Machine Learning based drug renal clearance predictor development and application

Natalia Łapińska (1), Sebastian Polak (1,2)
(1) Jagiellonian University, Medical College, Kraków, Poland, (2) Certara Predictive Technologies, Certara UK, Sheffield, UK

CERTARA Simcyp™

The developed model allows to predict human renal clearance based on relatively simple independent parameters

Background

Renal clearance is one of determinants of systemic exposure for many small molecules and their metabolites. Robust estimation of renal clearance remains a key objective from early drug candidates screening up to late-stage clinical pharmacology. At the pre-clinical level, renal clearance is characterized using in silico models, in vivo mass-balance and PK studies, kidney tissue distribution, and mechanistic in vitro systems (e.g., transporter assays and renal proximal tubule models) to dissect secretion and reabsorption pathways. At the clinical level, renal clearance is commonly estimated from urinary excretion data combined with plasma concentration profiles, supported by population PK, physiologically based PK, and biomarker-based assessments of renal function.

Objective

The objective of the study was **to develop and validate** machine learning (ML) based, empirical model allowing **human renal clearance prediction**. The developed model was further **applied** in the PBPK model of metformin HCl.

Materials and Methods

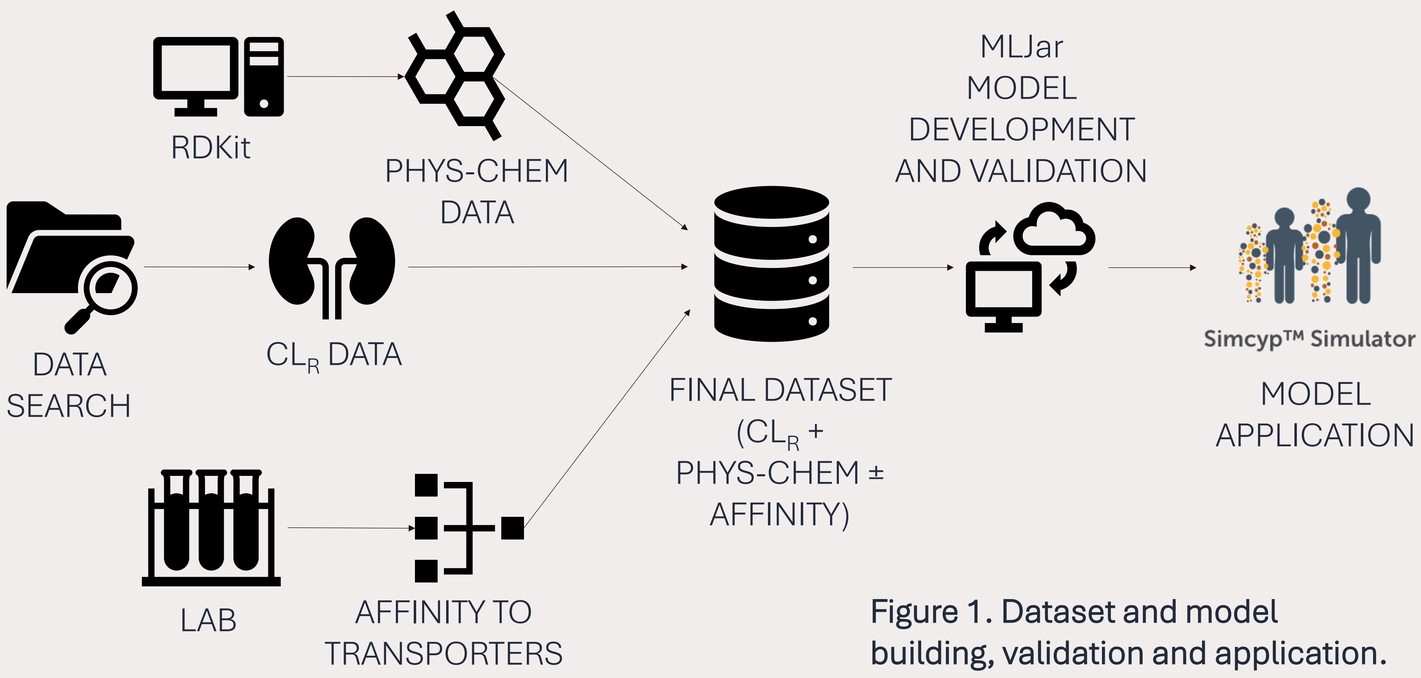


Figure 1. Dataset and model building, validation and application.

- final dataset - 1540 experimental renal clearance (CL_R) observations
- 462 unique compounds [1]
- molecular structures encoded with over 200 RDKit descriptors capturing physicochemical, topological, and functional group information [2]
- model incorporated study-level covariates including body weight, administered dose, route of administration, clearance determination method, and health status
- transporter interaction data (influx/efflux) were retrieved from the Drug Interaction Database (DIDB®) and encoded to reflect active renal transport mechanisms [3]
- model development was performed using an Automated Machine Learning (AutoML) MLJAR tool with compound-level splitting (90% training, 10% test set) [4]
- various scenarios with varying input vector were tested with the use of PBPK model built in Certara's Simcyp simulator (V25) was used to simulate kinetics of example drug, metformin, under various conditions: 1) without, 2) with clinically estimated, 3) with mechanistically modelled [5], 4) with the developed QSAR predicted CLR as an input [6]. Simulation of metformin kinetics after single dose was then compared versus observed data.

