BioData Mining

RESEARCH Open Access

Understanding predictions of drug profiles using explainable machine learning models

Caroline König[1,2*] and Alfredo Vellido[1,2]

*Correspondence:
ckonig@cs.upc.edu

[1] Intelligent Data Science and Artificial Intelligence (IDEAI-UPC) Research Centre, Universitat Politècnica de Catalunya (UPC Barcelona Tech), Jordi Girona 1-3, Barcelona 08034, Catalonia, Spain
[2] Department of Computer Science, Universitat Politècnica de Catalunya (UPC Barcelona Tech), Jordi Girona 1-3, Barcelona 08034, Catalonia, Spain

Abstract

Purpose: The analysis of absorption, distribution, metabolism, and excretion (ADME) molecular properties is of relevance to drug design, as they directly influence the drug's effectiveness at its target location. This study concerns their prediction, using explainable Machine Learning (ML) models. The aim of the study is to find which molecular features are relevant to the prediction of the different ADME properties and measure their impact on the predictive model.

Methods: The relative relevance of individual features for ADME activity is gauged by estimating feature importance in ML models' predictions. Feature importance is calculated using feature permutation and the individual impact of features is measured by SHAP additive explanations.

Results: The study reveals the relevance of specific molecular descriptors for each ADME property and quantifies their impact on the ADME property prediction.

Conclusion: The reported research illustrates how explainable ML models can provide detailed insights about the individual contributions of molecular features to the final prediction of an ADME property, as an effort to support experts in the process of drug candidate selection through a better understanding of the impact of molecular features.

Keywords: ADME properties, Explainable machine learning, Molecular descriptors, Drug design

Introduction

The analysis of Absorption, Distribution, Metabolism, and Excretion (ADME) properties is of great interest in early drug design as they directly determine the drug's effectiveness at its target location. Over the last decades, significant progress has been made in developing machine learning (ML)-based predictive models for quantitative structure-activity relationship (QSAR) [1, 2], in general with important contributions to ADME property prediction [3–5].

The availability of publicly accessible experimental data is paramount for the advances in such ML-based QSAR prediction for ADME properties [6–9]. From an ML point of view, the prediction of chemical properties can be accomplished with either conventional methods, which use a fixed-size feature representation, usually calculated from molecular descriptors, or, alternatively, by relying on graph-based models, such as graph convolutional neural network (CNN) models [10], or message-passing networks [11], which implement a graph-based representation of the molecule.

Such complex, often non-linear, ML models are popular as they outperform simpler models in terms of predictive power [12, 13]. Nevertheless, they are often characterized as black-boxes, as the complexity of the underlying algorithms and functional representations does not provide a human-understandable explanation of the reasoning behind the model. This is a strong limitation when the interpretability of the model is a requirement, for example when experts are interested in exploring the feature space for determining feature importance [14]. To alleviate this shortcoming, much research on Explainable Artificial Intelligence (XAI) [15, 16] has been conducted in recent years, aiming to add the ability to get human-understandable explanations of the model's reasoning; that is, making black-box models more transparent [17, 18]. Amongst the most popular explainable ML approaches, we find Local Interpretable Model-agnostic Explanations (LIME) [19] and SHapley Additive exPlanations (SHAP) [20]. They provide *post-hoc* explanations of the predictions of the model, either for the model in general (global level), or for an observation in particular (individual level). Partial Dependence Plots (PDP) [21] examine the marginal effect of a variable on the predictions of the model. SHAP analysis is a common approach for discovering feature relevance from ML models in different domains [22–24].

In this study, we focus on the investigation of feature impact for the prediction of several ADME properties with ML models. The study is carried out on a recently published data set by Fang et al. [7]. This study offers curated data sets tailored for predicting six internal in vitro ADME endpoints based on the calculation of a set of physicochemical descriptors. We follow an explainable ML approach by first quantifying feature importance, thus identifying the most important features, and then investigating the feature impact of those most relevant features on the prediction of the ML model. The applied method is based on the use of SHAP explanations and dependence plots analysis from the best-performing ML model in each case.

Materials section describes the data set under study, while Methods section elaborates on the explainable ML approach and Results and Discussion section respectively present and discuss the feature relevance results for ADME properties prediction.

Materials

The ADME properties define the drug's pharmacokinetic profile on-site. The data under study are part of a public ADME properties data set released by Fang et al. [7] that includes important information to measure the effectiveness of drugs in cells. This publicly available data set incorporates 3,521 non-proprietary small-molecule compounds selected from different available compound libraries, mainly from *eMolecules*[1], but also from the *ChEMBL*, *Enamine*,*WuXi LabNetwork* and *Mcule* database. The public data set comprises data for six ADME in vitro endpoints, which are described using 316 molecular descriptors calculated from the RDKit library.

[1] https://search.emolecules.com

Table 1 Main statistics of ADME activity values in logarithmic scale. The abbreviatons *N*, *25%*, *50%*, *75%* stand for number of compounds, 25^{th}, 50^{th}, and 75^{th} percentiles

	Sol	**MDR1**	**rPPB**	**hPPB**	**RLM**	**HLM**
N	2,173	2,642	885	1,808	3,054	3,087
Mean	1.267	0.399	0.715	0.636	2.253	1.309
Std	0.668	0.681	0.750	0.719	0.751	0.625
Min	-1	-1.046	-1.403	-1.593	0	0
25%	1.130	-0.157	0.211	0.164	1.691	0.675
50%	1.546	0.156	0.781	0.659	2.320	1.183
75%	1.681	0.903	1.278	1.170	2.831	1.802
Max	2.179	2.725	2	2	3.969	3.339

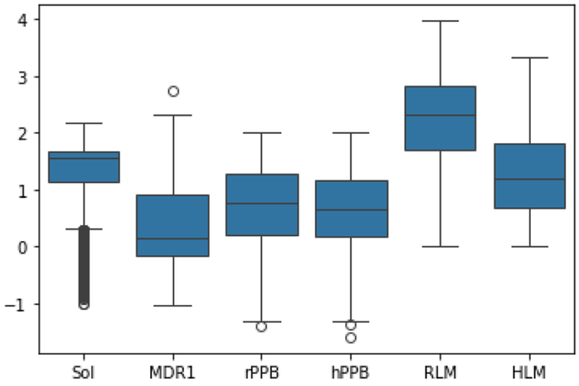


Fig. 1 Boxplot representation of the distribution of activity values for each ADME endpoint in logarithmic scale. The box represents the first and third quartile, the whiskers extend to the minimum and maximum within 1.5 times the interquartile range and outliers are shown as circles

The in vitro endpoints include human and rat liver microsomal (HLM/RLM) stability reported as intrinsic clearance expressed in mL/min/kg, human and rat plasma protein binding (hPPB/rPPB) values expressed as percent unbound, solubility at pH 6.8 (Sol) expressed in ug/mL. The MDR1-MDCK efflux ratio (MDR1-MDCK ER) is expressed as the B-A/A-B ratio, i.e. the ratio between the basolateral-to-apical permeability and vice-versa. Activity values are provided in logarithmic scale. According to the data set curators [7], the public ADME data set shows a structural diversity of compounds in terms of the number of scaffolds and singletons and is rich in experimental observations by covering a large range of experimental values for all six in vitro ADME endpoints. Table 1 and the boxplots in Fig. 1 show, in turn, the main statistics and the distribution of values for each ADME endpoint. Evaluating the compounds of the public ADME dataset from a point of view of general drug likeliness, the compounds meet the requirements of a set of known relevant physicochemical properties. According to Ghose's rule for drug likeliness [25], the molecular weight should be in range from 180 to 480 Da, the partition coefficient descriptor (logP) should have values in the range from -0.4 to 5.6, the molar refractivity in the range from 40 to 130 $m^3 mol^{-1}$, and, according to Veber's rule [26], the topological surface should have values no greater than 140 Å^2. Figure 2 shows that the distribution of

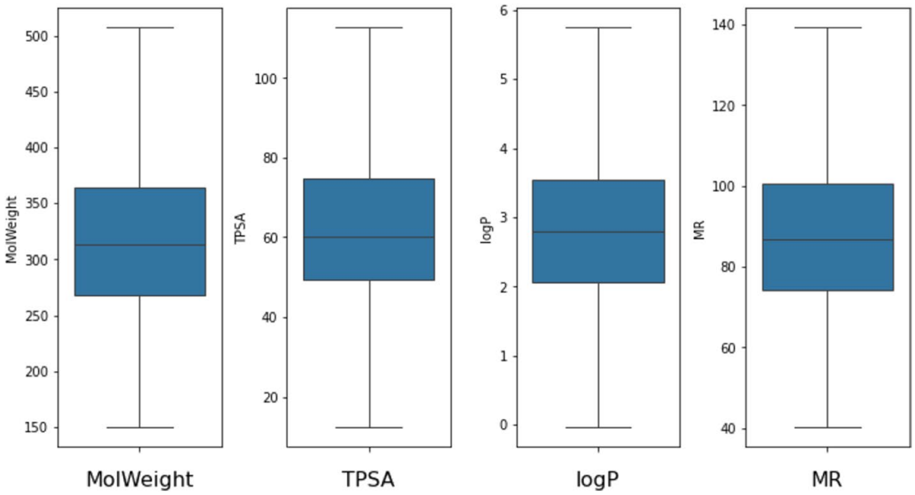


Fig. 2 Boxplot representation of the distribution of molecular weight (MolWeight), topological polar surface area (TPSA), partition coefficient (logP), and molar refractivity (MR). The box represents the first and third quartile and the whiskers extend to the minimum and maximum within 1.5 times the interquartile range

values of the 3,521 compounds of the data set is mainly inside the specified ranges for drug likeliness for these physicochemical properties.

The molecular representations used in this study comprise the 316 2D topological descriptors calculated from the RDKit library[2]. The information about the 1024-bit structural fingerprint descriptor is disregarded, as fingerprints need to be analyzed entirely and a division into sub-elements is meaningless. The data set provides training and test sets for each ADME endpoint comprising, in turn 80% and 20% of observations.

