Predictive models of target organ and systemic toxicities

Imran Shah,[1] Matt Martin,[1] Chris Grulke,[1] LyLy Pham,[1,2] Jie Liu,[1,2,3] Antony Williams,[1] Grace Patlewicz,[1] Russell S. Thomas[1]

[1]NCCT/ORD/US EPA, [2]ORISE, [3]UALR

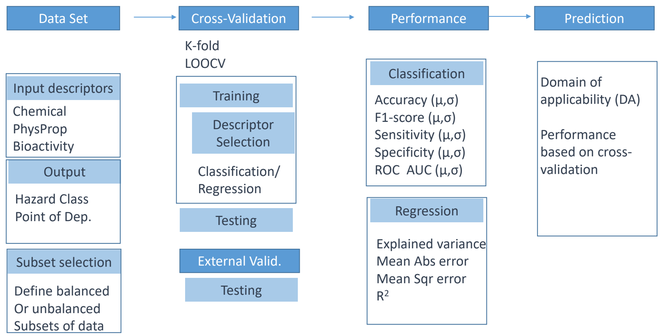
Presenting author: orcid.org/0000-0003-3863-9689| ?? | email: ??@epa.gov |

Abstract

The objective of this work is to predict the hazard classification and point of departure (PoD) of untested chemicals in repeat-dose animal testing studies. We used supervised machine learning to objectively evaluate the predictive accuracy of different classification and regression algorithms using chemical structure information, physico-chemical properties, and in vitro bioactivity data. The mean F1 score for predicting 20 target-organ hazard classes across three guideline study types was 0.69, and the R2 for predicting the PoD for systemic toxicity was 0.38. These models can be used to efficiently prioritize tens of thousands of environmental chemicals by hazard and by PoD.

Approach

General-purpose supervised machine learning (ML) pipeline



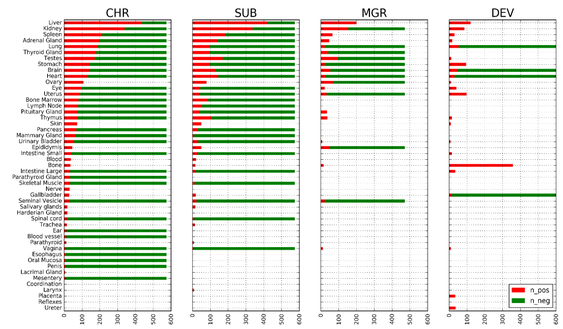
Target organ toxicity prediction

- Chemicals represented by: Morgan fingerprints (chm), Toxprint chemotypes (ct), ToxCast in vitro bioactivity (bio), and hybrid descriptors
- 5-fold cross-validation testing to evaluate the performance of hazard classification using Naïve Bayes (NB), k-nearest neighbor (KNN), random forest (RF), classification and regression trees (CART), and support vector classification (SVC). Model performance assessed using F1-score
- Fixed effects modeling to relate variance in F1 score to target organ outcome, descriptor type, classification algorithm and interactions between these

Point of departure estimation:

- Chemicals represented by: In-House physical and chemical properties, PaDEL descriptors, ToxPrint chemotypes, ToxCast Bioactivity, High-throughput Toxicokinetic parameters (httk)
- Training and testing (5-fold CV with 10X bootstrapping) PoD values were collected from ToxRefDB, HESS-DB, and COSMOS-DB
- Collection of Random Forest Models developed and calibrated

Results



Distribution of positive and negative chemicals across the *in vivo* guideline toxicity testing studies and target organs.

