Challenges

1. Improve model performance on sleep stage prediction
    a. The best results are achieved using more complex feature construction, which requires significant time to load the data and train the model.
    b. Neural network does not outperform simple model like random forest
        i. Grid search on hyperparameter did not improve the model.
        ii. Difficult to design the right architecture for various neural network.
    c. Find a similarity factor of sleep data of a single patient relative to cluster aggregations. This was key to be able to make a logical prediction of CVD risk prediction.
    d. Subject clustering required several stages of data processing to eventually return CVD risk predictions.

Group contribution

	Dylan Kapp	Blake Macnair	Rong Cao	Daniel Cazzaniga
% contribution	25%	25%	25%	25%

Areas of special focus:

Dylan:
- Determine subject similarity output using scala
- Python helper scripts to work shhs data into a good format for scala
- Video editing

Blake:
- Raw EDF file processing
- Python sleep stage annotations
- Power Spectral Clustering

Rong:
- Neural Network training models
- Random forest training models
- Comparison findings of sleep stage annotation performance

Daniel:
- Python helper scripts for spectral clustering
- Subject cluster assignment
- Research into tensorization approaches