Methods

The relevance of individual features for the prediction of the ADME activity is based on predictive models. Several regression models were trained and the best performing one was used for the feature relevance analysis. More in detail, regression trees (RT), Nearest Neighbor regressors (NN), Random Forest regressor (RF) [27] and LightGBM [28], an efficient variant of Gradient Boosting machines [21], were trained for each ADME data set. RF and LightGBM are ensemble models, which use a set of *weak* classifiers to construct a stronger model. While RF uses a *bagging approach* to combine weak classifiers, LightGBM uses a *boosting* approach by sequentially adapting weak classifiers to build an improved model [29].

The training and test data set provided in [7] were used in our experiments to build and evaluate the model. The regression models were trained using 5-fold cross-validation on the training set. Results are reported from the prediction error on the test set measured by the mean squared error (*MSE*) and, additionaly, the average Pearson correlation coefficient (*Pearson's r value*) for comparability with the results of the original study by the authors of the data set.

[2] https://www.rdkit.org/

For ensemble models (RF and LightGBM), feature importance is calculated from random feature permutation [30], based on the reduction in prediction accuracy elicited by the permutation. The most influential features in predicting each ADME property, are then shown in a feature relative relevance plot.

Feature relevance is further investigated using SHAP analysis [31], a method derived from Shapley values in game theory [32]. SHAP builds a surrogate model for the predictions of the original black-box ML model. The surrogate model aims to assess the sensitivity of each feature on the prediction of the model by representing it as Shapley additive values. Therefore, the surrogate model can break down the final prediction of the model into feature-specific contributions, the so-called Shapley additive values (abbreviated as SHAP values). In consequence, these additive explanation models can evaluate the impact of single features on the overall prediction of the model. The explanations can be either local for the prediction of a single observation or global for the model in general. In this study, we focus on the analysis at the global level as the interest of the study is the general relevance of features in the prediction of each ADME endpoint. In the following, several useful graphics of the SHAP analysis are described to illustrate the analytical approach applied in the experiments.

Beeswarm plots describe the feature importance on the prediction of the model. As an illustrative example from the results reported in Results section, the beeswarm plot in Fig. 4 describes the features' relevance according to their SHAP value in the model's prediction. These plots show the absolute impact a feature can have on the predicted value, i.e. the impact to increase or decrease the predicted value represented with either positive or negative SHAP values. For example, the partition coefficient (logP) calculated by the Crippen descriptor can at most change by 0.4 the model's prediction for the HLM activity either with positive or negative SHAP values. Hence, the topological polar surface area (TPSA) descriptor has a lower impact on the predictions of the HLM activity as it can only change the predicted value by 0.2 at most. Beeswarm plots also explain the relationship between the descriptor and the predicted value. For this, the plot uses a blue-red color scheme. In the abovementioned illustrative example, higher Crippen partition coefficient values (red-colored) cause an increase in SHAP values, while lower values (blue-colored) cause lower SHAP values. The relationship of the feature with regard to the prediction of the model can be analyzed with the marginal impact of the feature by means of dependence plots.

Again, as an illustration, Fig. 5 shows the dependence plot of the partition coefficient Crippen descriptor on the HLM activity prediction. The plot explains the model's expected predicted value is 1.3 for HLM activity (horizontal line). Additionally, the plot describes how the prediction varies (blue line) according to the Crippen descriptor feature's value. In this case, the Crippen descriptor can change the expected predicted value from 0.9 to 1.6 at most. The relationship between the descriptor and the predicted value is positive, as lower descriptor values imply lower activity and higher descriptor values increase the activity. Dependency plots describe with precision the absolute impact of the variable on the prediction. For example, the impact of the TSP descriptor is quite low as the marginal impact of the predicted feature only causes changes in it the range between 1.30 and 1.34.

Table 2 Performance of the prediction of the ADME endpoints by ML model evaluated by the MSE and Pearson's *r* value (MSE/*r*-value)

Data set	RT	RF	NN	LightGBM
HLM	0.31/0.47	0.25/0.59	0.36/0.30	**0.23**/**0.63**
RLM	0.44/0.52	0.35/0.63	0.52/0.32	**0.33**/**0.66**
hPPB	0.38/0.58	0.27/0.74	0.49/0.32	**0.25**/**0.76**
rPPB	0.38/0.57	0.30/0.67	0.55/0.21	**0.28**/**0.70**
MDR1	0.32/0.60	0.26/0.71	0.38/0.48	**0.23**/**0.74**
Sol	0.50/0.35	0.40/0.54	0.55/0.16	**0.37**/**0.59**

Best results are highlighted in bold

Results

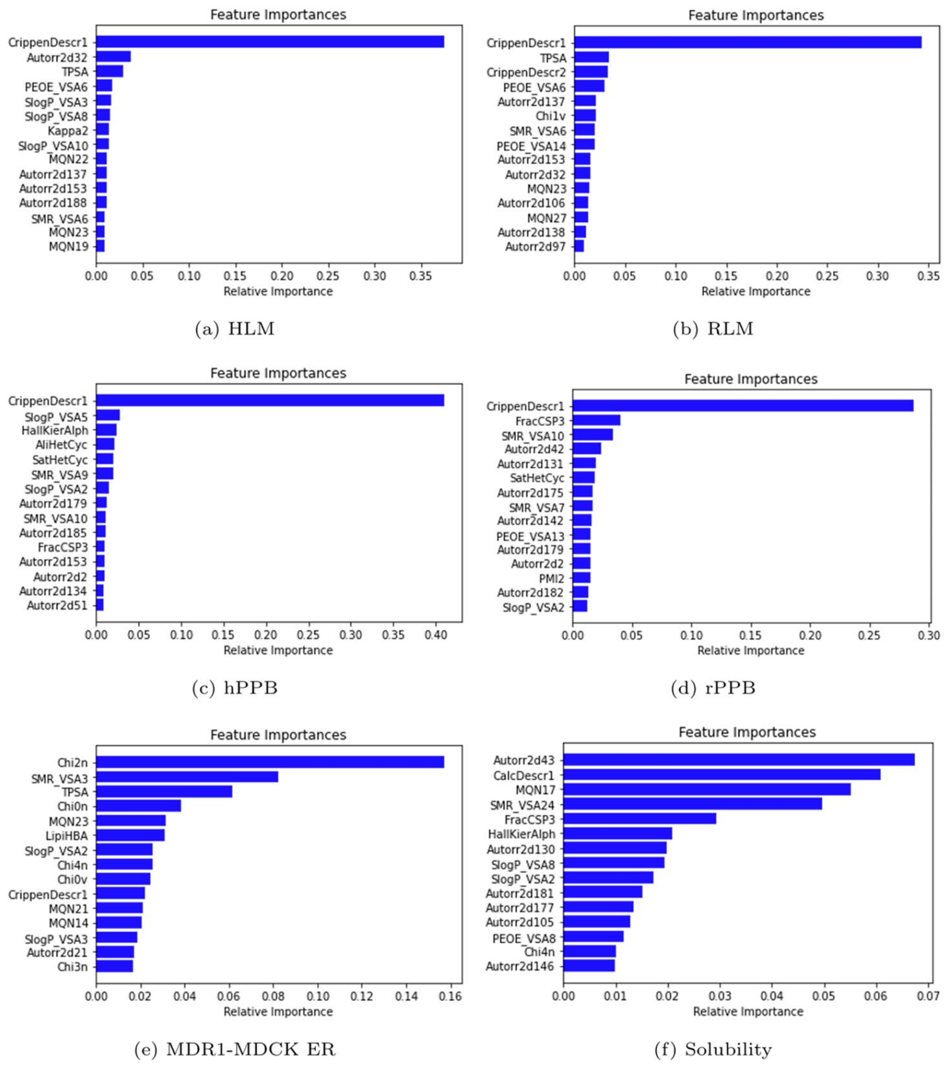
Table 2 shows the prediction performance for each ADME endpoint and for all models on the test set. Results were evaluated by the *MSE* and *Pearson's r* value. LightGBM consistently performed best for all endpoints, a result which is in line with the findings reported in [7] for the public ADME data set presented in their study. The best-performing model (LightGBM) in our study has a slightly lower performance compared with those obtained by the authors of the data set by a difference on average of 0.02 in the Pearson's correlation coefficient. This consistent variation in performance may be attributed to the difference in the data set by excluding the 1024-bit structural fingerprint, which constituted a non-interpretable feature.

Study of feature relevance

An analysis of LightGBM feature permutation reveals information about the relative relevance of features. Figure 3 shows the 15 most relevant features obtained with LightGBM for the prediction of each ADME endpoint. The relevance of features is further analyzed with SHAP analysis derived from the surrogate models built from the trained LightGBM model of each ADME property, using additive explanation models implemented in the SHAP library [20]. SHAP explanations and dependency plots are used to understand the impact of each feature on the respective model's prediction in the subsequent sections.

Human liver microsomal stability

For the prediction of human liver microsomal (HLM) stability, the most relevant features are the partition coefficient (logP) calculated by the Crippen descriptor [33] (CrippenDescr1) followed by the 2D autocorrelation coefficient and the topological polar surface area (TPSA) coefficient [34] (Figs. 3 and 4). The prediction of HLM activity depends mostly on the value of the logP Crippen descriptor and to a lesser extent on the TPSA descriptor, the 2D autocorrelation descriptor, the Partial Equalization of Orbital Electronegativity (PEOE) VSA descriptor (PEOE_VSA), the Molar Refractivity (SMR_VSA) and partition coefficient (SlogP_VSA) calculated by the Van der Waals Surface Area (VSA) descriptor [35] and the number of saturated heterocycles. According to the dependency plots in Fig. 5, the expected HLM activity of the SHAP model is 1.3 $\log_{10}$



(a) HLM

(b) RLM

(c) hPPB

(d) rPPB

(e) MDR1-MDCK ER

(f) Solubility

Fig. 3 Feature relevances for the ADME property prediction obtained from LightGBM models-based feature permutation

(mL/min/kg). The logP Crippen descriptor is the most relevant feature as its impact is a variation of the HLM activity in the range from 1.0 to 1.6. Higher values of the Crippen partition coefficient descriptor increase HLM activity and higher TPSA values lower the HLM activity according to the dependency plots.

