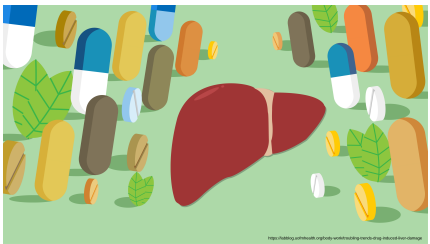
A fully *in Silico* framework to predict Drug-induced liver injury in early stage of drug development

Vahid Nassiri - Open Analytics
joint work with **Mohan Rao**, Helena Geys, Onyi Irrechukwu, James Damewood, Freddy Van Goethem, Jan Snoeys, and Yvonne Will (Janssen Pharmaceuticals)

October 20, NON-CLINICAL STATISTICS CONFERENCE NCS 2022,
Louvain-la-Neuve, Belgium

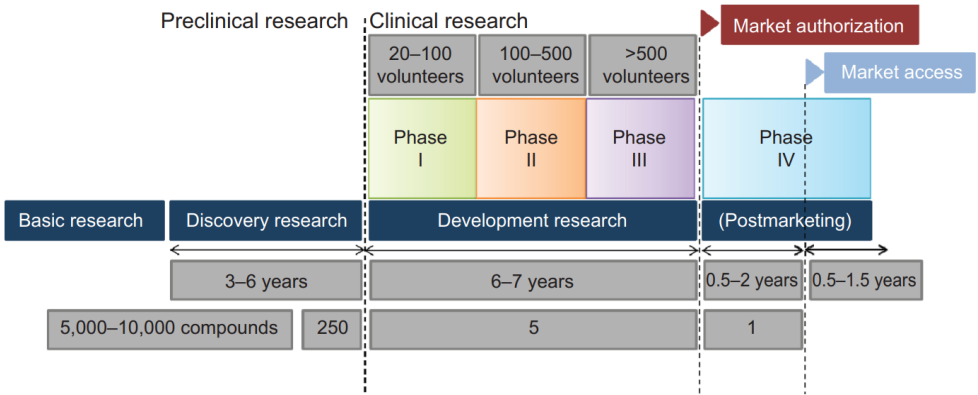
Drug-induced Liver Injury (DILI)



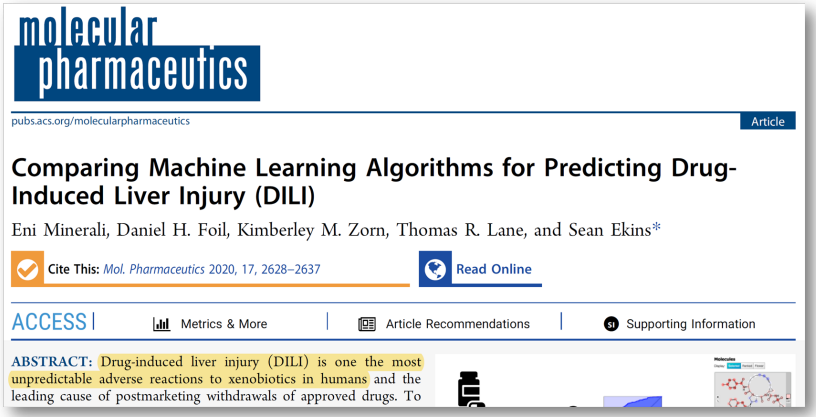
The liver plays a central role in transforming and clearing chemicals and is susceptible to the toxicity from these agents (Wikipedia!).

After Cardiotoxicity, DILI is the second leading reasons for drug-withdrawal.

Fail Early, Fail Cheap!



Ciani, Oriana, and Claudio Jommi. "The role of health technology assessment bodies in shaping drug development." Drug design, development and therapy (2014): 2273-2281.



molecular pharmaceutics

pubs.acs.org/molecularpharmaceutics

Article

Comparing Machine Learning Algorithms for Predicting Drug-Induced Liver Injury (DILI)

Eni Minerali, Daniel H. Foil, Kimberley M. Zorn, Thomas R. Lane, and Sean Ekins*

Cite This: *Mol. Pharmaceutics* 2020, 17, 2628–2637

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Drug-induced liver injury (DILI) is one the most unpredictable adverse reactions to xenobiotics in humans and the leading cause of postmarketing withdrawals of approved drugs. To

DILI predictive modelling

It is difficult!

- Small number of compounds with known clinical DILI severity class
- Unknown and/or multifaceted mechanisms
- Systematically biased set of compounds (all already approved by FDA)

DILI predictive modelling

It is difficult!

- Small number of compounds with known clinical DILI severity class
- Unknown and/or multifaceted mechanisms
- Systematically biased set of compounds (all already approved by FDA)

Consequence: AI/ML Magic Fails!

DILI predictive modelling

It is difficult!

