**Challenge #1 Drug Classification - BCRP Inhibitors vs. Non-inhibitors**

Generating data at scale from lab experiments on specific questions, such as classifying drugs as BCRP inhibitors or non-inhibitors, can be costly and time-consuming. To overcome this challenge, the project aims to scrape and build relevant data models from the scientific literature to develop a predictive model

Reference Paper

The reference paper, "ADMET evaluation in drug discovery. 20. Prediction of breast cancer resistance protein inhibition through machine learning" by Dejun Jiang, et al., provides a dataset of BCRP inhibitors and substrates. The paper can be accessed at [https://jcheminf.biomedcentral.com/articles/10.1186/s13321-020-00421-y#MOESM1](https://jcheminf.biomedcentral.com/articles/10.1186/s13321-020-00421-y%23MOESM1)

Task Overview

* Performed web scrapping using Python and found BCRP inhibitors and substrates
* Collected the drugs which are BCRP inhibitors and substrates and performed ML to train the data and test the data.
* Found the correlation between reported value and classification (IC50) and did EDA.
* Created an SVM classifier with an accuracy of 97% and with accuracy of unseen data of 93%

Conclusion

An ML algorithm for scrapping drugs was developed and drugs were classified as inhibitors and substrates according to their IC50 value.