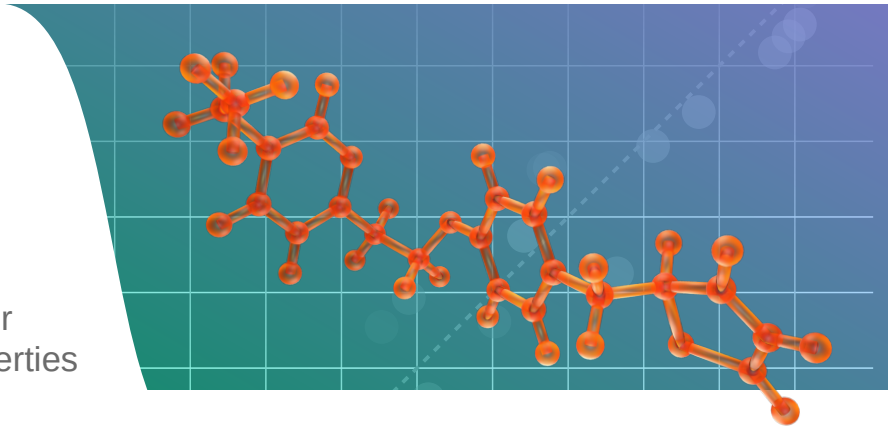
CYCLICA

Evaluating Cyclica's AI-based ADMET Prediction tool

The ADMET Prediction tool consistently outperforms traditional QSAR approaches for predicting pharmacokinetic and toxicity properties



PROBLEM

Available predictive tools lack universal standards across different pharmacokinetic and toxicity properties, leading to difficulty in implementation and reproducibility, risk of overfitting, and model bias.

TECHNOLOGY

ADMET Prediction

SOLUTION

ADMET Prediction (AP) implements a universal approach for predictive chemoinformatics, enabling model building for any dataset outlining a relationship between chemical structure and activity, which significantly outperforms traditional QSAR approaches.

INTRODUCTION

Developing a new drug is a long, expensive process that can take 10-15 years[1] and can cost upwards of $2.8 billion[2]. The 10 year success rate for new drugs is less than 10%, with a large percentage of molecules failing due to poor Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) properties. By the early 2000's, with the increased availability of curated public data, and metrics of chemical similarity like fingerprints, many researchers started building Quantitative Structure-Activity Relationship (QSAR) models for predicting ADMET properties of potential drugs. Standard practice for building QSAR models involves (1) iterating through different chemical fingerprint/descriptor sets, (2) iterating through *traditional classifiers*, i.e. supervised machine learning algorithms, such as Random Forest or Support Vector Machines, and (3) optimizing the hyper-parameters for supervised learning approach. This practice leads to different model types across different datasets, which can be very difficult to interpret, and is prone to overfitting due to the high dimensionality of the data.

Cyclica has developed a novel and proprietary QSAR method that leverages high-dimensional information from several chemical fingerprint/descriptor sets simultaneously to increase accuracy. Here we show how this method has been implemented in our ADMET Prediction (AP) tool, and that it reliably outperforms traditional classifiers for most datasets, without the need for case-by-case tweaking or performance optimization.

METHODOLOGY

We tested AP alongside five traditional classifiers to predict 17 ADMET properties (Table 1). For each of the five traditional approaches, we performed five-fold cross-validation using 80% of the dataset, leaving 20% for blind testing. This process iterated through 8 types of chemical fingerprints, and tuned hyperparameters using a grid-search strategy. For AP, we provide the 80%/20% testing split for direct comparison (*AP Test*), as well as 'leave-one-out' full cross-validation (*AP Full*) to assess predictive robustness with respect to dataset size. Cyclica's QSAR method is specifically designed to withstand the pitfalls of high-dimensional information and uses 10 different fingerprints simultaneously to build one model, compared with only one fingerprint per traditional model in the standard approach. In our tests of traditional classifiers, we built models separately for each fingerprint, selected the one performing best in cross-validation, and then measured its performance on the test set. The predictive power for all models was assessed using the area under the Receiver Operator Characteristic curve (ROC AUC). The datasets used for the 17 different models were obtained from curated, previously published QSAR studies, each focused on building an optimized model for a particular ADMET property. For example, the BBB is from Shen *et al* [3], and the HOB dataset is from Kim *et al* [4]. A full list of sources for the data sets will be provided upon request.

