Predicting Drug-Induced Liver Injury Using Machine Learning: A Novel Approach Integrating Multiple *In Vitro* Assay

○Hatsue Furukawa, Qinge Liang, Yoriko Ueda, Masamitsu Ando, Shun Saito, Yutaka Nakanishi, Kazunobu Aoyama
Axcelead Drug Discovery Partners, Inc.

Introduction and Objective

Drug-induced liver injury (DILI) is one of the major causes of drug development discontinuation and post-marketing withdrawal. Therefore, early prediction/mitigation of DILI is critically important in drug discovery. However, due to the diverse underlying mechanisms of DILI, a single assay cannot ensure high predictive performance. In this study, we first evaluated the predictive performance of six parameters obtained in drug discovery using receiver operating characteristic (ROC) curve analysis. In addition, the predictive performance was also evaluated using clinical Cmax values for the same compounds. Furthermore, we constructed a machine learning (ML) model incorporating molecular descriptors and evaluated the predictive performance of an integrated DILI risk assessment approach. Adjusted descriptor weighting optimized the predictive performance and stratification into distinct DILI-risk groups.
In conclusion, our ML model enabled not only DILI risk classification with Cmax-based scores but also determination of plasma drug concentrations indicative of DILI risk. This approach is applicable even for DILI risk assessment of compounds with unknown efficacious Cmax required in humans.

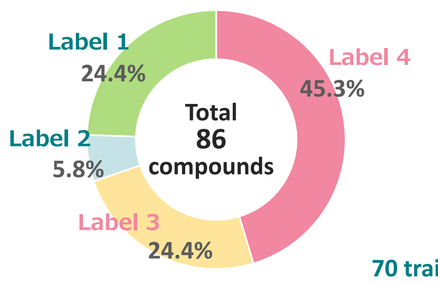
Methods

1) Compound dataset

A total of 86 compounds were selected based on the DILIrank 2.0 dataset [1] and literature data [2,3]. In DILIrank, compounds are classified into four categories and assigned labels 4, 3, 2, and 1, respectively. For the literature data [2,3], labels were assigned according to the reported category information. The compounds were further classified into DILI-positive and DILI-negative groups for validation. The classification criteria of DILIrank are shown in the table below.

	DILIrank	DILI classification
Label 4	vMost-DILI	Positive
Label 3	vLess-DILI	
Label 2	Ambiguous-DILI	Negative
Label 1	vNo-DILI	

Distribution of compounds by DILI risk category



2) Cmax information

Cmax values were mainly obtained from literature data [4] summarizing therapeutic or toxic plasma concentration.

3) Predictive performance of each parameter

Prior to machine learning, parameters derived from four assays in the lead optimization stage, together with Cmax, were selected. Their predictive performance was assessed by ROC analysis using 70 compounds (Table 1).

4) Molecular descriptors (MDs)

A total of 209 MDs were calculated from SMILES using RDKit. Based on predefined selection criteria, nine MDs were selected. Together with AlogP, a total of ten MDs were used for DILI evaluation.

MDs	MD family
AlogP	Lipophilicity, MR Drug-likeness
AMW	Size, Composition
BCUT2D MRHI	Fingerprint BCUT
BertzCT	Topology
HallKierAlpha	
EState VSA3	Surface VSA
PEOE VSA6	
MinAbsEStateIndex	Electronic Charge
MinPartialCharge	
FractionCSP3	Flexibility, Ring properties

5) Machine Learning (ML) model

A machine learning model was constructed based on the following methods and subsequently evaluated using test compounds. In addition, the weighting of MDs was adjusted to optimize the predictive performance. Furthermore, DILI risk scores were calculated across the Cmax sweep. DILI-positive/negative classifications were determined based on the threshold.

