**Predicting Atrial Fibrillation in Embolic Stroke of Undetermined Sources**

**Hypothesis**

Machine learning method, especially tree-based models, is an effective strategy for AF screening among ESUS patients.

**Data**

Our patient cohort is provided by Dr. Nahab, who kept a record of his ESUS patients. To confirm these patients, we used international disease classification (ICD) to ensure individual ESUS patient had a history of ischemic stroke in electronic health records. Our exclusion criterion is any ESUS patients who had AF on or before index ESUS stroke.

We then collected demographic data (age, gender), physical characteristics (height, weight), lab records (cholesterol, hemoglobin, glucose, red and white blood cell count, etc.), radiographic reports (electrocardiogram and echocardiogram), and other common risk factors (smoking and obesity). These data were collected closest to the index ESUS stroke date in order to reflect the most appropriate condition of the ESUS patient for AF prediction. In addition, we collected medical comorbidities history and prescribed medicines history for each ESUS patient.

**Methodology**

First, we external validate existing AF risk scores (HAVOC, Brown ESUS-AF, CHA2DS2-VASc, C2HEST).

Second, we develop several machine learning tree-based classifier, including random forest classifier, extremely randomized trees classifier, gradient boosting classifier, and histogram gradient boosting classifier.

Third, we plan to incorporate an embedding vector from a pretrained Hypergraph Neural Networks into our predictor space, which inputs dummy variables from medical comorbidities history and prescribed medicines history.

**Expected Results**

We expect that our machine learning risk score to outperform existing risk scores such as HAVOC, Brown ESUS-AF, CHA2DS2-VASc, and C2HEST due to richer data and more complex modeling method.