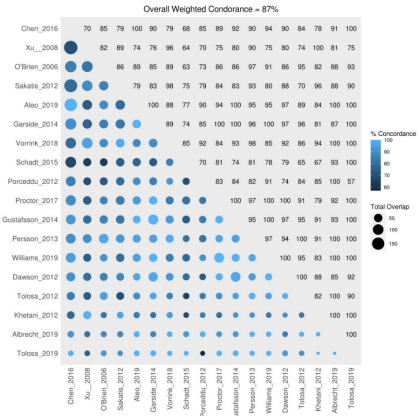
- Small number of compounds with known clinical DILI severity class
- Unknown and/or multifaceted mechanisms
- Systematically biased set of compounds (all already approved by FDA)

Consequence: AI/ML Magic Fails!

- Identify possible DILI predictors.
- Use already approved and validated AI/ML methods to predict them.
- Build a predictive model based on them.

DILI severity classification

Chen et al. (2016) used FDA approved drug labels to classify compounds



Martin, Matthew T., et al. "Early Drug-Induced Liver Injury (DILI) Risk Screening:"Free", as good as it gets." Toxicological Sciences (2022).

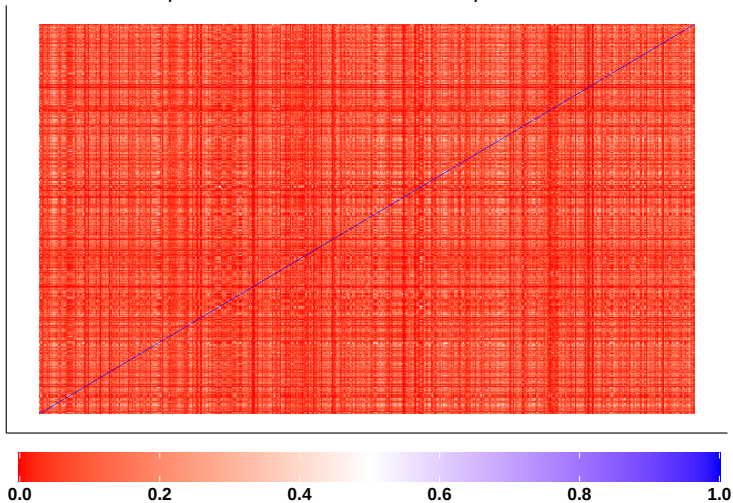
No DILI-concern

Less DILI-concern

Most DILI-concern

Selected compounds

194 No DILI-concern, 245 Less DILI-concern, and 164 Most DILI-concern

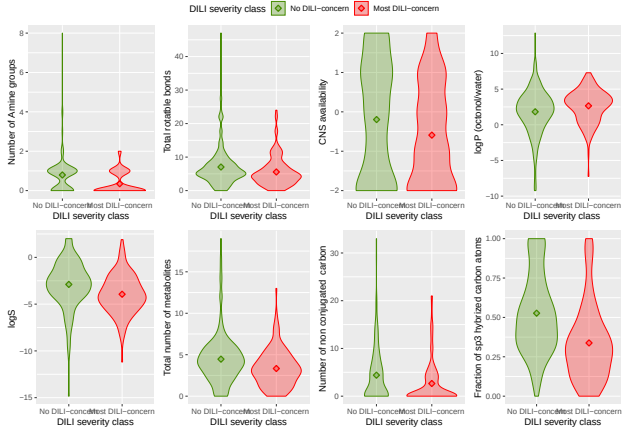


Physico-chemical properties

Physico-chemical properties are among the ever-present descriptors of DILI.

The main properties considered in the literature are Lipophilicity, and, fraction of sp 3 carbon atoms see e.g. Norman (2020).

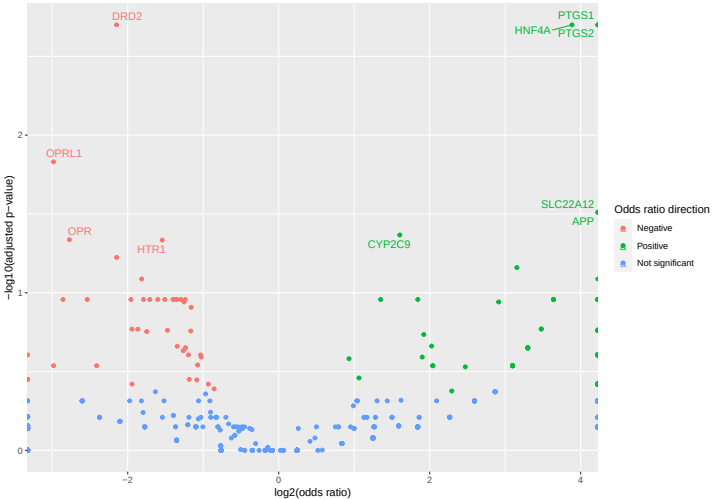
We consider 25 different Physico-chemical properties computed fully *in Silico* using Schrödinger's QikProp.



