Methodology Report

- Feature space: MAP4 (r=3, 2048 bits), MACCS (167 fp) and Mordred (1658 descriptors after fixing) were used;
- Models: (1) PCA was applied independently on each feature space; the principal components from each feature space were concatenated as features for XGBoost models with 5-fold CV for hyperparameter optimization
- Models: (2) Feature selection on each component of the feature space, followed by concatenation of selected features; followed by XGBoost models / Random Forest models
- Models: (3) single-task GNN models
- Models: (4) Handcrafted feature space of ~25 descriptors used in XGBoost model.
- Both scaffold and random splits were used

Performance comments

Random splits showed marginally better performance on the internal holdout test set. GNN performed well on LogD. Ensemble was used for the other endpoints.

Team members:
Ashok Palaniappan, Ph.D
Ida Titus (Doctoral candidate)
Tamilazhagan, Masters student
Deepa Lakshmi, UG student