Input Data Preparation

70 training compounds
• 49 DILI-positive
• 21 DILI-negative

In-house wet-lab screening data
Data Acquisition
Screening assays
↓
Final Feature Set
- Cytotox 24hr Glucose IC_{50}/Cmax
- Cytotox24hr Glucose/Galctose ratio
- BSEP IC_{50}/Cmax
- Phospholipidosis relative AUC (%) compared with amiodarone (positive control)
- dGSH trapping: reactive metabolite formation rate (%)

MD data
Data Acquisition
SMILES (source: PubChem) → RDKit Descriptors
↓
Feature Selection
Low-variance and highly correlated MDs removed
Selected by stable importance and consistent directionality
↓
Final Feature Set
9 RDKit MDs + AlogP (source: PipelinePilot)

Model Construction
Weighted Ensemble Machine Learning Prediction
Two Random Forest Regressor models:
Wet-lab model + Descriptor model (balance adjustable)

Internal Validation
3-fold repeated cross-validation
OOF (Out-of-Fold) predictions (OOF-AUC) for model performance evaluation

Model evaluation using 16 test compounds
• 11 DILI-positive
• 5 DILI-negative

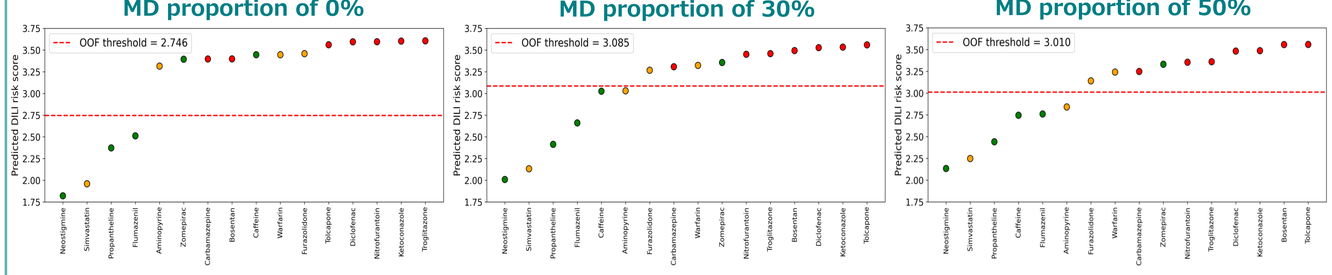
Results

Table 1 Predictive performance of individual parameters

Mechanisms of toxicity	Parameters	AUC	Classification Performance		
			Sensitivity	Specificity	Youden index
Cytotoxicity	Cytotox 24hr IC_{50}/Cmax	0.72	76%	71%	0.47
Mitochondrial toxicity	Cytotox 24hr Glu/Gal ratio	0.56	18%	95%	0.14
Cholestasis	BSEP IC_{50}/Cmax	0.75	63%	81%	0.44
Phospholipidosis	Relative value to the AUC of amiodarone (%)	0.44	80%	29%	0.08
Reactive metabolite（RM）	RM Formation rate (%)	0.52	41%	76%	0.17
Plasma concentration	Log Cmax	0.74	92%	52%	0.44

Sensitivity and specificity at the cutoff value that maximizes the Youden index were calculated, and the area under the ROC curve (AUC) was evaluated to assess the predictive performance of each parameter.

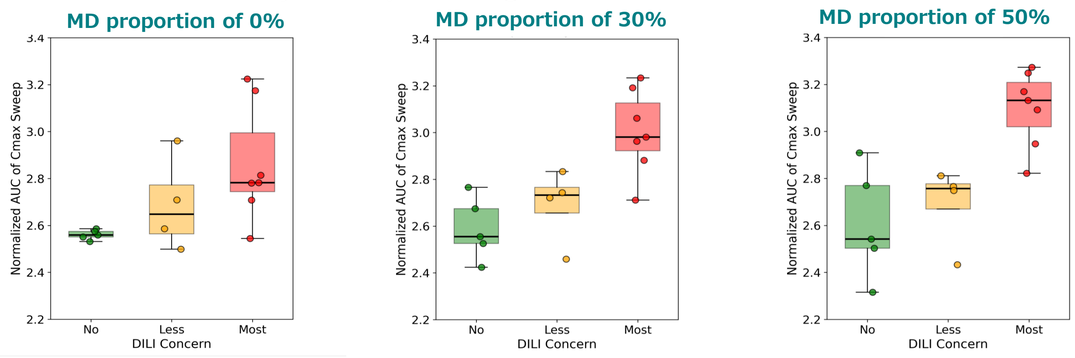
Table 2 Validation results of the ML model with different MD weightings