Rat liver microsomal stability

For RML stability, the partition coefficient (logP) calculated by the Crippen descriptor (CrippenDescr1) is most relevant, while the TPSA descriptor, the molar refractivity (MR) calculated by the VSA descriptor (SMR_VSA), and the Partial Equalization of Orbital Electronegativity (PEOE) VSA descriptor are relevant to a lesser extent (Figs. 6 and 7). The mean expected RLM activity is 2.25 $\log_{10}$(mL/min/kg). The logP Crippen descriptor is the feature with the highest impact as it may change the activity in a range from 1.7 to 2.5, while TPSA, SMR_VSA and the autocorrelation descriptor have

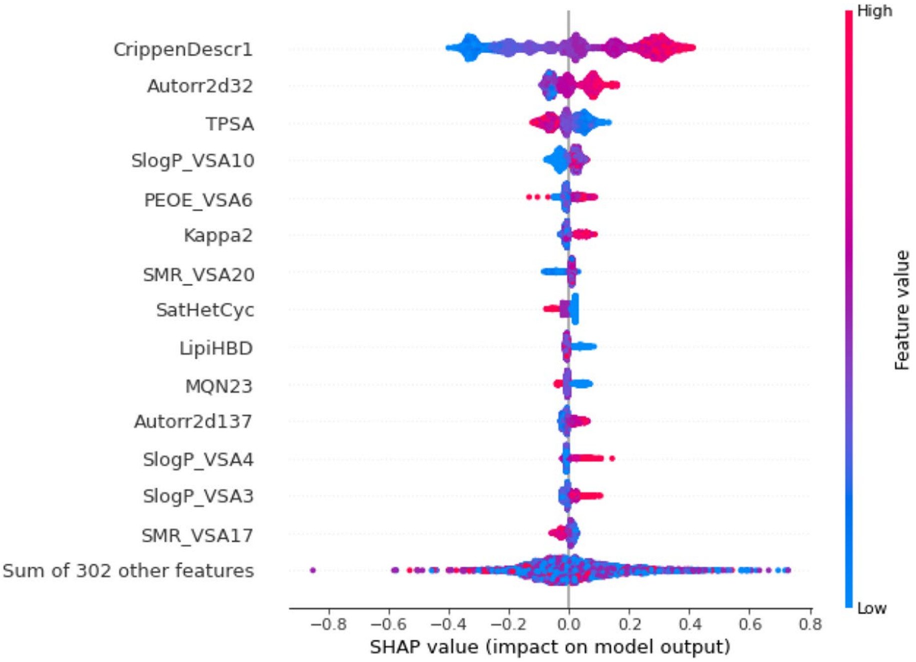
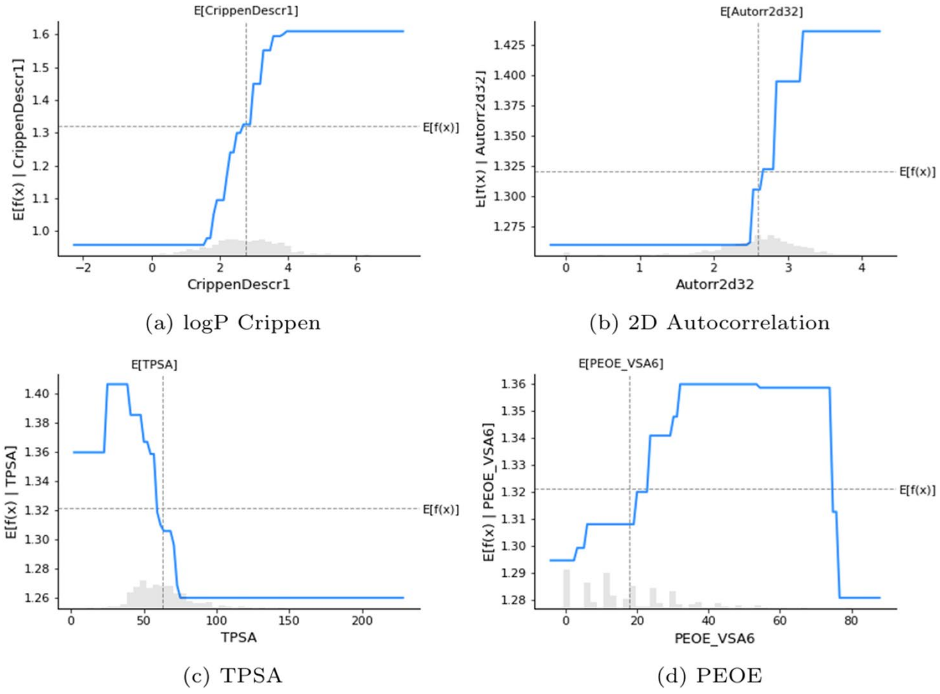


Fig. 4 Mean SHAP values by features for HLM prediction



(a) logP Crippen

(b) 2D Autocorrelation

(c) TPSA

(d) PEOE

Fig. 5 Dependence plot for HLM activity prediction

a substantially lower individual impact. The dependency plots of the most relevant features (Fig. 7) describe the positive relationship of the logP Crippen description and the SMR_VSA descriptor on the prediction of RLM activity, while the TPSA descriptor has a negative relationship with RLM activity.

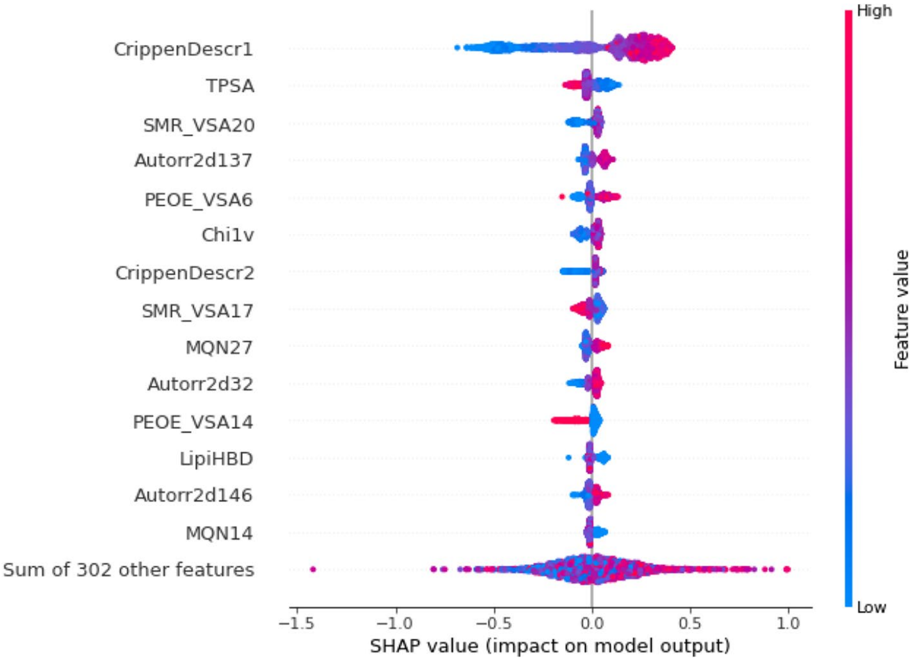
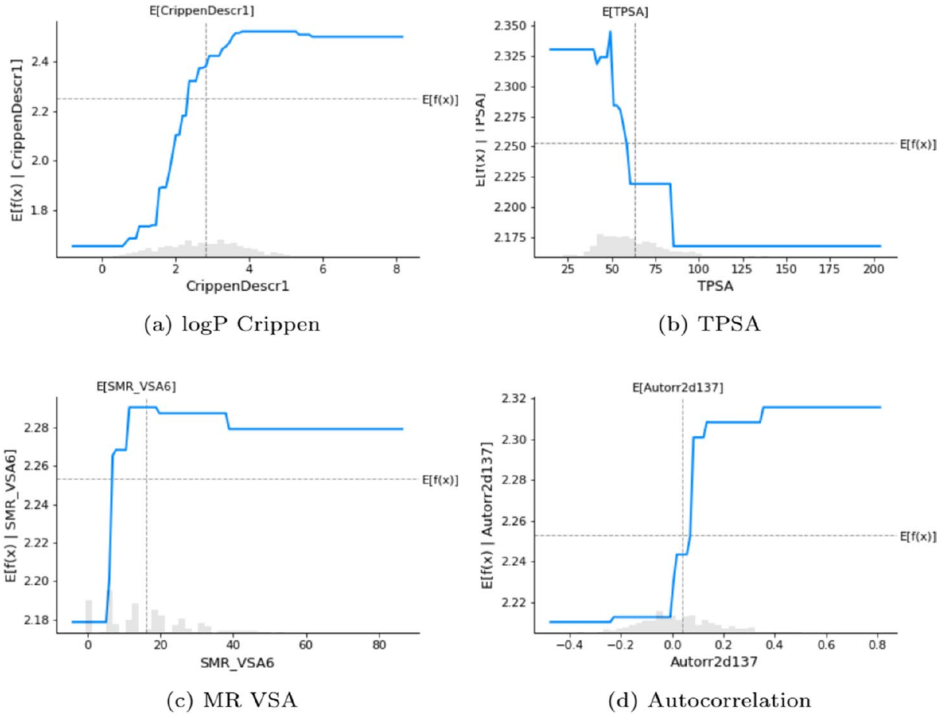


Fig. 6 SHAP values mean contribution of features to RLM prediction



(a) logP Crippen

(b) TPSA

(c) MR VSA

(d) Autocorrelation

Fig. 7 Dependence plot for RLM activity predictions

Human plasma protein binding

For the prediction of human plasma protein binding (hPPB), the most relevant feature is the partition coefficient of the Crippen descriptor. To a lower extent, the logP

