AI/ML AND VIRTUAL HUMAN PLATFORMS FOR THREAT AGENT HAZARD ASSESSMENT AND MEDICAL COUNTERMEASURE DISCOVERY AND DRUG DEVELOPMENT

Machine Learning Models For Predicting Human Liver Microsomal Metabolism Of Organophosphate Pesticides

Thomas Lane Collaborations Pharmaceuticals, Inc **Garima Agarwa** Ohio State University, Columbus, OH **John Corbett** Collaborations Pharmaceuticals, Inc **Fabio Urbina** Collaborations Pharmaceuticals, Inc **Kemal Ozlap** Collaborations Pharmaceuticals, Inc **Craig McElroy** Ohio State University, Columbus, OH **Sean Ekins** Collaborations Pharmaceuticals, Inc

Little is known about the human metabolism of organophosphate pesticides (OPPs). Accidental exposure and intentional poisoning by OPPs kill many thousands of people every year. Over the past 20 years, large quantities of in vitro and in vivo data have accumulated on drug metabolism in humans. The availability of drug metabolism data enables the building of predictive computational models using molecular structure. Machine learning (ML) methods have been applied to many of these datasets to enable prospective prediction and increase efficiency of screening. We have curated data for human liver microsome (HLM) metabolism from public sources, followed by building of classification ML models based on intrinsic clearance values (Clint) with our Assay Central software. Next, we generated clearance rates for 54 OPPs and then used these as an external test set. The results showed poor predictions with classification models. As OPPs had very low representation in the model training sets, it was unsurprising that they were poorly predicted. We therefore built classification and regression models for the 54 OPPs alone with 8 classification (deep learning (DL), adaboosted decision trees (ada), Bernoulli naïve bayes, k-nearest neighbors (knn), random forest (rf), support vector classifier (svc), logistic regression and XGBoost (xgb)) and 7 regression (Ada, Bayesian, elastic net, knn, RF, support vector machine and xgb) different ML algorithms, respectively. This approach still showed poor predictive ability for OPP metabolism likely due to the small, limited training sets, therefore we explored expanded our methodology.

Most QSAR revolves around predicting single point values of a molecule-target interaction. Single-point predictions do not take advantage of the full dose-response curve often generated for these drug-target interactions by collapsing these data into a single-value with potentially critical information loss. Recently, we have shown the ability to predict UV-vis spectra curves, which is continuous data, from molecular structure using recurrent neural network (RNN)-based models. We have applied the RNN long short-term memory (LSTM) algorithm to predict full dose-response curves based on molecule structure allowing us to predict not only the clearance of these molecules, but also the maximal response at different concentrations. In addition, we have also used both multivariate Random Forest (MRF) and multivariant linear regression (MLR) to establish if these more classical approaches were able to enhance the predictive abilities of the model. Utilizing these various methods, we tried to different options to optimize this method using multiple types of molecular descriptors (e.g. Morgan fingerprints with variable radii and bit length, MACCS keys and simple chemical descriptors such as MW and polarizability). Each of these methods were assessed using a 5-folded nested cross validation approach. These models outperformed the more classical approaches but our data suggests that more optimization is still needed. Although this dose-response-curve method still requires additional development it shows promise as an approach for the accurate computational prediction of metabolism of OPPs and potentially other compounds.