Submitter Name: Shuo Chen
Submitted email: shuochen@som.umaryland.edu

Deep learning model using network topology of linkage disequilibrium patterns increases the accuracy of polygenic risk scores for smoking

Shuo Chen[1,2], Charles Ma[3], Peter Kochunov[2], and Elliot Hong[2]

[1]Division of Biostatistics and Bioinformatics, University of Maryland, School of Medicine; [2]Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland, School of Medicine; [3]Department of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park

Complex traits are often associated with a set of correlated single-nucleotide polymorphisms (SNPs) in a genomic region. The linkage disequilibrium (LD) patterns between these multivariate SNPs are in organized (yet sometimes latent) network/graph topological structures (e.g. interconnected communities). We develop novel shrinkage algorithms to identify the latent graph topological structures of LD patterns of SNPs in a genomic region. Next, we propose a new deep learning model using graph convolutional networks where the graph structure is determined by the detected LD network topology. The phenotype-related SNPs (based on statistical screening) in multiple genomic regions are considered as input of the deep learning model, while polygenic risk scores (PRS) are the outputs. Therefore, the deep learning model integrates the LD patterns and effect sizes of multiple SNPs in genomic regions, and better explain the phenotypic variance by fully leveraging the interactive and nonlinear relationships between correlated SNPs. We apply this approach to GWAS data for nicotine addiction. We train the deep learning model based on a training data set of 600 subjects and validate the method using an independent testing data set of 500 subjects. The results show that the area under the curve (AUC) of receiver operating characteristic (ROC) curve based on smoking status PRSs using the proposed deep learning model is significantly higher than competing methods.