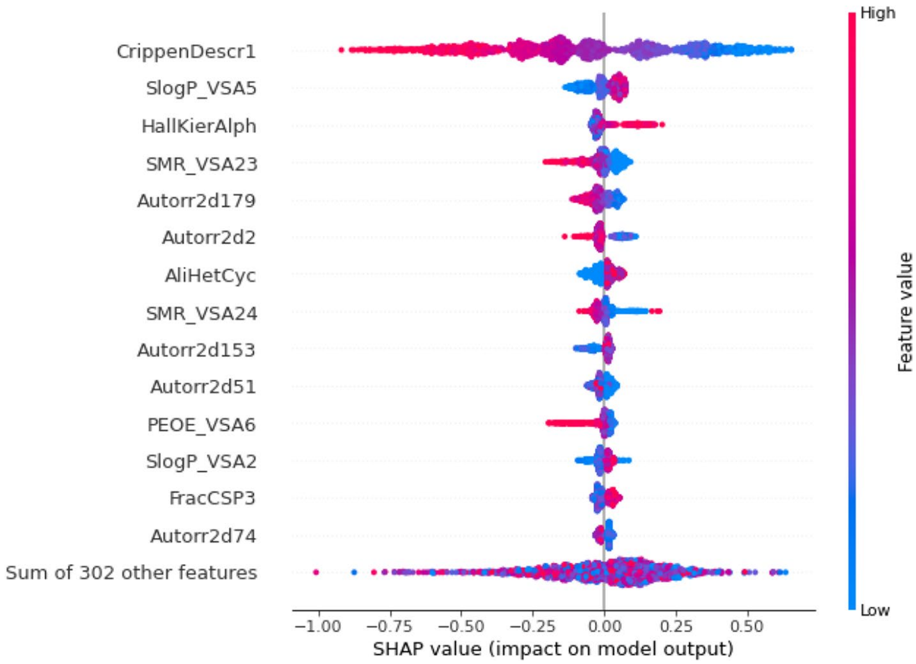
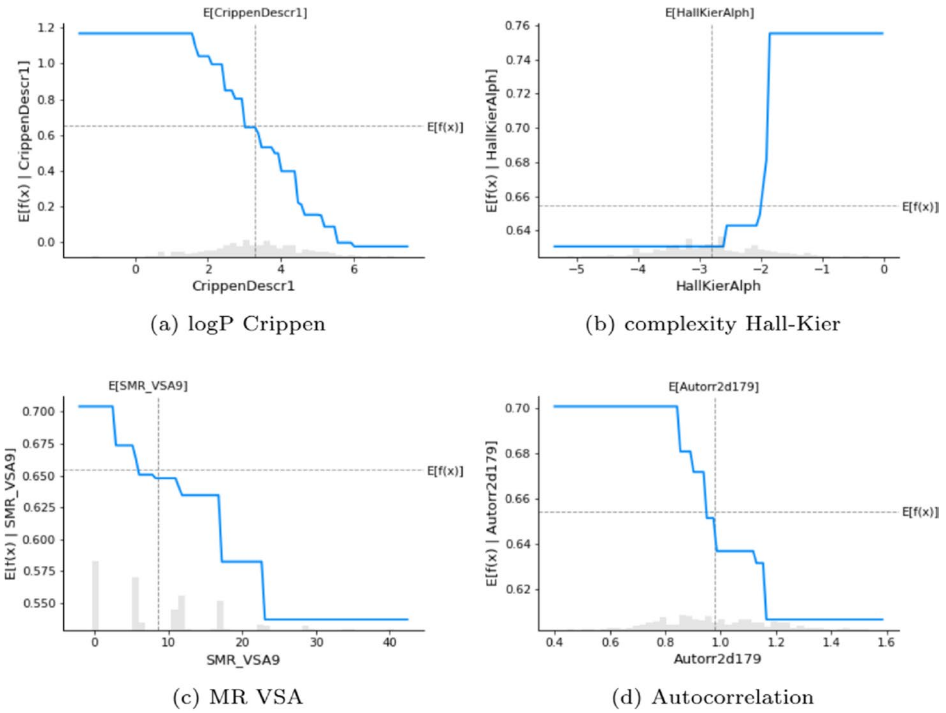


Fig. 8 SHAP values mean contribution of features to hPPB prediction



(a) logP Crippen

(b) complexity Hall-Kier

(c) MR VSA

(d) Autocorrelation

Fig. 9 Dependence plot for hPPB activity prediction

VSA descriptor, the topological HallKier descriptor [36] as an index of molecular complexity, the MR VSA descriptor, the autocorrelation descriptor and the number of aliphatic heterocycles are also relevant (Figs. 8 and 9). According to the

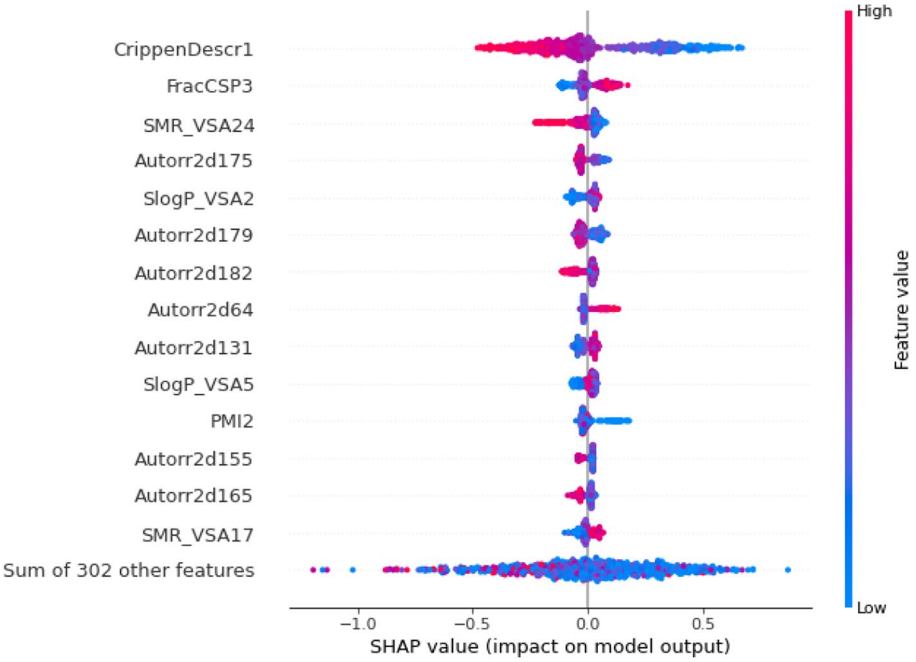


Fig. 10 SHAP values mean contribution of features to rPPB prediction

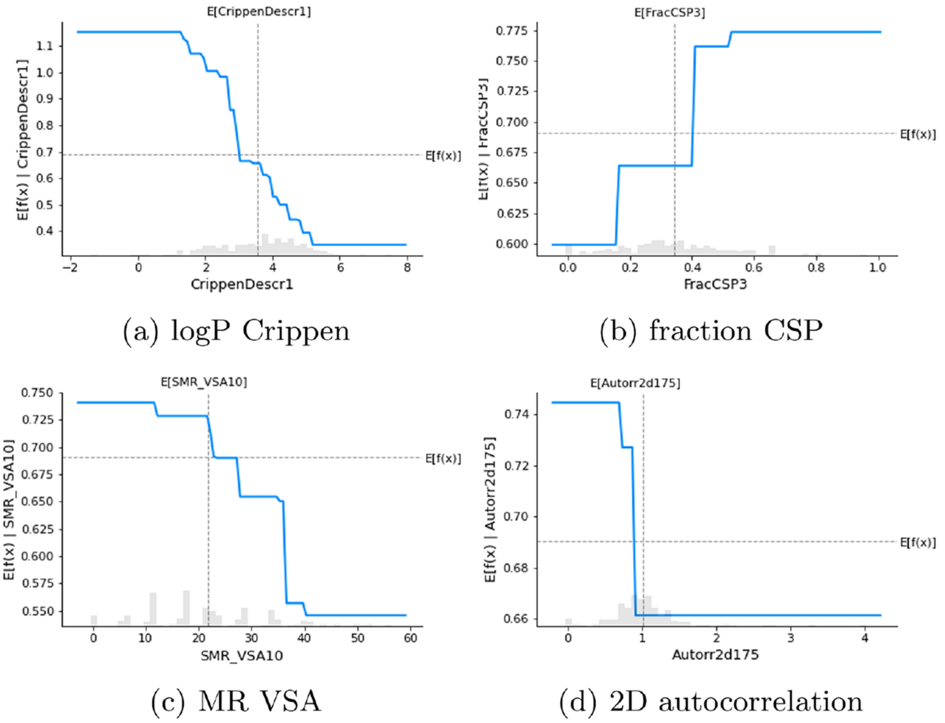
SHAP model, the mean expected activity value is 0.65 measured as $\log_{10}$ of percent unbounds. The logP Crippen descriptor has the highest impact on the prediction varying the hPPB prediction in a range from 0 to 1.2 (Fig. 9). For the other relevant features, the absolute value of the impact is substantially lower. The Hall-Kier descriptor has a positive relationship with hPPB activity, while the logP Crippen descriptor and the ML VSA descriptor has a negative relationship with the prediction of the hPPB activity.

Rat plasma protein binding

For rPPB the partition coefficient calculated by the Crippen descriptor is the most relevant feature. The fraction of SP3-Hybridized Carbon Atoms descriptor (frac_CSP3), the MR VSA descriptor (SMR_VSA), and the 2D autocorrelation coefficient are also relevant but to a lower extent (Figs. 10 and 11). The mean expected rPPB value is approximately 0.68, measured as $\log_{10}$ of percent unbounds. According to the dependency plots in Fig. 11, the impact of the logP Crippen descriptor on the activity prediction is high by varying the target value in the range of 0.4 to 1.2 approximately. The other relevant features have a much lower impact. The logP Crippen descriptor and the MR SVA descriptor have a negative relationship with the prediction of the activity. Hence, the fraction of CSP descriptor has a positive relationship on the activity prediction.

