Toxicology 389 (2017) 139–145

Contents lists available at ScienceDirect

Toxicology

journal homepage: www.elsevier.com/locate/toxicol

Predicting drug-induced liver injury: The importance of data curation

Eleni Kotsampasakou, Floriane Montanari, Gerhard F. Ecker*

University of Vienna, Department of Pharmaceutical Chemistry, Althanstrasse 14, 1090 Vienna, Austria

ARTICLE INFO

Keywords:
Drug-induced liver injury
Random Forest
2-class classification
Liver transporters
Data curation
Toxicity reports

ABSTRACT

Drug-induced liver injury (DILI) is a major issue for both patients and pharmaceutical industry due to insufficient means of prevention/prediction. In the current work we present a 2-class classification model for DILI, generated with Random Forest and 2D molecular descriptors on a dataset of 966 compounds. In addition, predicted transporter inhibition profiles were also included into the models. The initially compiled dataset of 1773 compounds was reduced via a 2-step approach to 966 compounds, resulting in a significant increase (p-value < 0.05) in model performance. The models have been validated via 10-fold cross-validation and against three external test sets of 921, 341 and 96 compounds, respectively. The final model showed an accuracy of 64% (AUC 68%) for 10-fold cross-validation (average of 50 iterations) and comparable values for two test sets (AUC 59%, 71% and 66%, respectively). In the study we also examined whether the predictions of our in-house transporter inhibition models for BSEP, BCRP, P-glycoprotein, and OATP1B1 and 1B3 contributed in improvement of the DILI mode. Finally, the model was implemented with open-source 2D RDKit descriptors in order to be provided to the community as a Python script.

1. Introduction

Drug-induced liver injury (DILI) is the term used for liver damage that is caused by drugs, herbal agents or nutritional supplements (Ghabril et al., 2010; Watkins and Seeff 2006). DILI has gained increasing attention in recent years (Raschi and De Ponti, 2015), as it is one of the main causes for attrition during clinical and pre-clinical studies and the main reason for drug withdrawal from the market or for labeling with a black box warning (Ballet 1997; Chen et al., 2011; O'Brien et al., 2006; Regev 2014). Thus, great effort has been invested towards elucidating the toxicological processes and mechanisms that result in manifestations of DILI (Vinken, 2015). It is widely accepted that, together with metabolizing enzymes, liver transporters play an important role for maintaining the integrity and proper function of the liver, and also influence the ADMET (absorption, distribution, metabolism, excretion and toxicity) profile of drugs (Faber et al., 2003; Shitara et al., 2013). Actually, there are several recent publications suggesting that inhibition of liver transporters might result in manifestations of DILI. For cholestasis in particular, strong evidence towards the role of the bile salt export pump (BSEP) (Aleo et al., 2014; Dawson et al., 2011; Padda et al., 2011; Qiu et al., 2016; Vinken 2015; Vinken et al., 2013; Welch et al., 2015) has been posed. There is also evidence for the multidrug resistance-associated protein 2 (MRP2) (Padda et al., 2011; Pauli-Magnus and Meier 2006), breast cancer resistance protein (BCRP) (Padda et al., 2011; Pauli-Magnus and Meier 2006), P-glycoprotein (Padda et al., 2011; Pauli-Magnus and Meier 2006) and multidrug resistance-associated protein 3 and 4 (MRP3 and MRP4) (Padda et al., 2011; Pauli-Magnus and Meier 2006; Welch et al., 2015) to be involved. For hyperbilirubinemia, another possible manifestation of hepatotoxicity, involvement of organic anion transporting polypeptides 1B1 and 1B3 (OATP1B1 and OATP1B3) (Chang et al., 2013; Sticova and Jirsa 2013), MRP2 (Sticova and Jirsa, 2013) and to a smaller extent BCRP (Sticova and Jirsa, 2013) is discussed.

Although *in vitro* predictive methods are efficient for many toxic endpoints, they are time-consuming and expensive (Bowes et al., 2012; Whitebread et al., 2005). In addition, for assessing hepatotoxicity, experimental methods such as *in vitro* tests and animal models, have been shown to share low concordance (< 50%) with human hepatotoxicity (Chen et al., 2011; Liu et al., 2011; Olson et al., 2000).

This led to the development of predictive computational methods, which are summarized in two recent reviews by (Chen et al., 2014) and (Ekins, 2014). Although all these models generally perform quite well,

Abbreviations: Acc, Accuracy; ADMET, absorption, distribution, metabolism, excretion, toxicity; AUC, area under the curve; BA, balanced accuracy; BCRP, breast cancer resistance protein; cpd(s), compound(s); CV, cross validation; DILI, drug-induced liver injury; EV, external validation; IV, internal validation; MCC, Matthews correlation coefficient; MDR3, multidrug resistance protein; MRP2, multidrug resistance-associated protein 2; MRP3, multidrug resistance-associated protein 3; OATP1B1, organic anion transporting polypeptide 1B1; OATP1B3, organic anion transporting polypeptide 1B3; P-gp, P-glycoprotein; RF, Random Forest; SMO, sequential minimal optimization; sd, standard deviation; Sen, sensitivity; Spec, specificity; SVM, support vector machines

* Corresponding author.
E-mail address: gerhard.f.ecker@univie.ac.at (G.F. Ecker).

http://dx.doi.org/10.1016/j.tox.2017.06.003
Received 15 March 2017; Received in revised form 10 June 2017; Accepted 15 June 2017
Available online 23 June 2017

Table 1
Classification models for DILI reported in literature. Acc stands for accuracy, Sen for sensitivity, Spec for specificity, BA for balanced accuracy, CV for cross validation, EV for external validation and IV for internal validation.