Predicted off-targets

- Transcriptomics gained less attention.
- Based on Rao et al. (2019), AI/Ml models were produced using more than 21 million small molecules and six different methods to predict interactions with ~6000 human genes fully *in Silico*.
- We have identified 1666 unique targets; and 161 that could significantly separate No from Most DILI-concern.

Predicted off-targets v.s. DILI



Model building

- **Begin** with (1666+25) predictors.
- **Variable selection:** a LASSO logistic regression together with expert knowledge.
- **Prediction:** quantify the DILI risk in terms of (pseudo-)probability of Most DILI-concern using:
  - Ridge logistic regression
  - Neural network
  - Random forest
  - Support vector machines

Model building

- **Single Probability Score:**.
  - A weighted average (using accuracies from 10-fold CV as weights): $Score_{WA} = w_1 RLR + w_2 NN + w_3 RF + w_4 SVM$,
  - Ensemble learning (using a penalized logistic regression equation): $Score_{EL} = \frac{1}{1+\exp(-s)}$, where $s = \beta_0 + \beta_1 RLR + \beta_2 NN + \beta_3 RF + \beta_4 SVM$.
- **Class Prediction:** $Score > 0.6 \Rightarrow \mathrm{DILI+}$.

Black ~~— Box~~, Explainable but also Explained

Select Key Off-Target Biology	
Kinases	VEGFR1, ABL1, RET, AURKA and FYN
Enzymes	COX1, COX2, AKR1C3 and XDH
Nuclear Receptor	AR, PPARG and RXRA
GPCR	DRD2, OPRL1, OPR, CHRM, HTR2, HRH1 and ADRA2
Transporter	SLC22A12 & SLC6A4
Cytochrome(s)	Cyp1A2 and Cyp2C9

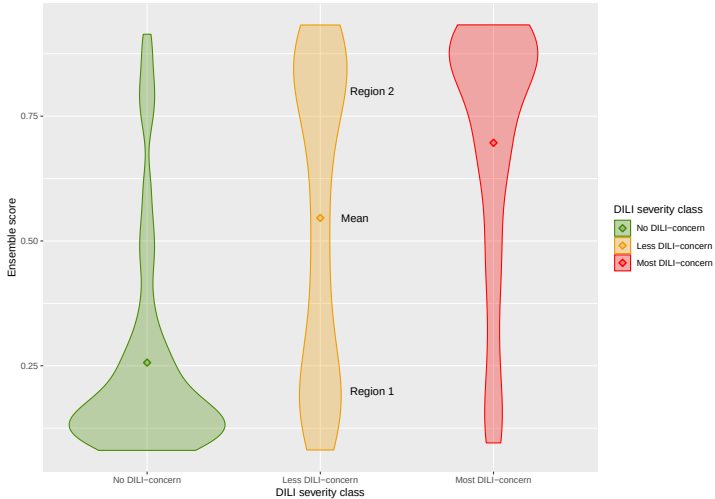
Chemical Properties

1. Csp^3 (Flatness index)
2. logS (Solubility) or logP (lipophilicity)
3. Number of metabolites
4. Number of free amines
5. Number of reactive functional groups

Performance

Method	Sensitivity	Specificity	PPV	NPV	LR+	Accuracy
Penalized logistic regression	0.658	0.912	0.872	0.765	8.197	0.796
Neural network	0.689	0.891	0.848	0.775	7.406	0.798
Random forest	0.598	0.907	0.850	0.731	7.056	0.763
Support-vector machine	0.698	0.877	0.832	0.779	7.008	0.796
Weighted average	0.694	0.907	0.866	0.784	8.341	0.821
Ensemble learning	0.732	0.897	0.862	0.804	8.064	0.809

