Project Proposal

Computational Medicine, Fall 2021

Team Members: **Yuxuan Wu**   *Andrew ID:* **yuxuanwu**
Xin Wang   *Andrew ID:* **xinwang3**

Introduction: Emergency department (ED) visit is one of the most common ways to get medical support and EDs represents the largest source of hospital admissions. In order to improve the running efficiency of ED, optimize the resource allocation, as well as to maximize the number of patients that get appropriate treatment, we train machine learning models to predict hospital admission at the time of ED triage.

Data Source: The Electronic Health Record (EHR) data we use was from a paper published on PLOS one (https://doi.org/10.1371/journal.pone.0201016). The original retrospective data was obtained from three Emergence Department from March 2013 to July 2017, each ensuring 1 year of historical timeframe. We obtained the data from the author's Github Repository (https://github.com/yaleemmlc/admissionprediction), which had already been cleaned. And the data includes 560,486 patient visits with 972 variables.

Methods: First of all, we will do exploratory data analysis (EDA) to inspect the data and find its characteristics. Then we will perform feature selection on the training and validation data with cross validation, and we may try PCA as well. After that, we will build several machine learning models from different categories, and adjust the parameters in the models based on their performance. Our choice for now: Logistic Regression, Random Forest, SVM, XGBoost, Gradient boost Tree and deep learning model CNN.

Model evaluation: For the evaluation part, we will measure confusion matrix, AUC and F1-score.

Discussion:
1. Because of the limitation of our laptops in CPU and GPU and the large scale of the data, we may not be able to train a lot of models.
   Strategy: We may use Kaggle to train our model, which offers free GPU resources.
2. The practical value of this project might be limited because the models we will train can hardly be used for later projects without encapsulation.
   Strategy: if time is permitted, we would develop a web interface like a shiny application in R, which takes patient's data as input and returns the predicted label.