Excessive alcohol drinking impacts the structural development of brain in adolescents, but its impact on the functional activity or connectivity of the brain has not yet been explored.

Our goal is to design a classification model to predict if a subject is a heavy drinker based on their resting-state fMRI data (stored as blood oxygen-level dependent (BOLD) signals). We used logistic regression of pre-processed data as a baseline for CNN/RNN-based models and SVMs.

Surprisingly, we found that using derived features with logistic regression yielded far better results than applying the simple, processed data to complex models.

Dataset
- Source: National Consortium on Alcohol and Neurodevelopment in Adolescence (NCANDA) database
- fMRI scans of \( m = 715 \) adolescents and young adults (16-19 y/o), measured as BOLD signals from each voxel every between each \( T = 269 \) timesteps (2.2 seconds / timestep)
- Dataset was imbalanced (122 (17%) heavy drinkers out of 715)

Pre-processing
- Parcellate brain into regions (\( N \)) to reduce noise
- Brain activity was normalized to z-score
- Downscaling of majority class (size(1) == size(0))

Raw Features (\( m \times T \times N \))
- \( m = 715 \) subjects / 244 after downscaling
- \( T = 269 \) timesteps
- \( N = \) Variable (brain regions)

Derived Features (\( m \times N \))
- Dynamic range per brain region in ICA
  - \( \Delta_{\text{derived}}(N) = \max(x_{\text{derived}}(N)) - \min(x_{\text{derived}}(N)) \)
- Demographics
  - sex, age, scanner type

Results

- High variability of the deep-learning model makes adjustments of hyperparameters difficult
- Risk of overfitting deep-learning models and SVMs is high
- Small dynamic range of prediction values in deep-learning models suggests low sensitivity
- Many instances of ‘nan’ or bias only toward one class
- Overall suggests that our current amount of data may be insufficient to train deep-learning models
- Fairly good results from logistic regression alone when using derived features including demographics
- Removing of age as a feature decreases performance of logistic regression. Highlights the influence of demographic information toward making correct predictions

References