Predicting Less DILI-concern



Predicting Less DILI-concern

DILI score	Warnings and precautions
< 0.6	23
> 0.6	47

Retrospective prediction

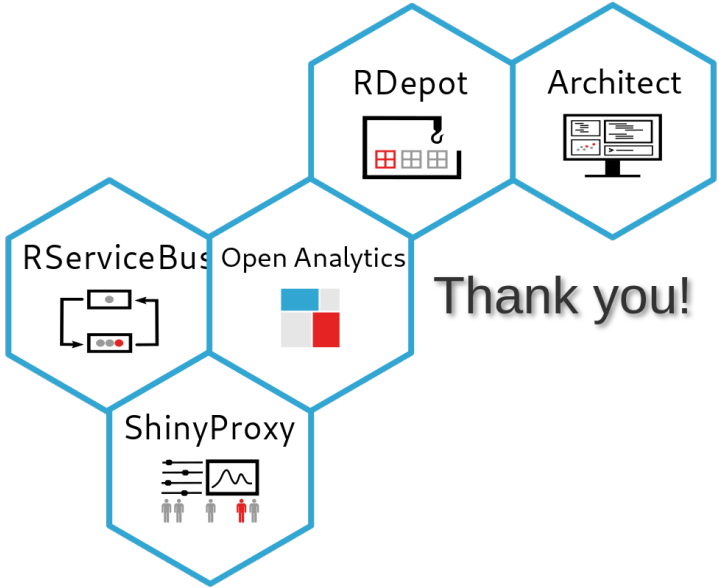
Drug	Mode of Action	DILI score	Predicted risk class	Status for Findings
Entacapone	COMT	0.91	High Risk	Less DILI concern
Thelin	ETA	0.67	Moderate Risk	Hepatobiliary disorder in clinic
Macitentan	ETA	0.61	Moderate Risk	ALT, AST increased in Post marketing
Ambrisentan	ETA	0.23	No Risk	High risk
LY-2409021	GCGR	0.82	High Risk	Transaminase elevation in clinic
MK-0893	GCGR	0.67	Moderate Risk	Transaminase elevation in clinic
PF-05020182	KCNQ1	0.23	No Risk	High risk
Flupirtine	KCNQ1	0.69	Moderate Risk	Warning with precaution- Most DILI
PF-04895162	KCNQ1	0.78	High Risk	Terminated at clinical development
Montelukast	LTD4	0.91	High Risk	Marketed with DILI warning
Verlukast	LTD4	0.91	High Risk	Terminated at clinical development
CP-85958	LTD4	0.55	Moderate Risk	Terminated at clinical development
Pioglitazone	PPAR	0.75	High Risk	Warning with precaution
Rosiglitazone	PPAR	0.94	High Risk	Warning with precaution
Ciglitazone	PPAR	0.77	High Risk	Terminated at clinical development
Englitazone	PPAR	0.57	Moderate Risk	Terminated at clinical development
Fasiglifam	GPR40	0.84	High Risk	Terminated due to DILI

References

Chen, Minjun, Ayako Suzuki, Shraddha Thakkar, Ke Yu, Chuchu Hu, and Weida Tong. 2016. "DILIrank: The Largest Reference Drug List Ranked by the Risk for Developing Drug-Induced Liver Injury in Humans." *Drug Discov Today* 21 (4): 648–53.

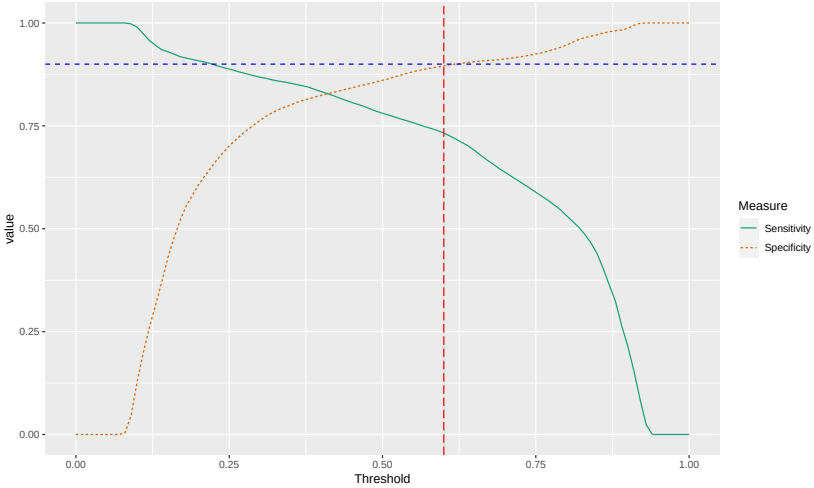
Norman, Bryan H. 2020. "Drug Induced Liver Injury (Dili). Mechanisms and Medicinal Chemistry Avoidance/Mitigation Strategies." *Journal of Medicinal Chemistry* 63 (20): 11397–419.

Rao, Mohan S, Rishi Gupta, Michael J Liguori, Mufeng Hu, Xin Huang, Srinivasa R Mantena, Scott W Mittelstadt, Eric AG Blomme, and Terry R Van Vleet. 2019. "Novel Computational Approach to Predict Off-Target Interactions for Small Molecules." *Frontiers in Big Data* 2: 25.



Thank you!

Threshold selection



Cross-validation

- Split the 358 compounds randomly into k chunks of roughly equal sizes.
- For *iFold* from 1 to 10
  - Predict the compounds in chunk iF old using all the compounds in 9 remaining chunks,
  - Compute and store performance measures for these predictions.
- Compute and store mean of 10 computed performance measures Repeat steps 1–3, 100 times
- Report the median of mean