MDR1-MDCK efflux ratio

There are several relevant features for the prediction of the MDR1-MDCK efflux ratio (MDR1-MDCK ER), namely the MR VSA descriptor, the topological Chi descriptor for the quantification of the molecule's complexity, the molecule quantum number (MQN)



(a) logP Crippen

(b) fraction CSP

(c) MR VSA

(d) 2D autocorrelation

Fig. 11 Dependence plot for rPPB activity prediction

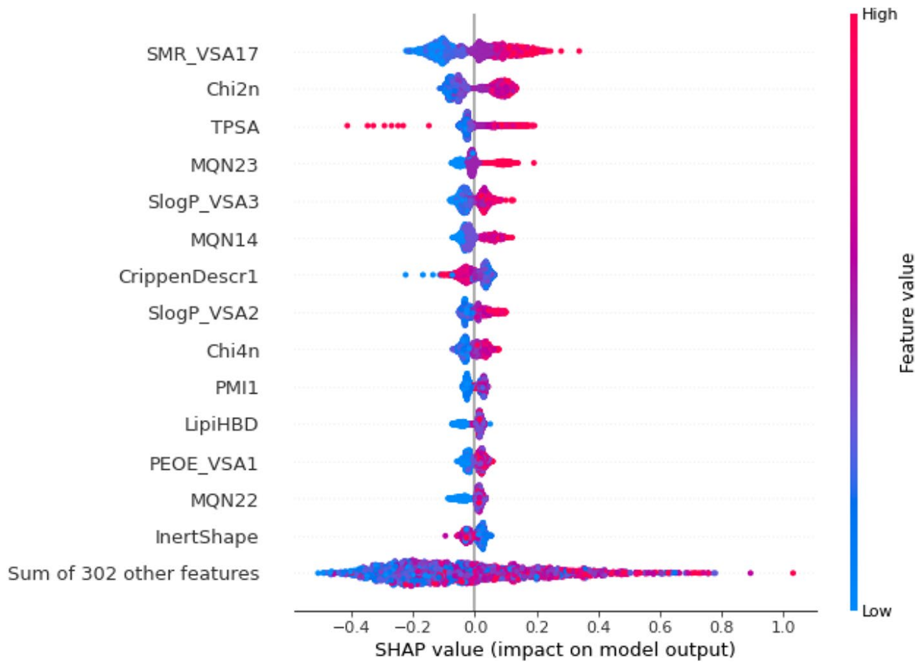
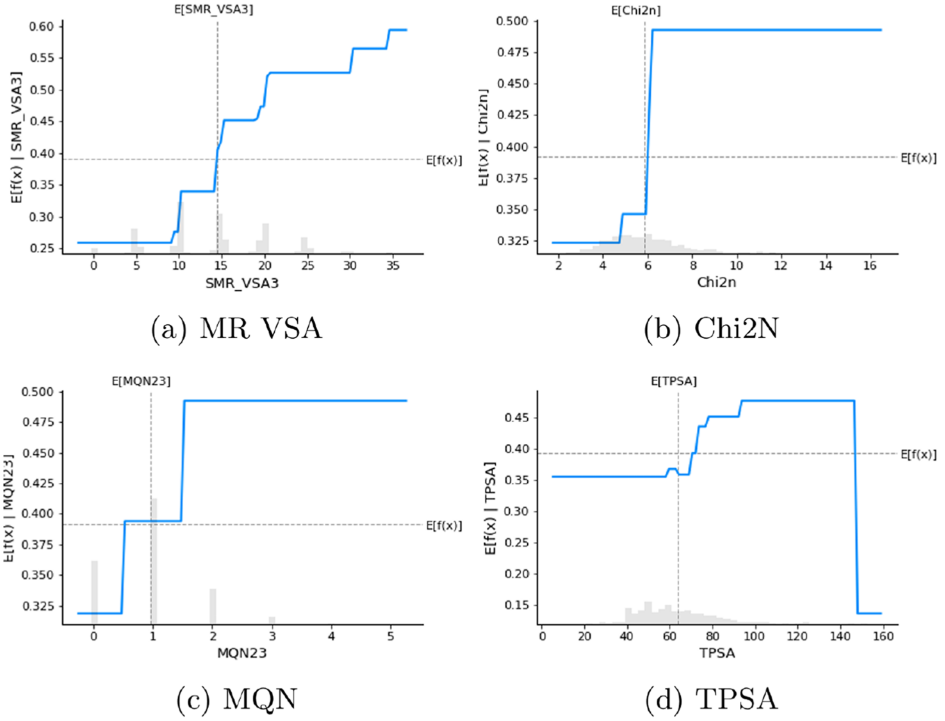


Fig. 12 SHAP values mean contribution of features to MDR1 prediction

descriptor with information about the molecule's structure and properties, the topological TPSA descriptor, the partition coefficient VSA descriptor (SlogP_VSA) and Crippen descriptor (Figs. 12 and 13).



(a) MR VSA

(b) Chi2N

(c) MQN

(d) TPSA

Fig. 13 Dependence plot for MDR1 ER activity prediction

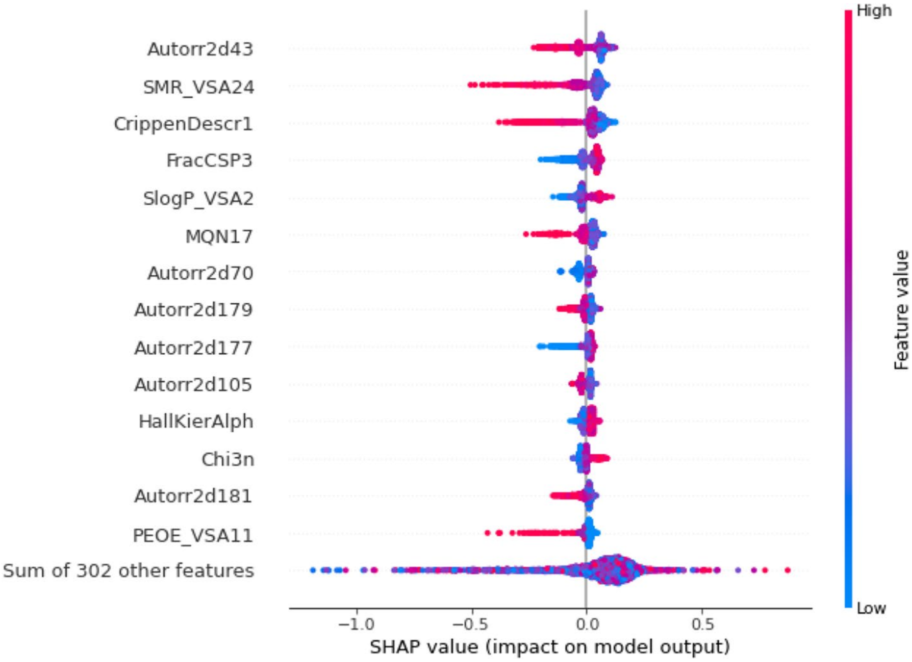
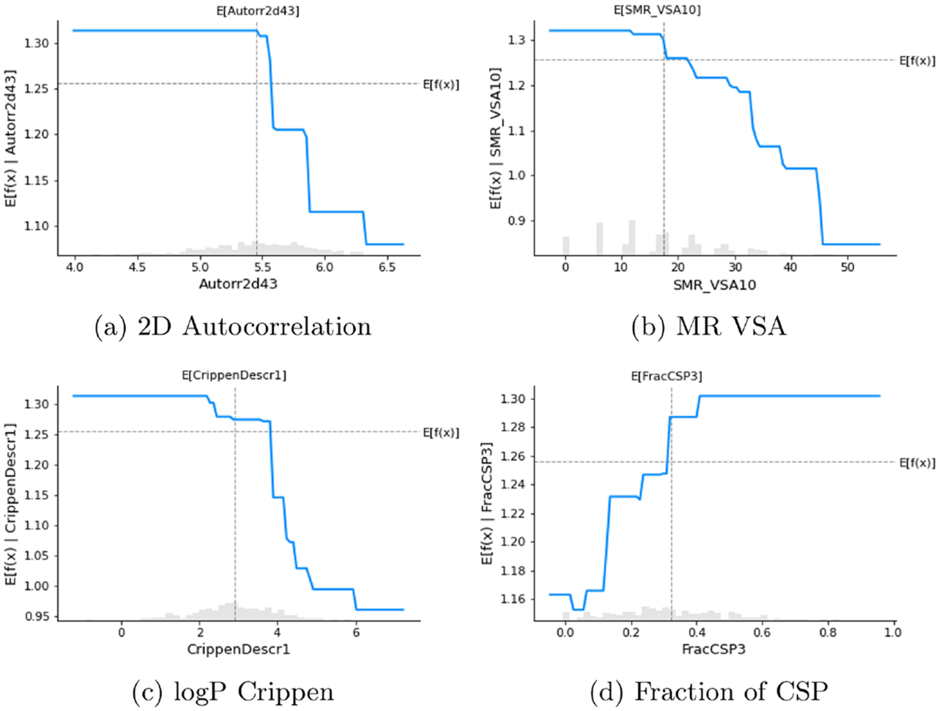


Fig. 14 SHAP values mean contribution of features to solubility prediction

The mean expected MDR1-MDCK activity for the efflux ratio is 0.4 (measured as $\log_{10}$ (B-A/A-B) ratio). The MR VSA descriptor has a high impact varying the target value from 0.25 to 0.6. In this case, the other three most relevant features have a similar impact



(a) 2D Autocorrelation

(b) MR VSA

(c) logP Crippen

(d) Fraction of CSP

Fig. 15 Dependence plot for solubility activity prediction

on the model's prediction, as the prediction varies in a similar range of 0.3 to 0.5 approximately (Fig. 13). As described in the dependency plots, all features have a positive relationship with MDR1 activity.

