SUD findings, the massive amount of existing SUD data (i.e., millions of genetic variants and millions of samples) pose tremendously computational challenges to data analysis. Furthermore, SUD are complex processes likely influenced by many genetic variants in a sophisticated way. Most existing methods have certain assumptions (e.g., the linear and additive assumptions) on the relationship between SUD and genetic variants, and are not suitable to model their complex relationship, which adding another layer of difficulty to the discovery process. In this talk, I will introduce a new artificial intelligence (AI) tool, Kernel Deep Neural Network (KDNN), for SUD research. The KDNN is developed based on the idea of the most popularly AI technology, deep learning, which has been widely used in high-tech industries (e.g., self-driving cars and robotics). Despite its successes in many areas, deep learning has been rarely used in genetics and population health research. The proposed KDNN method addresses the limitations of existing methods (e.g., linear model and linear mixed model) for capturing the complex relationship between genetic variants and SUD phenotypes, and solves barriers of using deep learning for analyzing high-dimensional genetic data. Through rigorous statistical proof and preliminary simulations, we show that the new method attains better performance than the conventional linear mixed model, especially when there is a complex relationship between genotype and phenotype. In this talk, I will also briefly discuss how to improve its computational efficiency for large-scale data analysis and its future extension to meta-analysis of data from multiple sites.

Supported by: NIDA K01DA033346 and NIDA R01DA043501