Reference	Descriptors	Classification algorithm	Data used	Reported performance
Cheng and Dixon (2003)	2D molecular descriptor	Ensemble recursive partitioning	382 drugs for CV	CV: 76% Acc; 76% Sen; 75% Spec
			54 drugs for EV	EV: 81% Acc; 70% Sen; 90% Spec
Cruz-Monteagudo et al. (2008)	Radial distribution function	Linear discriminant analysis	74 drugs for CV	CV: 84% Acc; 78% Sen; 90% Spec
	molecular descriptors		13 drugs for EV	EV: 82% Acc
Matthews et al. (2009)	Molecular descriptors	4 commercial QSAR programs	~1600 drugs for CV	CV: 39% Sen; 87% Spec
			18 drugs for EV	EV: 89% Sen
Rodgers et al. (2010)	topological indices of molecular structures (MolConnZ) and Dragon molecular descriptors	k-nearest neighbor	37 drugs for EV	84% Acc; 74% Sen; 94% Spec
Fourches et al. (2010)	2D fragments and Dragon molecular descriptors	Support vector machine	531 drugs for CV 18 compounds for EV	CV: 62–68% Accs EV: 78% Acc
Ekins et al. (2010)	extended connectivity functional	Linear discriminant analysis	295 compound for CV	CV: 59% ACC; 53% Sen; 65% Spec
	class fingerprints of maximum diameter 6 (ECFC_6)		237 compounds for EV	EV: 60% Acc; 56% Sen; 67% Spec
Liew et al. (2011)	PaDEL molecular descriptor	Ensemble of mixed learning	1087 compounds for CV	CV: 68% Accs; 67% Sen; 70% Spec
			120 compounds for EV	EV: 75% Acc; 82% Sen; 65% Spec
Liu et al. (2011)	functional class	Bayesian models	888 drugs for training3 data sets with 40–148 drugs for EV	EV: 60–70% Accs
	fingerprints (FCFP_6)			
Chen et al. (2013)	Mold2 chemical descriptor	Decision Forest	197 drugs for CV Three data sets with 190–348 drugs for EV	CV: 70% Acc EV: 62–69% Accs
Liu et al. (2015a)	physicochemical descriptors and fingerprints	Ensemble classifier	677 compounds for CV	81% BA; 66% Sen; 95% Spec
Muller et al. (2015)	physicochemical descriptors and fingerprints	Ensemble classifier	677 compounds for CV	81% BA; 66% Sen; 95% Spec
Muller et al. (2015)	ISIDA fragment descriptors	SVM	424 drugs for CV	66% BA
Xu et al. (2015)	Encoding layers based on SMILES, PaDEL descriptors	Deep Learning	190, 475 & 1065 compounds for CV	CV: 70–88% Accs; 70–90% Sens; 70–87% Specs
			185,320, 236,198 & 119 compounds for EV	EV: 62–87% Accs; 62–83% Sens; 62–93% Specs
Mulliner et al. (2016)	2D and 3D physicochemical descriptors	SVM with a genetic algorithm	3712 compounds for training	IV: 75% Acc; 73% AUC
			221 compounds for IV 269 compounds for EV	
Zhang et al. (2016)	FP4 fingerprints	SVM	1317 compounds for training	Training set: 66% Acc; 85% Sen; 34% Spec; 55% AUC
			88 compounds for EV	EV: 75% Acc; 93% Sen; 38% Spec; 61% AUC

they sometimes suffer from low statistical performance, imbalanced sensitivity vs specificity, or small data sets (Table 1).

In this study we generate *in silico* classification models for DILI by compiling multiple and diverse datasets from literature. We carefully curated these data regarding the chemotypes, as well as the accuracy of the class label. In addition, we are exploring the importance of hepatic transporter inhibition on DILI by using the predictions of a set of in-house *in silico* classification models as additional descriptors for the DILI model.

2. Methods

2.1. Data compilation

2.1.1. Training set

Searching PubMed, 2017 (http://www.ncbi.nlm.nih.gov/pubmed), Google, 2017 (https://www.google.at) and Scopus, 2017 (https://www.scopus.com/) using the terms: "drug-induced liver injury", "DILI", "drug-induced hepatotoxicity" identified 9 unique datasets for human DILI/hepatotoxicity (Table 2).

For visualizing the data structures and for converting the names into structures Marvin from ChemAxon, 2013 (http://www.chemaxon.com 2013) was used.

2.1.2. External test sets

After compiling the training set and generating the DILI model, we came across one more human DILI dataset that had initially escaped our attention (Liew et al., 2011). Additionally, there were two more datasets published after the model development (Chen et al., 2016; Mulliner et al., 2016) (Table 3).

All datasets (training set, the three external test sets and the merged test set) are provided in the Supplementary material.

2.1.3. Chemical curation

For each dataset we applied the following chemotype curation:

- Check for inorganic compounds using MOE 2014.09. (MOE, 2015) and remove any occurring.
- Using the Standardiser tool (Atkinson, 2014) created by Francis Atkinson; all salt parts and any compounds containing metals and rare or special atoms are removed from the dataset and the structures are standardized.