

# CSGS Hackathon - Binary Bandits

Dominik Soós

David Calano

October 31, 2024

For this assignment, I watched the following videos:

## 1 Introduction

The classification of drug interactions comes in two primary varieties, Drug-Target Interaction (DTI) and Drug-Drug Interaction (DDI).

## 2 Data

One of the challenges was to get our hands on data. DrugBank is the most

### 2.1 DrugBank

The most popular dataset for Drug-Drug Interaction (DDI) has been used extensively to test the performance of multiprediction tasks. There are 1706 unique drugs in this dataset with 191,808 drug-drug pairs. first dataset we experimented with was t We were able to use Drugbank through TDC

The drugbank dataset is very imbalanced with two classes dominating the distribution. Figure ?? illustrates the class imbalance. We can observe the difference in the distribution after we performed the data augmentation.

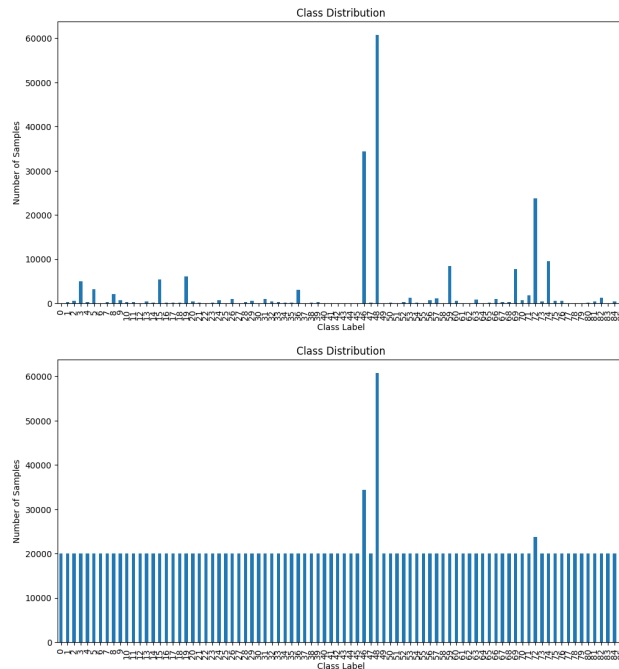


Figure 1: Drugbank dataset before (top) and after (bottom) data augmentation

## 2.2 TWOSIDES

The TWOSIDES dataset seems to be a better option as it contains more balanced classes. However, it is a much more challenging dataset containing more than 4 million DDI for only 645 drugs. For this reason, we sorted the classes and took the top  $k$  classes, where  $k = 10$ .

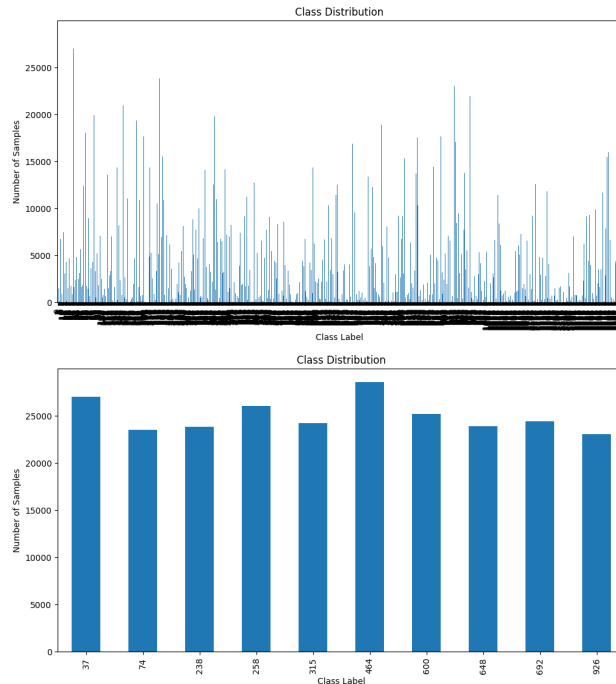


Figure 2: TWOSIDES data before (top) and after (bottom) trimming the dataset. The downsampled distribution still contains nearly 250 thousand samples.

### 3 Methodology

We implemented a graph convolutional network since the data itself is in a graph format.

