Team Contributions:

Shanshan Jiang

In this project I mainly responsible for building up components in the pipeline of training deep neural network. My contribution can be break down to following acceptives

1. Getting medical notes extracted from upstream spark data generation step. I use NLP approaches to process the notes, including clean the notes, tokenize the notes, remove stop works from notes.

2. To start with, I built and train a baseline model using random forest algorithm. I generated bag of words features out of pre-processed notes from last step.

3. For building the deep learning algorithm. I finished following steps:

3.1. Our deep learning algorithm using ELMo feature as embedding of medical description. So, I implement an approach to extract ELMo features for our extracted medical notes using a pretrained ELMoEmbedder which is from a 3rd party python package.

3.2. Given ELMo features extracted, I prepared training, validation and testing data for the proposed deep learning algorithm and I built dataloaders for these data which will be used in our deep learning algorithm

3.3. I built a GRU based neural network and get it trained using generated ELMo embeddings.

3.4. I tune the network by comparing results of many different hyperparameter configurations. Finalized the model we are submitting by choosing the one with best performance.