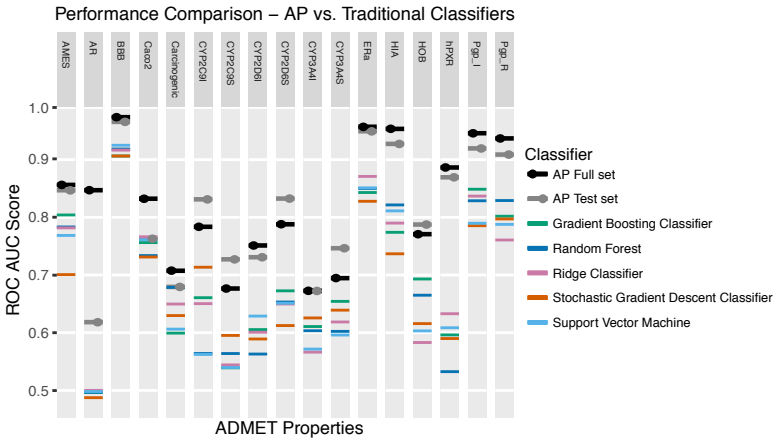


Figure 1. Comparison of AP versus the five traditional classifiers by plotting ROC AUC score for 17 ADMET Properties.

RESULTS

AP consistently outperforms the five traditional classifiers for all 17 ADMET properties, as shown in **Figure 1**. The ability of Cyclica's QSAR method to use all the data for training without the need for the 80/20% split needed with the traditional approaches, leads to higher predictive power for most properties as shown by the difference between AP Full and AP Test in **Figure 1**. Furthermore, even when using the same 80/20% data split that enables direct comparison of AP and traditional classifiers, AP Test still outperforms each of the traditional classifiers. **Figure 2** shows cross validation versus blind testing ROC AUC scores for AP and traditional QSAR to assess overfitting. Almost all properties represented by the different points in the AP plot fall on or very close to the diagonal line, versus the traditional classifiers where validation performance tends to be higher than test performance, indicating overfitting. Androgen Receptor (AR) binding models appear to overfit with both tools; however, AP still performs better (Test: 0.62) than the best traditional classifier (Test: 0.50). The inability of traditional classifiers to use many fingerprints at the same time leads to inconsistency in the type of fingerprints/classifiers generating the best performing models, as shown in **Table 1**. This inconsistency is clearly shown in Cytochrome P450 isoforms where for each isoform, a different fingerprint/classifier outputs the best performing classifier.

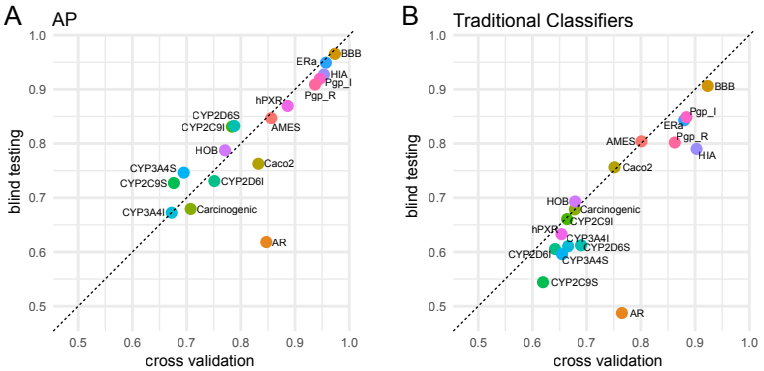


Figure 2. Overfitting between AP and Traditional Classifiers visualized by a dot plot with a diagonal line a) Comparison of ROC AUC scores generated by the AP Test and AP Full models b) The ROC AUC scores of cross validation & blind testing for traditional classifiers.