Solubility

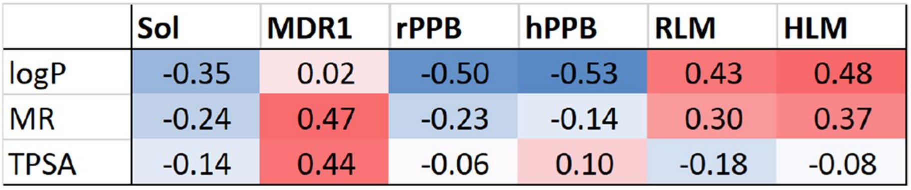
For the prediction of solubility, there are several relevant features, namely the 2D autocorrelation coefficient, the MR VSA descriptor, the partition coefficient calculated by the Crippen descriptor or VSA descriptor, the fraction of CSP descriptor, and the molecule quantum number (MQN) descriptor (Figs. 14 and 15). The mean expected solubility is 1.25 $\log_{10}$(ug/mL). The impact of the most relevant features is similar, as described in the respective dependency plots (Fig. 15). The autocorrelation descriptor, the MR VSA descriptor, and the logP Crippen descriptor have a negative relationship concerning solubility, while the fraction of CSP descriptor has a positive relationship (Fig. 15).

Summary of the feature relevance study

Table 3 provides an overview of the most relevant features for the prediction of the ADME endpoints under study derived from the SHAP analysis of the surrogate models built on the trained LightGBM model for each ADME property. The partition coefficient (logP Crippen), the autocorrelation descriptor, and the MR VSA descriptor are found to be relevant to all ADME predictions. Interestingly, there is a set of common features for both HLM/RLM prediction as well as for hPPB/rPPB prediction. The TPSA and MQN descriptors are highly relevant for both HLM/RLM prediction, but not for the hPPB/rPPB prediction. Hence, the Fraction CSP descriptor and saturated heterocycle

Table 3 Summary of most relevant features for ADME prediction

	HLM	RLM	hPPB	rPPB	MDR1	Sol
logP Crippen	x	x	x	x	x	x
TPSA	x	x			x	
MR Crippen		x				
Autocorr.	x	x	x	x	x	x
PEOE_VSA	x	x	x		x	x
logP VSA	x		x	x	x	x
MR_VSA	x	x	x	x	x	x
Kappa	x					
MQN	x	x			x	x
Chi		x			x	
Hall-Kier			x			x
Aliph. HC			x			
Sat. HC			x	x		
Fraction CSP			x	x		x
PMI				x		



	Sol	MDR1	rPPB	hPPB	RLM	HLM
logP	-0.35	0.02	-0.50	-0.53	0.43	0.48
MR	-0.24	0.47	-0.23	-0.14	0.30	0.37
TPSA	-0.14	0.44	-0.06	0.10	-0.18	-0.08

Fig. 16 Pearson's correlation coefficient of descriptors with ADME activity

descriptor is relevant for the hPPB/rPPB prediction, whereas these descriptors are not between the relevant ones for the HLM/RLM prediction. The partial charge descriptor (PEOE_VSA) is found to be relevant for several ADME properties.

Furthermore, the analysis of relevance has highlighted different composite descriptors (Chi, MQN, Autocorrelation, VSA or MQN) comprising a larger number of scores with detailed information about the molecule's property under study. This result suggests the inclusion of these descriptors into the feature set comprising all their subscores.

The findings about the SHAP explanations at the global level described in the previous sections are compared with statistical correlations between variables. The correlation coefficients calculated by Pearson's correlation (Fig. 16) confirm a moderate positive and negative relationship of the logP Crippen descriptor with the HLM/RLM endpoints and the rPPB/hPPB endpoints. The MR Crippen descriptor has a moderate positive correlation with the MDR1 activity values. The correlation of this descriptor with the activity value of the RLM/HLM endpoints is positive and for the rPPB/hPPB prediction, there is a weak negative correlation. TPSA shows only in the case of the MDR1 activity a moderate positive correlation.

Model optimization

To evaluate whether the most relevant features suffice to predict the ADME property properly, feature selection was performed to evaluate the predictiveness of the models

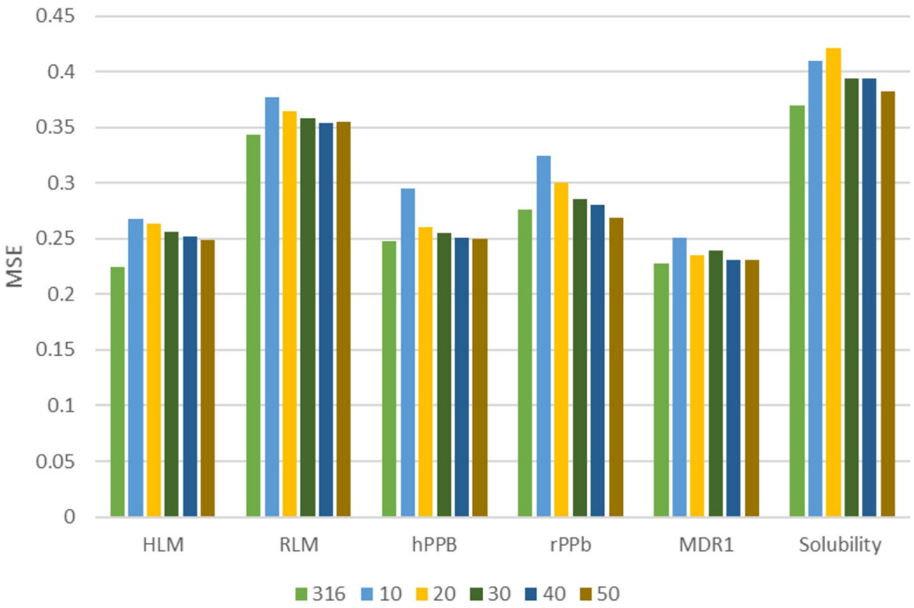


Fig. 17 MSE with LightGBM trained on reduced feature sets

Table 4 MSE of LightGBM for ADME endpoints on reduced feature sets

Data set	316	10	20	30	40	50
HLM	**0.23**	0.268	0.263	0.256	0.252	0.249
RLM	**0.33**	0.377	0.364	0.358	0.354	0.355
hPPB	0.25	0.295	0.260	0.255	0.251	**0.250**
rPPb	0.28	0.324	0.300	0.286	0.280	**0.269**
MDR1	0.23	0.251	0.235	0.239	0.231	**0.231**
Solubility	**0.37**	0.410	0.421	0.394	0.394	0.380

Best results are highlighted in bold

built on reduced feature sets. For each ADME prediction model, a subset of features was selected according to feature relevance. Figure 17 shows the prediction accuracy of the LightGBM model trained on different-sized reduced feature sets. The models are trained on the complete feature set (316 descriptors) and with different-sized reduced feature sets including between 10 and 50 features. The comparison of the performance of the models trained using different-sized feature sets shows that nearly identical results can be obtained with a reduced feature set (Fig. 17). Interestingly, in the case of hPPB, rPPB and MDR1-ER prediction, feature selection allowed to actually improve the performance of the model, as highlighted in Table 4.

Prediction analysis for molecules

While the former sections focused on the analysis of SHAP explanations at the global level, i.e. for the mean prediction of the model to discover the most relevant features, in this section, the predictions are analyzed at the individual level for single molecules. The SHAP explanations are obtained from the surrogate models built from the LightGBM model for each ADME property. The individual Shapley-based explanations for all compounds of the data set are available at the GitHub repository of the project. In the following, examples from the HLM, MDR1 ER, and Solubility data sets are analyzed to illustrate the explanations of the predictions for individual molecules. Molecules are compared by calculating the Tanimoto similarity between the extended-connectivity fingerprint ECFP4 representation [37], namely a 2,048-bit long Morgan fingerprint [38] assuming a maximum path of length seven.

HLM data set

Human liver microsomal intrinsic clearance influences the bio-availability and the half-life of a drug, which are important for the dosing regimen of a drug [39]. Depending on the compound and the therapeutic level a higher or lower intrinsic clearance is desirable. Figure 18 shows three examples of explanations for HLM activity prediction at the individual level. These three molecules have an intrinsic clearance of 0.84, 0.79, and 1.64 respectively measured as $\log_{10}$(mL/min/kg). The similarity between the molecules is lower than 0.25 according to the pair-wise Tanimoto coefficient.

Molecules ID 177444153 and ID 49964398 are predicted to have a similar intrinsic clearance of 6.9 and 6.16 (mL/min/kg) respectively. Nevertheless, due to their structural differences, their molecular features are quite different. The partition coefficient found as the most relevant feature specifically has very different values. As a consequence, the contributions of the SHAP values differ in the respective predictions. The previous case is an example where structurally different molecules yield similar HLM activities, and SHAP values explain in detail the contributions of the molecular features in the prediction. The third molecule with ID 53827576 has a high HLM activity of 1.46 $\log_{10}$ (mL/min/kg) equivalent to 28.84 (mL/min/kg), which indicates efficient metabolism to remove the drug. The breakdown of the prediction into individual contributions provides helpful insights about the logic of the predicted HLM activity. Although the impact of the Crippen partition coefficient is negative in terms of SHAP values, other less relevant molecular features have positive contributions.