Results

The best model was an ensemble composed of XGBoost, Random Forest, LightGBM, and ExtraTrees single models. The model achieved R^2 = 0.84 (RMSE = 3.15) for the training set and R^2 = 0.61 (RMSE = 5.36) on the external test set, demonstrating moderate but robust generalization performance. SHAP analysis identified health status as the dominant covariate, while dose and route of administration showed minimal influence, supporting the hypothesis that intrinsic elimination mechanisms outweigh administration-related factors under linear pharmacokinetic conditions. Structural descriptors contributed through complex, non-linear interactions rather than simple monotonic relationships.

The developed model was used to predict CL_R for metformin (25 L/h) and compared versus the clinically estimated value (32.5 L/h). The observed Cmax and AUC_{inf} after single dose of 500 mg metformin HCl in healthy individuals varies between 1.0-1.4 µg/ml and 5.5-8.5 µg/ml•h [7,8]. The Simcyp simulated Cmax and AUC, single dose 500 mg metformin HCl for the above mentioned 4 scenarios were as follows: 1) 3.16/62.88 (no CL_R), 2) 0.95/6.52 (clinically estimated CL_R=32.5 L/h), 3) 1.06/7.55 (mechanistically modelled CL_R), 4) 1.10/8.07 (the developed QSAR predicted CL_R=25 L/h) µg/ml and µg/ml•h respectively.

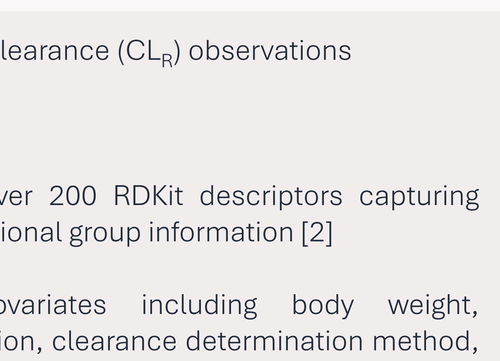
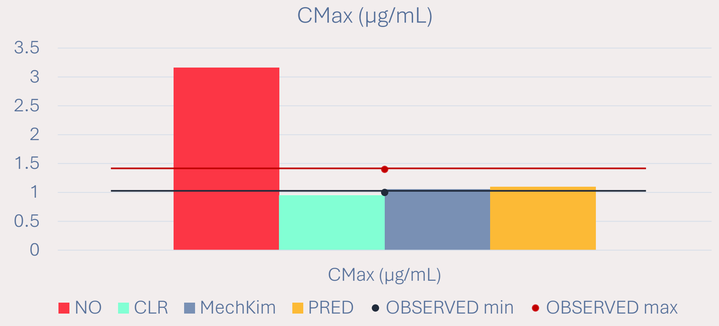


Figure 2. SHAP analysis results.



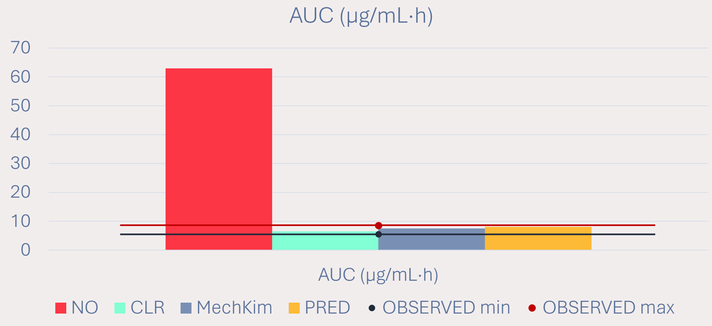


Figure 3. Metformin PK parameters for 4 tested scenarios

Conclusions

The proposed empirical model allows to predict CL_R based on **relatively simple** and **easy to obtain** independent **parameters**. The models' precision increases with the increase of input data details level, i.e. **even qualitative information on API's affinity to transporters allows for more robust CL_R estimation**. As part of the current study the predicted CLR value was used to parametrise example PBPK model built for metformin. In this case the predicted PK parameters, namely Cmax and AUC were overpredicted when CLR was not considered. Addition of the QSAR predicted CLR allowed to simulate realistic drug kinetics comparable with the clinical data.

The developed QSAR model can be also used for early prediction of CLR, even at the discovery stage.

References

[1] Łapińska, N; Polak, S, Human renal clearance of xenobiotics, Mendeley Data, V2, 2025, doi:10.17632/3427x3wzzc.2; [2] RDKit version 2025.3.3 https://www.rdkit.org/ ; [3] Drug Interaction Database (DIDB®) https://www.certara.com/drug-interaction-database-didb/; [4] Plonska A., Plonski P. MLJAR: State-of-the-Art Automated Machine Learning Framework for Tabular Data, version 0.10.3. https://github.com/mljar/mljar-supervised ; [5] Neuhoff, S. et al. (2013). Accounting for Transporters in Renal Clearance: Towards a Mechanistic Kidney Model (Mech KiM). In: Sugiyama, Y., Steffansen, B. (eds) Transporters in Drug Development. AAPS Advances in the Pharmaceutical Sciences, https://doi.org/10.1007/978-1-4614-8229-1_7; [6] Simcyp simulator https://www.certara.com/software/simcyp-pbpk/ ; [7] Santos-Caballero N et al., Comparative Pharmacokinetic Study between Metformin Alone and Combined with Orlistat in Healthy Mexican Volunteers", Pharmacology & Pharmacy 3(3)2012, doi:10.4236/pp.2012.33040; [8] Lucía Montoya-Eguía S. et al. Comparative Pharmacokinetic Study Among 3 Metformin Formulations in Healthy Mexican Volunteers: A Single-Dose, Randomized, Open-Label, 3-Period Crossover Study, Current Therapeutic Research, 77, 2015, doi:10.1016/j.curtheres.2014.09.003

Want to learn more?
<< Scan Here

certara.com