To address concerns related to the implementation of traditional classifiers, we also compared AP performance with previously reported results in the literature. For example, Shen *et al* report on AdmetSAR, a tool used by DrugBank and assessed for performance on Blood Brain Barrier (BBB) and Human Intestinal Absorption (HIA) properties. Based on their report AdmetSAR ROC AUC scores (BBB: 0.9517, HIA: 0.9458) are outperformed by AP (BBB: 0.9738, HIA: 0.9533)[3].

Table 1. A summary of the performance (ROC AUC score) of 17 ADMET properties of AP and best-performing, traditional, classifier-fingerprint combination.

Property	AP Full	AP Test	Traditional	Classifier	Fingerprint
AMES Toxicity	0.8562	0.8469	0.8041	Gradient Boosting	Pattern
Androgen Receptor	0.8470	0.6183	0.4874	Stochastic Gradient Descent	Morgan Rad:2
Blood Brain Barrier	0.9736	0.9655	0.9063	Stochastic Gradient Descent	Pharm Gobbi
Caco-2 permeability	0.8323	0.7627	0.7562	Gradient Boosting	Pharm Gobbi
Carcinogenic	0.7074	0.6794	0.6786	Random Forest	Layered
CYP450 2C9 Inhibitor	0.7836	0.8312	0.6607	Gradient Boosting	Layered
CYP450 2C9 Substrate	0.6766	0.7272	0.5442	Ridge	Layered
CYP450 2D6 Inhibitor	0.7512	0.7308	0.6054	Gradient Boosting	Pharm Base
CYP450 2D6 Substrate	0.7879	0.8325	0.6124	Stochastic Gradient Descent	Pharm Base
CYP450 3A4 Inhibitor	0.6726	0.6724	0.6107	Gradient Boosting	Pattern
CYP450 3A4 Substrate	0.6946	0.7464	0.5960	Support Vector Machine	Atom Pair
Estrogen Receptor alpha	0.9570	0.9492	0.8430	Gradient Boosting	Pharm Base
Human Intestinal Absorption	0.9533	0.9273	0.7900	Ridge	Layered
Human Oral Bioavailability	0.7708	0.7875	0.6931	Gradient Boosting	Pharm Base
Human Pregnane X Receptor	0.8864	0.8694	0.6329	Ridge	Daylight
P-glycoprotein Inhibitor	0.9458	0.9195	0.8487	Gradient Boosting	Pharm Gobbi
P-glycoprotein Recognition	0.9368	0.9088	0.8019	Gradient Boosting	Pattern

SUMMARY

Available tools for activity prediction of pharmacokinetic and toxicity properties use traditional classifiers and machine learning algorithms to build models that predict molecular properties based on measures of chemical similarity. Typically, models are generated with multiple distinct classifiers and fingerprints and the top performing models are selected. This process can lead to inconsistent accuracies and increased risk of overfitting due to the increase number of hyper-parameters and through selection bias. Furthermore, traditional classifiers lack the ability to use multiple fingerprints simultaneously in a model, and so risk including any systematic bias of the fingerprinting method used. Cyclica's AP tool uses a proprietary, multi-fingerprint, machine learning method to mitigate many of the issues introduced by the traditional classifiers. We compared the performance of AP with that of five traditional classifiers in predicting 17 ADMET properties. AP consistently outperforms the traditional classifiers, both in our hands and in comparison with previously published performance studies.

RESOURCES

1. Van Norman G. A., *et al.* Drugs, Devices, and the FDA: Part 1: An Overview of Approval Processes for Drugs. *J Am Coll Cardiol Basic Trans Science.* **1**, 170-179 (2016)

2. DiMasi, J. A., *et al.* Innovation in the pharmaceutical industry: new estimates of R&D costs. *J. Health Econ.* **47**, 20–33 (2016)

3. Shen J., *et al.* Estimation of ADME Properties with Substructure Pattern Recognition. *J. Chem. Inf. Model.* **50**, 1034-1041 (2010)

4. Kim M., *et al.* Critical Evaluation of Human Oral Bioavailability for Pharmaceutical Drugs by Using Various Cheminformatics Approaches. *Pharm Res.* **31**, 1002-1014 (2014)

CYCLICA INC.

18 King St East, Suite 801
Toronto, Ontario, M5C 1C4, Canada
1-416-304-9201