MDR1 ER data set

The MDR1-MDCK efflux ratio evaluates if a compound is a substrate of the P-glycoprotein (P-gp) efflux transporter, which relates to the intestinal absorption of drugs and the permeability to the central nervous system (CNS) [40]. In general, a low MDR1-MDCK efflux ratio indicates minimal efflux by P-gp and makes the compounds candidates to cross the blood-brain barrier and reach the central nervous system [41, 42]. Figure 19 shows three examples of explanations of the MDR1-MDCK ER predictions with values of -0.145, -0.129, and 1.28 respectively, measured as $\log_{10}$ value of the B-A/A-B efflux ratio. The similarity between the molecules is below 0.29 according to the pair-wise Tanimoto coefficients. Molecules with ID 00480727 and ID 5646720 have low predicted MDR1 ER, namely an efflux ratio of 1.39 and 1.34 respectively, indicating a potential for good drug absorption. The itemization of the prediction reveals the negative contributions of the most relevant features, namely the CHi2n descriptor, the MR VSA descriptor, and the MQN descriptor. The Molecule with ID 8318595, in turn, has a much higher predicted MDR1-MDCK ER of 19 indicating less optimal drug absorption

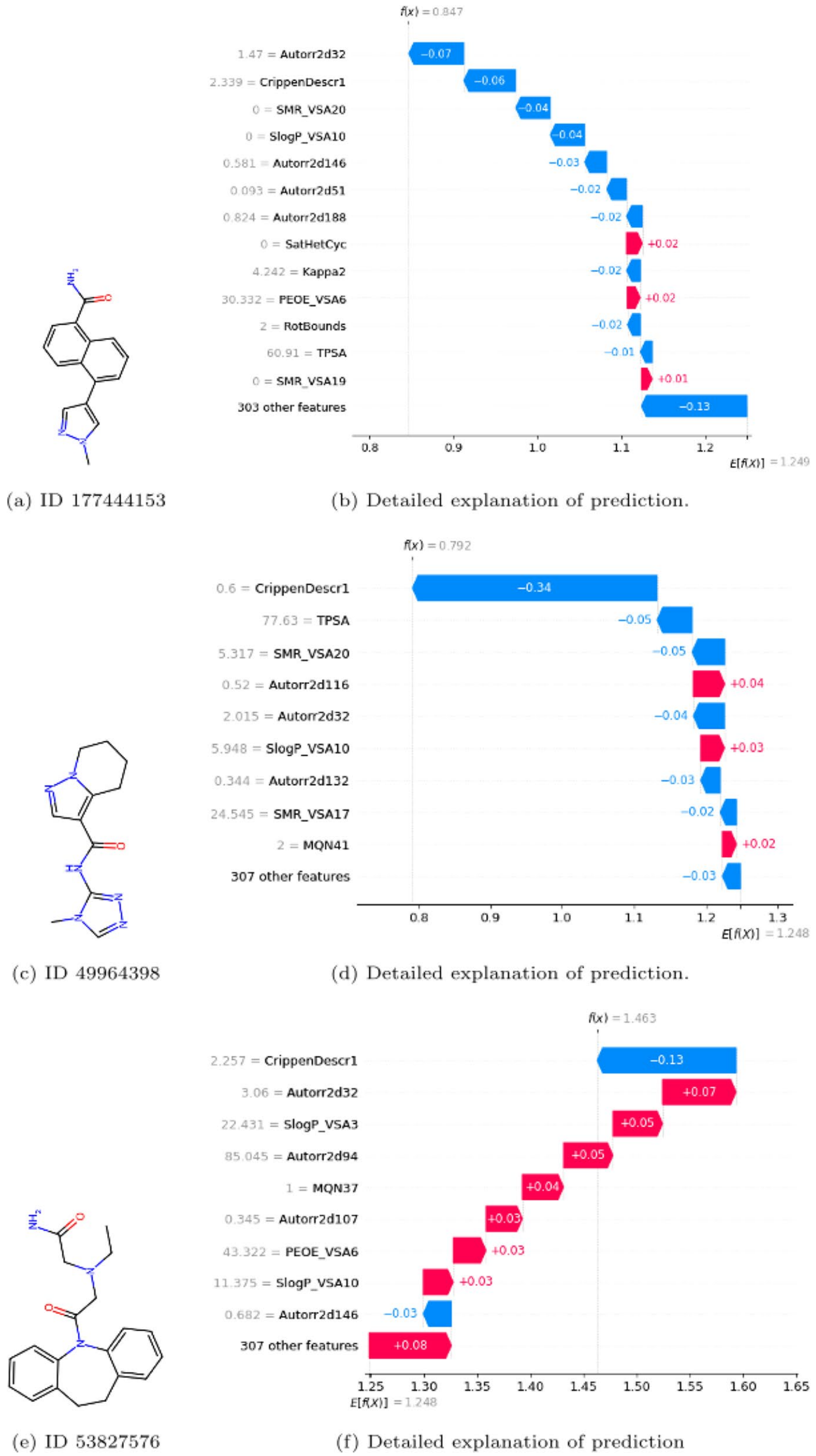
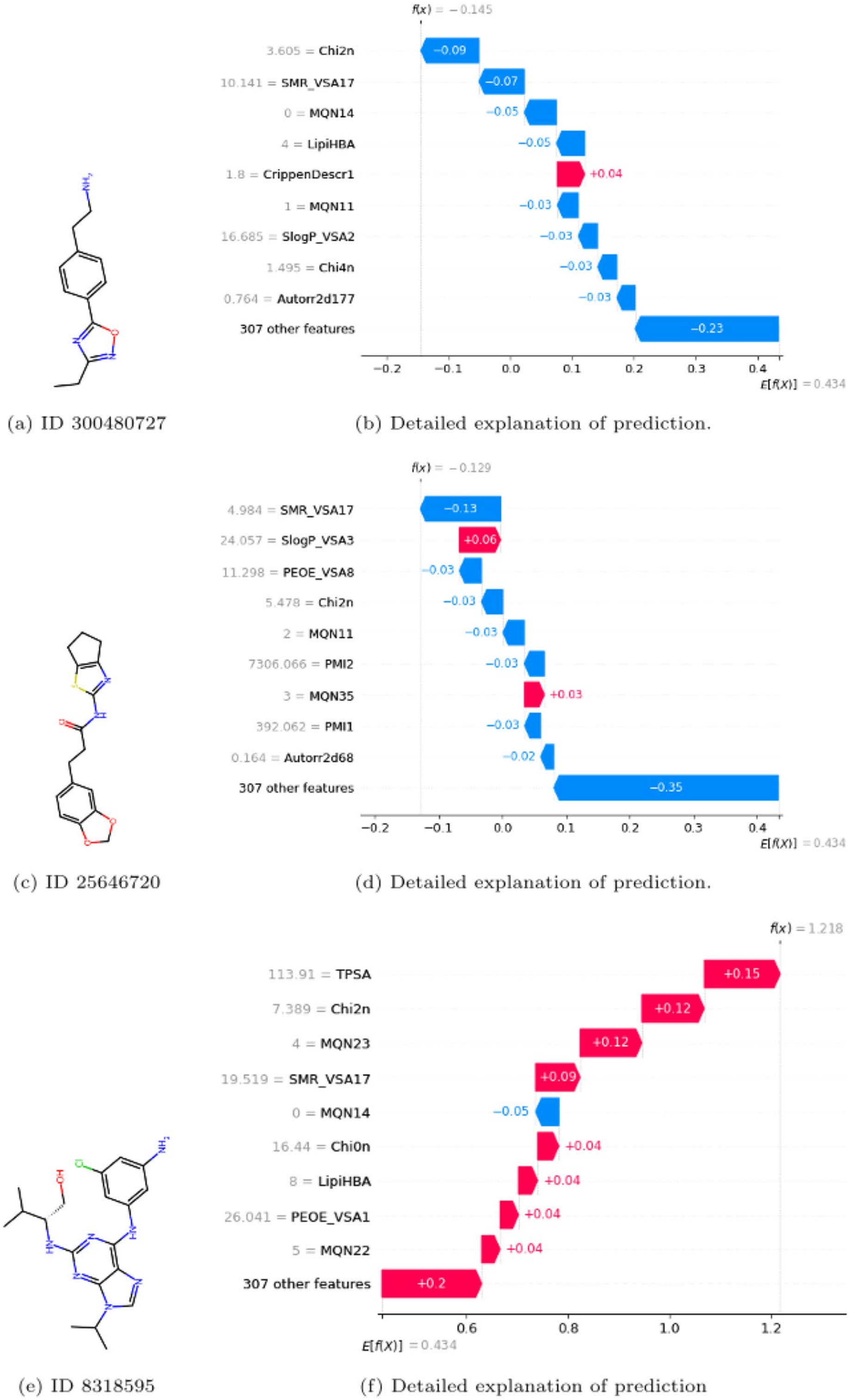


Fig. 18 Break-down of SHAP values for HLM activity prediction. Compound shown at the left and prediction at the right



(a) ID 300480727

(b) Detailed explanation of prediction.

