For measuring the ADME properties, the following datasets have been chosen:

* HIA (Human Intestinal Absorption), Hou et al.- For **Adsorption (A)**
* BBB (Blood-Brain Barrier), Martins et al. – For **Distribution (D)**
* CYP P450 2C19 Inhibition, Veith et al. - For **Metabolism (M)**
* Half Life, Obach et al. – For **Excretion (E)**

The above-mentioned datasets are chosen to benchmark the models. The first three are datasets allow binary classification. While the Half-life is a dataset for regression. These datasets are specifically chosen as they would help to tell which model could generalize better.

**Algorithms/Models used:**

**MPNN (supervised graph neural network)**

**Transformer**

**XGBoost (Tree model, ensemble)**

**Metrics**

While benchmarking models for classification, it’s best to use ROC Curve and AUC score over accuracy. As the AUC is well-suited to measure the model’s performance on an imbalanced set and to compare the performance. Overall accuracy is based on one specific Threshold, while ROC tries all of the Threshold and plots the sensitivity and specificity. So when we compare the overall accuracy, we are comparing the accuracy based on some Threshold. The overall accuracy varies from different thresholds.

For a better understanding of how to interpret the ROC graph please see the figure below:

Diagram

Description automatically generated

Figure 1: Reference ROC curve, the larger the better

Chart, line chart

Description automatically generated



Figure 2: ROC and AUC for XGBoost for A,D,M

Chart

Description automatically generated



Figure 3: ROC and AUC curve for ADM using MPNN and Transformer

The models’ performance could be seen from the fig 2 and fig 3.

Since the first three categories consist of Classification and the last (Excretion) consists of regression. the model benchmarking is done on the first three categories.

XGBoost is chosen over MPNN as XGBoost is able to generalize and produce decent results both for classification and regression.

Here is a graph showing the AUC score for classification:

Figure 4: Comparison of all three models

In real-world discovery, the drug structures of interest evolve over time. Thus, ADME prediction requires a model to generalize to a set of unseen drugs that are structurally distant from the known drug set. Therefore, XGBoost is a model which could do this task well both for regression as well as classification.