

# Capstone Project Proposal

Danlei Qian  
September 19, 2020

Mechanisms of Action (MoA) prediction is an active Kaggle competition, with the goal of advancing drug development through improvements to MoA prediction algorithms. In this new process, scientists seek to identify a protein target associated with a disease and develop a molecule that can modulate that protein target. As a shorthand to describe the biological activity of a given molecule, scientists assign a label referred to as MoA.

This is a multi-label classification problem. Drugs can have multiple MoA annotations which describe binary responses from different cell types in different ways. In this project, I would like to apply several machine learning algorithms to predict multiple targets of the MoA responses of different types of human cells to various drugs, given various inputs including gene expression data and cell viability data. The final submission deadline of this competition is November 30, 2020.

And the dataset of the competition is as follows:

<https://www.kaggle.com/c/lish-moa/data>

Neural Network (NN), which appear to be performing best in this competition so far, will be mainly explored. And other developed methods such as SVM, XGBoosting will be explored as well. And the main frameworks will be used are Keras and Pytorch. And the accuracy of solutions will be evaluated on the average value of the logarithmic loss function applied to each drug-MoA annotation pair.

In addition, I'll refer to some materials regarding the implementation of different techniques. Some links are shown below. More will be added during the process of doing the project.

Split Neural Network: <https://arxiv.org/pdf/1812.00564.pdf>

LSTM: <https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21>

TabNet: <https://towardsdatascience.com/modelling-tabular-data-with-googles-tabnet-ba7315897bfb>

This project is spanning 10 weeks and the schedule is as follows:

Week 1 Data preprocessing and EDA

Week 2 Explore SVM and XGBoosting

Week 3 Build baseline NN model and explore potential methods

Week 4 Explore Split Neural Network:

Week 5 Explore TabNet

Week 6 Explore LSTM

Week 7 Explore other architectures

Week 8 Explore other architectures

Week 9 Refine the models and submit the final version

Week 10 Complete the final report/paper