MACS 30200
Perspectives on Computational Research
Research Proposal
Jiaxu Han

Identify Risk Factors for Early-onset Depression: A Machine Learning Approach

Background

There has been a heightened awareness of adolescent depression among both the professional and the general media (Costello, Erkanli, & Angold, 2006). Early-onset depression tends to continue to adulthood and comorbid with other psychiatric disorders (Birmaher et al., 1996). Adolescent depression itself is a major risk factor that surrogates multiple influences on one's developmental trajectory and often leads to a poor academic outcome, substance abuse, suicide, and so on (Birmaher et al., 1996). Therefore, understanding what might be predictive of early-onset depression is critical for early detection and developing intervention and treatment plan. "It is now widely understood that singular, linear cause will rarely obtain" (Cicchetti & Sroufe, 2000). A tenet of developmental psychopathology is to acknowledge the complexity of causality. Instead of trying to identify a linear causal relationship, it emphasizes on understanding the interaction of risk and protective factors on mental health outcomes. However, it is challenging for classic inferential paradigms such as group comparison or correlation study to capture the complexity of human developmental trajectories. In addition, those inferential approach has raised replication and reproducibility issues (Dwyer, Falkai, & Koutsouleris, 2018). In the past, due to limited resource, a single research study can only examine so many variables. The classic inferential approach is well suited for this kind of research, but the results do not apply to each individual. In recent years, there has been increased collaboration between multiple sites across

the country. With collaborative efforts, the research study is now able to gather much more information about each participant and build a large dataset with thousands of participants. Machine learning provides an alternative approach to analyze those data with its own advantages such as enhanced predictability.

In light of the information above, the research question of the proposed study is: what are the risk factors that are most predictive of early-onset depression among adolescents?

Data

The proposed study will use the data from the Adolescent Brain Cognitive Development (ABCD) study. ABCD study is NIH funded and consists of 21 research sites across the country. So far, there have been 11,878 children ages 9-10 join the study. The study conducts physical, cognitive, social, emotional, environmental, behavioral, and academic assessments, as well as multimodal neuroimaging and biospecimen collection for hormonal, genetic, epigenetic, environmental exposure, and substance use analysis ("Protocols", n.d.). The demographics, ethnicity and socioeconomic status of the children that participated in the ABCD study have been very diverse (see figure 1, figure 2 and figure 3). There are many variables collected in the ABCD study and it requires some exploratory analysis to decide what variables to include in the final analysis. The proposed study will not analyze fMRI data.

Computational method and theory

The proposed study will use python 3 to analyze the data. There will be three major steps in data analysis: preprocessing, feature selection, and build a decision tree. In data preprocessing, we will drop the missing data, identify covariates (and combine some covariates if needed), merge columns if necessary, and so on. There are many methods for feature selection, we will do some exploratory analysis to find the one that is a better fit for the proposed study. Similarly,

there are several algorithms and associated packages for building a decision tree. We will use k-fold cross-validation and hyperparameter tuning to find the appropriate model.

The proposed study adopts a developmental psychopathology framework (Cicchetti & Sroufe, 2000) to explain the implication of the results.

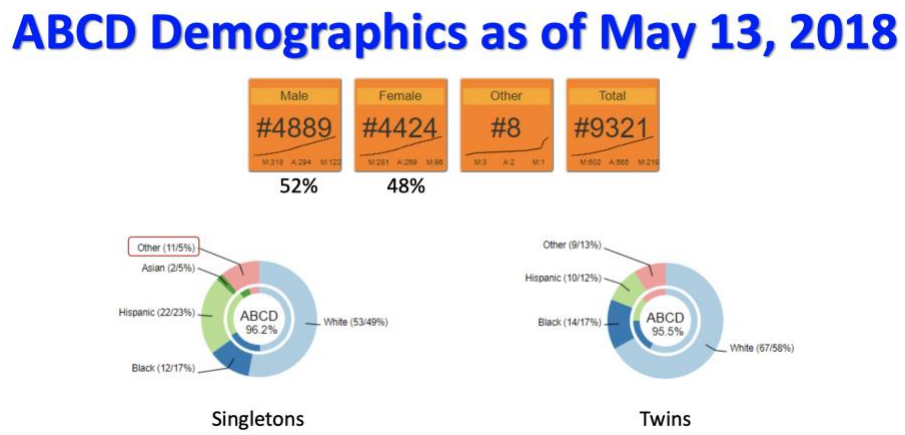

Figure 1. Demographics of participants in ABCD study as of May 13, 2018. (Retrieved from website https://deainfo.nci.nih.gov/advisory/joint/0518/Dowling.pdf)

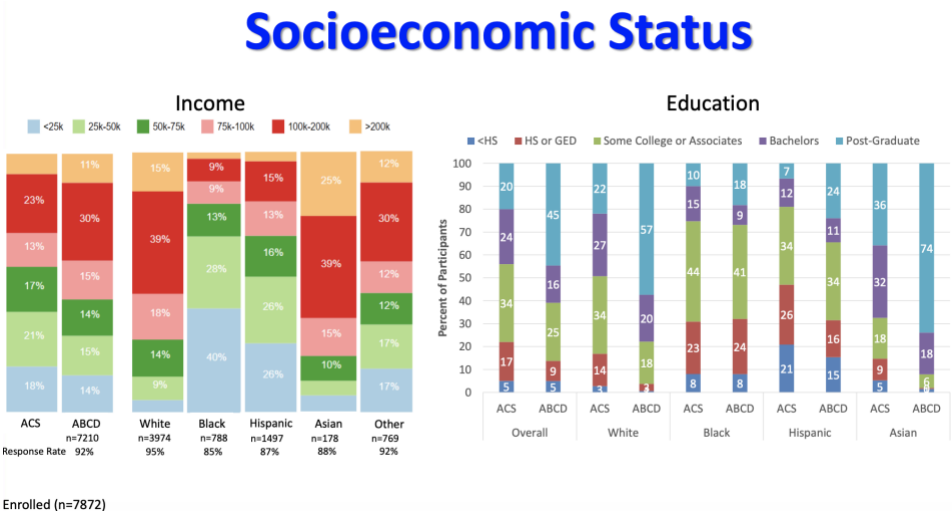

Figure 2. Socioeconomic Status of participants in ABCD study. (Retrieved from website https://deainfo.nci.nih.gov/advisory/joint/0518/Dowling.pdf)

Figure 3. Ethnic diversity of participants in ABCD study. (Retrieved from website https://deainfo.nci.nih.gov/advisory/joint/0518/Dowling.pdf)

References

Birmaher B, Ryan N.D., Williamson D.E., Brent D.A., Kaufman J, Dahl R.E., Perel J, Nelson B. (1996). Childhood and Adolescent Depression: A Review of the Past 10 Years. Part I. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(11), 1427–1439.

Cicchetti, D., & Sroufe, L. A. (2000). The past as prologue to the future: the times, they've been a-changin'. *Development and Psychopathology*, *12*(3), 255–264.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, *14*(1), 91–118.

Jane Costello, E., Erkanli, A., & Angold, A. (2006). Is there an epidemic of child or adolescent depression? *Journal of Child Psychology and Psychiatry*, *0*(0), 1263-71.