Hi workshop members,

My writing sample below is the last portion of my Prospectus introduction. In the Prospectus document, I proposed to use the ABCD dataset to build a deep learning model distinguishing children with and without family history (FH) of Alcohol Use Disorder (AUD). In previous sections, I’ve talked about what is known about anatomical and connectivity alterations in family history positive (FHP) subjects, as compared to family history negative (FHN) subjects. The writing sample below is my rationale why do I choose deep learning over traditional regression method. I have cut out many details such as description of/steps involved in deep learning, as I will mention them in the Method section. Please let me know if I need to further explain anything and if my justification is lacking

I have grayed out the portion that exceeds the word limit, which I included because I think it provides some context, i.e., how I think my study can address the unknown in the literature. I would appreciate your comments on this portion, but you don’t have to.

Thank you!

**3.2. Strengths of the Current Study**

In the extant literature, no quantitative meta-analysis has been run to find consistent findings of both brain morphometry and connectivity alterations across studies, and a qualitative review paper (McPhee et al 2018) pointed out inconsistent findings on brain morphometry changes, even for the most robust effect (reduced amygdalar volume). In addition, previous studies on both brain morphometry and resting-state functional connectivity had relatively small sample size (less than 100 FHP, except for Dager et al 2015). In many studies, FHP had substance use history, making it difficult to tease apart neural alterations associated FH status and personal substance use.

The Adolescent Brain Cognitive Development study provides a rich dataset from more than 11,000 children age 9-10 with minimal substance use. Using deep learning to analyze data from substance naïve FHN and FHP children from this dataset can elucidate changes in brain structure and connectivity associated with FH of AUD in a large group of participants.

**3.2.1 Strengths of machine learning**

Traditional whole brain analysis for both structural and connectivity data typically involves mass-univariate analysis, which has significant limitations (Vieira et al 2017). Firstly, mass-univariate analysis involves multiple independent comparisons (e.g., independently comparing amygdalar and hippocampal volumes between FHP and FHN). This results in stringent correction for multiple comparisons, which can increase Type II error. In addition, multiple comparisons assume that each brain regions are independent, which is not accurate. Vieira et al (2017) also noted that mass-univariate techniques can have statistical inferences at the group level, but not at the level of the individual. Diagnostic and treatment decisions, on the other hand, are made at the individual level.

These limitations can be addressed with machine learning, which is a multivariate method to detect trends and patterns in existing data, which can later be used to make predictions on new data. Machine learning also allows for prediction to be made at the individual level.

The gap in the literature that this project hopes to address is the inconsistent findings in brain structure and functional connectivity in FH of AUD. Using a large sample from the ABCD database is one of the strategies to achieve reliable results. Using machine learning is the additional strategy to improve reliability, as machine learning can extract replicable brain patterns that distinguish FHP from FHN. In particular, machine learning algorithms can pick up the overall patterns of a dataset while ignoring noisy, idiosyncratic characteristics that are specific to that dataset, increasing generalizability and replicability of the results.

**3.2.1 Strengths of deep learning**

One criticism of conventional machine learning techniques is not performing well on raw data, especially high dimensional brain imaging data, necessitating the “feature selection” step, which involves discarding redundant information to focus solely on informative features. Feature selection is oftentimes subjective and requires extensive prior knowledge deducing what features should be considered informative.

Deep learning, on the other hand, is more objective as it can learn features from raw data, bypassing the “feature selection” step. Inconsistent findings from previous studies on FH of AUD demonstrated the subtle and diffuse structural and connectivity abnormalities associated with family history status. Deep learning can outperform traditional machine learning in datasets with such subtle and diffuse patterns (LeCun et al., 2015). In addition, unlike univariate methods, deep learning automatically takes into account the inter-correlation between features and down-weight unmeaningful interactions.

Indeed, several studies have demonstrated deep learning’s superior performance relative to other machine learning methods (Koyamada et al. 2015; Plis et al. 2014). For instance, using Deep Neural Networks to classify brain states associated 7 different tasks (Emotion, Gambling, Language, Motor, Relational, Social and Working Memory) from 499 subjects, Koyamada et al. (2015) found better performance for a Deep Neural Networks model (mean accuracy of 50.74%) compared to supervised learning methods (mean accuracy of 47.97%) such as Linear Regression and Support Vector Machine. In a different study, using a Deep Belief Network model to classify T1-weighted structural data from 198 schizophrenic patients and 191 controls, Plis et al. (2014) achieved 90% classification accuracy with deep learning, while a Support Vector Machine model only achieved 68% classification accuracy.