	MDs 0%	MDs 30%	MDs 50%
ROC-AUC	0.845	0.855	0.855
Sensitivity	90.9%	81.8%	81.8%
Specificity	60.0%	80.0%	80.0%

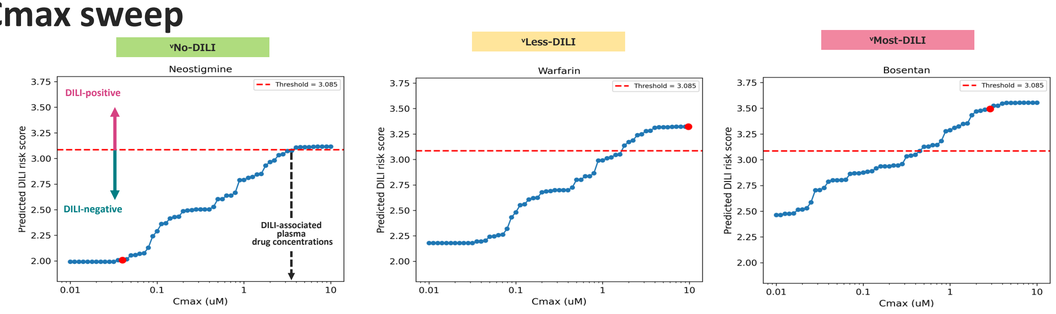
ML models constructed with different MD weightings were validated using 16 test compounds based on ROC-AUC, sensitivity, and specificity.

Figure 1 Improvement of Cmax sweep separation and DILIrank classification by incorporating MDs



The AUC values of DILI risk scores calculated within the Cmax sweep[1)] range of 0.01–10 µM were used to compare variability and group separation among DILIrank classes (vMost, vLess, and vNo-DILI) for each ML model constructed with different descriptor weightings.
1) Predicted model output using 61 inputs from 0.01 to 10 µM

Figure 2 Changes in risk scores and risk classification across the Cmax sweep



Changes in DILI risk scores calculated by the ML model across each Cmax sweep are shown (at 30% descriptor weighting). The red dashed line indicates the threshold for risk classification. If the predicted DILI risk score at the measured Cmax (red circles) exceeds the threshold, the compound is classified as DILI-positive. In addition, the plasma concentration corresponding to the threshold can be identified.

Summary

- Among the individual parameters, Cytotox 24hr IC_{50}/Cmax, Log Cmax, and BSEP IC_{50}/Cmax showed similarly high predictive performance based on ROC-AUC.
- The ML model with descriptor weighting of 30% and 50% showed relatively favorable ROC-AUC values and balanced sensitivity–specificity profiles; however, 30% showed a more favorable trend in terms of both stratification and variability of DILI risk scores in boxplot analysis.
- The ML model enables not only DILI risk classification based on known Cmax values but also the estimation of plasma drug concentrations associated with DILI risk. This approach allows DILI risk assessment even when human Cmax values are unknown.

Conclusion

In conclusion, the ML model provides a useful tool for DILI risk assessment in drug discovery, enabling both risk classification and estimation of plasma drug concentrations associated with DILI risk.

References
[1] Drug Induced Liver Injury Rank (DILIrank 2.0) Dataset
[2] Williams et al., Chem. Res. Toxicol., 2020（AstraZeneca）
[3] Aleo et al., Chem. Res. Toxicol., 2020（Pfizer）
[4] Schulz et al., Crit. Care, 2020

Acknowledgement
The authors would like to thank Hiroshi Kajino for his valuable support in machine learning.

COI Disclosure Information
Lead Presenter/Responsible Researcher:
Hatsue Furukawa
I have no financial relationships to disclose.