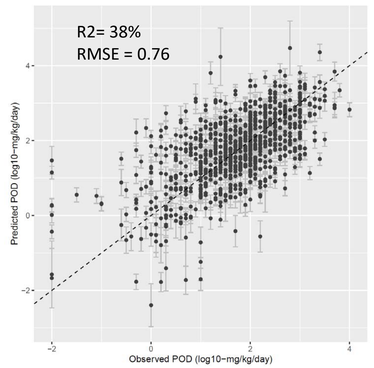
Summary of performance for a subset of target organ outcomes. The visualization shows the predictive performance for toxicity outcomes in rows (study:target organ) using eight machine learning algorithms (columns): naïve Bayes (NB), k-nearest neighbor classification (KNN0 and KNN1) classification and regression trees (CART0 and CART1) and support vector classifiers (SVCL0 and SVCR0). The predictive performance is compared across five different descriptors including: chemical (chm), chemotype (ct), *in vitro* bioactivity (bio), a combination of *in vitro* bioactivity and chemical (bc), and a combination of *in vitro* bioactivity and chemotype (ct).). The performance of a classification method for predicting an outcome using a descriptor type was measured using specificity (green), F1 score (red) and sensitivity (blue), which are visualized as vertical glyphs. The center, top, and bottom of the glyphs correspond to the mean ±1 SD.

Data for Target Organ Toxicity Prediction 985 Chemicals from ToxCast Phases I, II, III represented by the following descriptors: 2,048 Morgan fingerprints (chm), 729 Toxprint chemotypes (ct), 821 bioactivity assays (bio) and 574 target organ outcomes from ToxRefDB (tox). The outcomes from ToxRefDB were aggregated at the level of guideline study and target organ. There were 35 target organ hazard classes with at least 50 positive and 50 negative chemicals.

Terminal Variable Set	R^2	RMSE	Top 5 Features from Terminal Variable Set (Importance Rank)
In vivo (baseline)	21%	0.85	dose_spacing (1); strain_group (2); pod_qual (3); study_type (4); dose_no (5)
chemical descriptors (physchem, PaDEL, ToxPrint)	37%	0.76	ATSC4m (5); AATSC1m (9); SC-6 (10); ALogP (11); MDEO-11 (12)
biological descriptors (ToxCast)	37%	0.76	burst (45); estrogen_receptor (51); zebrafish_development (114); peroxisome_proliferator_activated_receptor_alpha (125); xenobiotic_metabolism_induction (172)
kinetic descriptors (httk)	38%	0.76	logmean (151); peak (276); fub (365); vdist (385); intcl (400)
mean POD (benchmark)	73%	0.5	podmn (1); strain_group (2); dose_spacing (3); study_type (4); pod_qual (5)

Predictive model of systemic toxicity points-of-departure

4382 studies across 1201 chemicals were used to develop (train/test) and externally evaluate the Random Forest Regression models. The table (above) shows each models performance based on R2 and RMSE as well as the top 5 most important variables for each of the added variable sets. The figure (right) shows the predicted PoD estimate with 95% confidence intervals versus the observed PoD in the external test set.



Summary

Target organ toxicity prediction

- Mean F1 score across all hazard classes ~ 0.69.
- Best predictions and uncertainty estimates for each target organ using different descriptors are provided for transparency
- Variance in F1 score explained in descending order: target organ > descriptor type > classification algorithm
- Combining bioactivity and chemical descriptors improves performance

Point of departure estimation

- Baseline models explain ~15-20% of the total variance
- Benchmark models explain ~60-70% of the total variance
- Final consensus model explains 38% of the total variance in the external validation dataset
- Developed model to be able to predict values with only chemical descriptors, or a combination of chemical, biological and kinetic descriptors
- Uncertainty estimates (95% CI) for each prediction are provided for transparency

Impacts

- The predictive models can be used to prioritize tens of thousands of untested chemicals using structure data alone
- Predictive models of target organ toxicity (classification models) will assist in: (a) grouping of chemicals with similar toxicity profiles, and (b) provide bioactivity signatures that can used to inform adverse outcome pathways (AOPs)
- Predictive models of systemic toxicity PoD (regression models) can be used to inform, as part of a weight of evidence approach, on the safety of chemicals with limited to no ability to be tested in traditional animal models (e.g., European Union Cosmetic Directive chemicals)
- Future directions: (a) Better characterize and quantify sources of variability in traditional animal studies to improve benchmarking of alternative test methods and models; and (b) continue to improve and expand our input variables (i.e., chemical descriptors)

References

Liu J, Patlewicz G, Williams, A, Thomas, R, Shah, I. Predicting Target Organ Toxicity using in vitro bioactivity and chemical structure. *Submitted*.