#### 3.1 Model Architecture

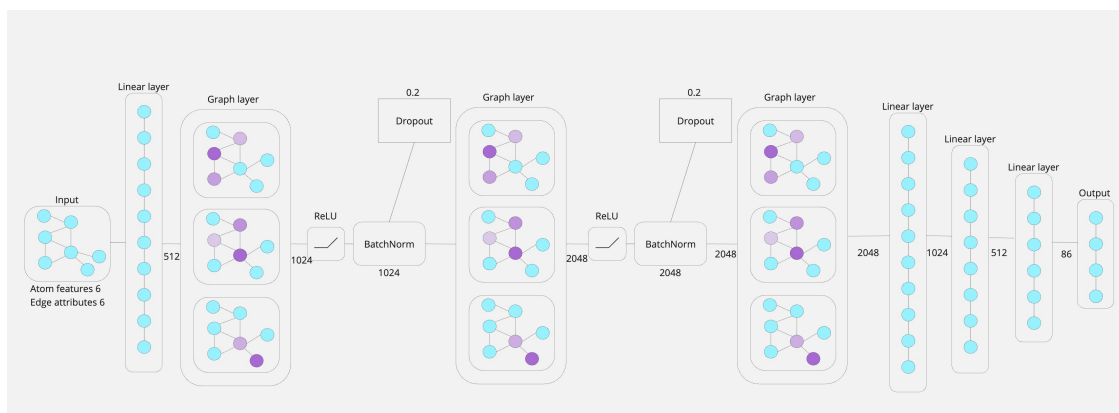


Figure 3: Model architecture

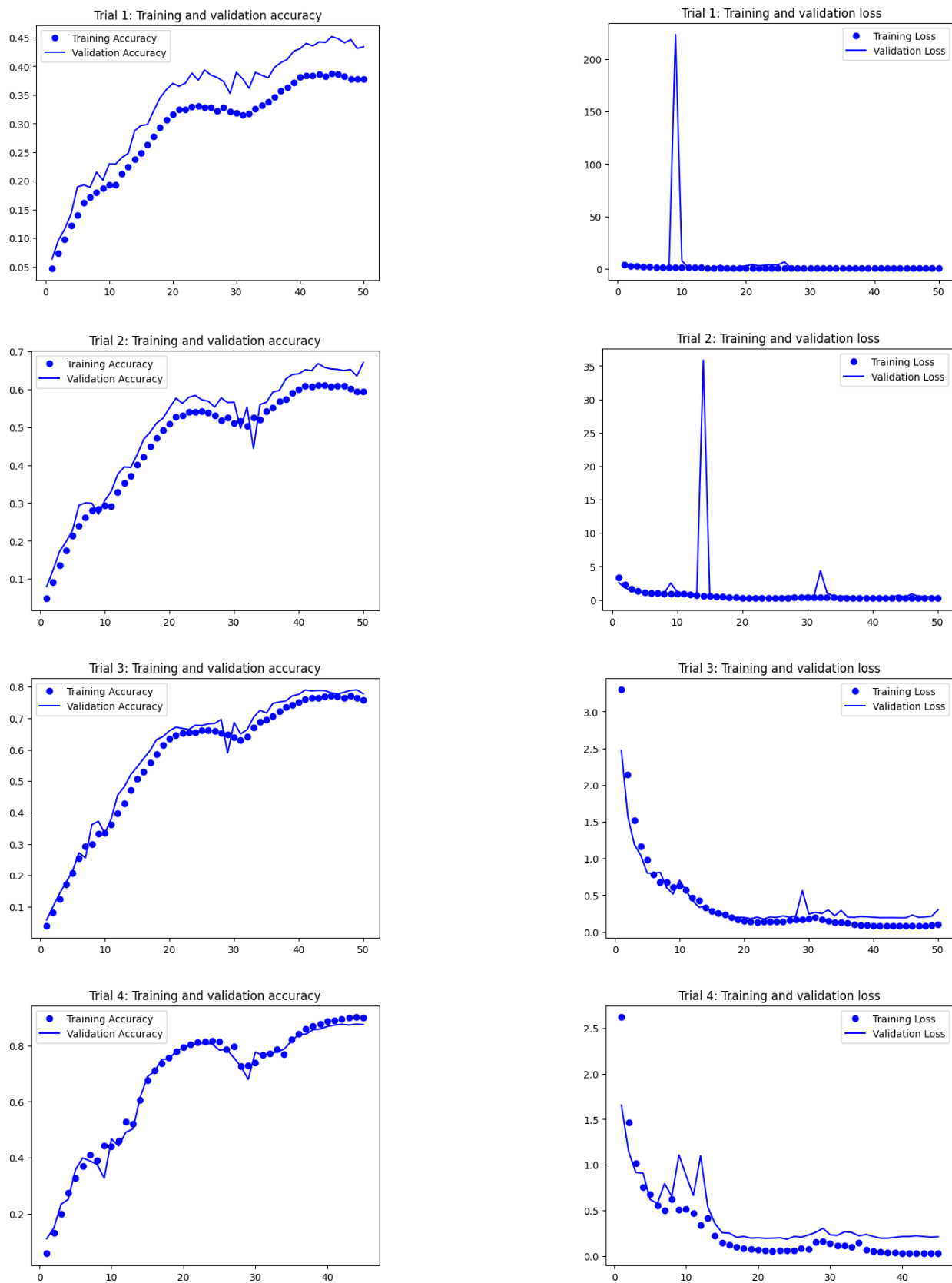


Figure 4: Training accuracy and loss data from DrugBank trials

### 3.2 Drugbank

## 4 References