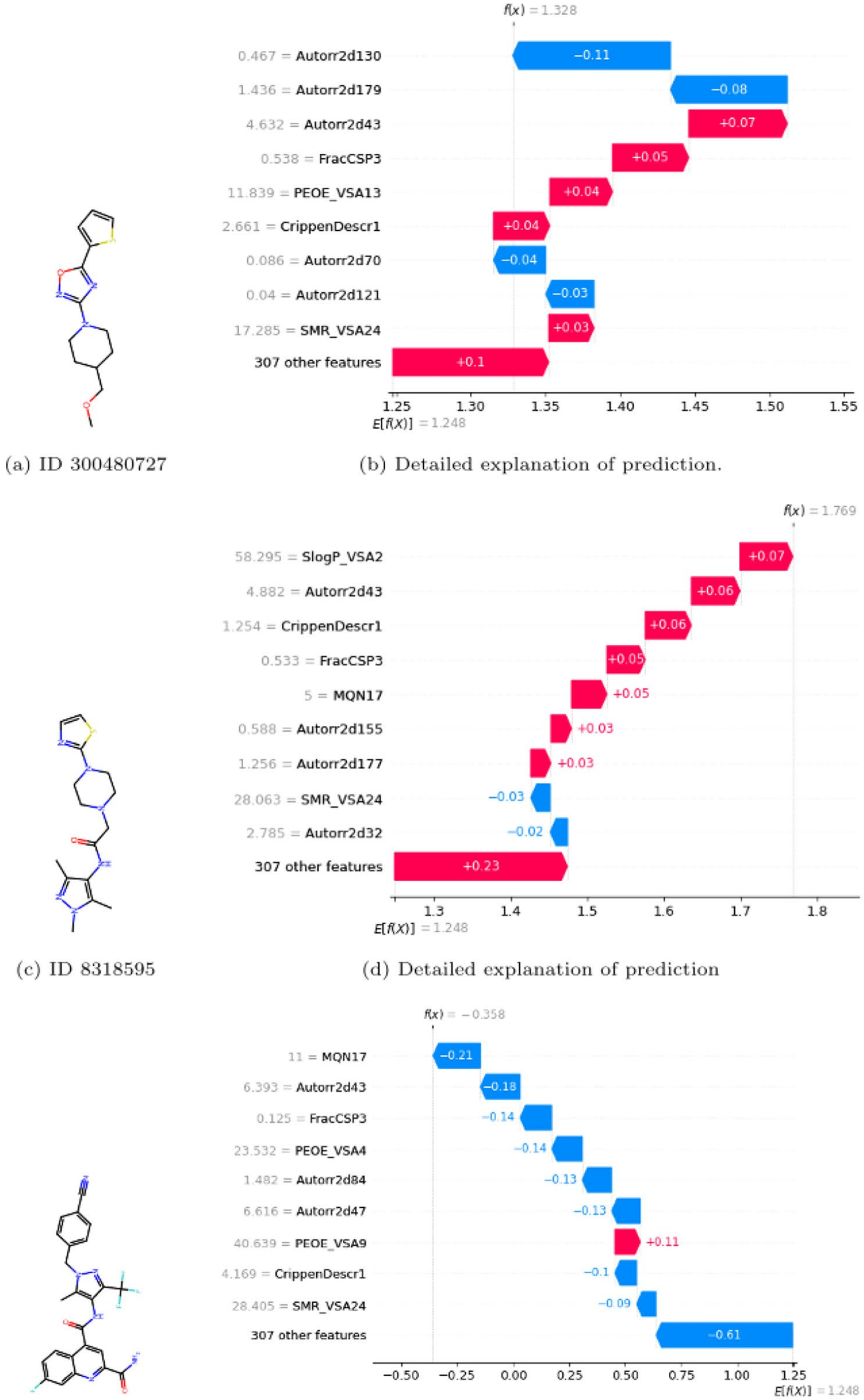
(c) ID 25646720

(d) Detailed explanation of prediction.

(e) ID 8318595

(f) Detailed explanation of prediction

Fig. 19 Break-down of SHAP values for MDR1 ER activity prediction. Compound shown at the left and prediction at the right



(a) ID 300480727

(b) Detailed explanation of prediction.

(c) ID 8318595

(d) Detailed explanation of prediction

(e) ID 25646720

(f) Detailed explanation of prediction.

Fig. 20 Break-down of SHAP values for solubility prediction. Compound shown at the left and prediction at the right

characteristics. For this molecule, the most relevant features have a positive contribution in terms of SHAP values in the explanation of the prediction.

Solubility data set

Solubility in water is an important property for oral-administrated drugs, as it has a direct impact in the capability of drug absorption in the body [43]. Figure 20 shows the explanations for the Solubility predictions for three molecules with solubility coefficients of 1.328, 1.77, and -0.36 $\log_{10}$(ug/mL), respectively. The similarity between the molecules is below 0.37 according to the pair-wise Tanimoto coefficients.

Molecules with ID 300480727 and ID 8318595 have a high predicted solubility, with solubility values of 21.28 ug/mL and 58.55 ug/mL considered as moderate/high type of solubility in other studies [43]. The itemization of the prediction into feature-related contributions reveals the positive contributions of the most relevant features, namely the autocorrelation coefficient, the logP VSA descriptor, the fraction of CSP descriptor, and the logP Crippen descriptor. The Molecule with ID 25646720 has a very low solubility of 0.46 ug/mL, representing an insoluble compound according to the minimum threshold of 10 ug/mL described by [44]. In this case, the most relevant features have a negative contribution in terms of SHAP values. The observed contributions of features in the previously explained examples agree with the impact of features outlined in the global analysis of features' impact on the predictions (Fig. 15).

Discussion

The experimental results based on the analysis of eXplainable ML models for the prediction of the different ADME endpoints revealed several insights about the most relevant descriptors for each property. First of all, less complex ML models, namely NN and RT, performed worse than more complex ensemble models (RF or LightGBM). While RT has an inherently interpretable logic for the prediction, RF and LightGBM are not straightforward to interpret, as they are an ensemble of individual models. The analysis of feature relevance based on feature permutation yielded information about the most relevant features for each ADME endpoint. The partition coefficient descriptor (logP), the molar refractivity descriptor (MR), and the topological polar surface area (TPSA) descriptor were found to be the most relevant features. These results agree with known rules to evaluate drug likeliness, such as those reported in Ghose et al. [25] and Veber et al. [26], which highlight these three physicochemical properties as relevant. Other known properties for drug likeliness such as the number of hydrogen bond donors (HBD), the number of hydrogen bond acceptors (HBA) described by the Lipinski's rule [45], or the number of rotatable bonds [26] are also relevant in the predictions. The number of hydrogen bond donors (HBD) is highlighted as a relevant feature in the SHAP analysis for the HLM, RLM, and MDR1 activity prediction. The number of hydrogen bond acceptors (HBA) and the number of rotatable bonds are often found to have a significant influence on the predictions for individual molecules. For a particular example, see the predictions illustrated in Figs. 18 or 19, or refer to the explanations of individual predictions at the GitHub repository of the project. The study by Honório et al. [46] uses several of the aforementioned descriptors to predict different ADME properties. This work describes the biological role of several physicochemical properties concerning

ADME properties. The partition coefficient (logP) represents a measure of lipophilicity and is therefore relevant for membrane permeability, absorption, distribution, and clearance. Both the polar surface area and the hydrogen bonding capability are decisive for membrane permeability. The state of ionization influences the solubility in water and membrane permeability and is related to several ADME properties. In particular, for the plasma protein binding capability, the partial charge descriptor (PEOE) was found to be an important descriptor [47]. From a structural point of view, the number of rotatable bonds determines the molecular flexibility, which in turn is related to the capability for absorption, permeability, and distribution. The information about the biological role of these physicochemical properties substantiates the use of related molecular descriptors in the respective predictive models.

SHAP analysis delved further into the relevance of these molecular descriptors measuring their absolute impact on the prediction of the ML model as well as the trend to increase or decrease each of the single ADME activities. These findings were compared with Pearson's correlation coefficients, which confirmed the either positive or negative relationship between the descriptor and ADME activity. The feature relevance study based on SHAP analysis provided useful insights about the reasoning of the predictive model by highlighting the contribution of the features and so identifying the most important features. Interpretability based on SHAP explanations was analyzed in a similar study for activity prediction [22] using structural features of the compounds, instead of physicochemical descriptor values.

The assessment of the predictiveness of these most relevant features was carried out through a study of the progressive inclusion of features, aiming to find the reduced feature set that equals its performance to that of the baseline model. Our experiments have shown that, in general, the ML models can not rely solely on the most relevant features but need to include a wider set of molecular descriptors to achieve the same performance as the baseline models built on all descriptors. A feature set including approximately 50 descriptors (15.8%) yielded the same results as the baseline model in most cases.

The analysis of predictions at the individual level illustrated well how SHapley additive explanations can be used as human-understandable explanations of the ML model's logic for the prediction. For three relevant ADME properties, namely HLM, MDR1-ER, and solubility, a set of differently-performing molecules was selected to illustrate the explanations of their predictions. The explanations are based on the breakdown of feature contributions concerning the mean expected prediction of the model for the ADME endpoint under study. The breakdown of feature contributions represents a useful tool for humans to practically understand the impact of certain molecular properties on the prediction. The expert can combine the insights of feature contributions for a single molecule with the feature's marginal impact at the global level to optimize the search for drug leads, for example in the molecule's neighbor space. The engineering of molecules from properties is an active field of research [48, 49], where issues such as the optimization of multiple properties are challenging [50]. The information about the contribution of a certain feature can help in the engineering of models for inverse molecular design or help to optimize the search for improved drug leads. As the variation in the ADME activity is quantifiable by the marginal impact on the SHAP values for each feature the search can be directed towards molecules satisfying these characteristics. Furthermore, the information about the most relevant molecular properties can also help to fix complementary criteria of evaluation for possible drug leads.

Conclusion

This study presents a preliminary contribution, based on SHAP analysis, to the understanding of the molecular feature's role in the ML model's prediction for the six specific ADME endpoints in a recently published public available data set. As the impact of the molecular descriptors on the ADME activity is quantifiable according to the marginal impact in terms of SHAP values, this model provides valuable information about the desired properties of drug leads. In future work, we plan to extend the proposed approach to propose a computational approach that systematically analyzes the marginal impact of a wider set of molecular descriptors and uses this information for automatic drug screening.

Acknowledgements
This research was partially funded by grant ALECTORS-2023 from the Polytechnic University of Catalonia (UPC).

Authors' contributions

C.K. conceptualization of the study, experiments, data analysis, figure generation, initial draft writing, edition, and revision of this manuscript. A.V. conceptualization of ths study, writing, and revision of the manuscript. All authors participated in the writing and editing of this manuscript.

Funding
C.K. supported partially by grant ALECTORS-2023 from the Polytechnic University of Catalonia (UPC).

Availability of data and materials

The dataset supporting the conclusions of this article is available in the GitHub repository https://github.com/carolinekonig/ADME_SHAPexplanations.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors approved the manuscript and agreed with its publication.

Competing interests

The authors declare no competing interests.

Received: 26 May 2024 Accepted: 26 July 2024
Published online: 01 August 2024

References

1. Keyvanpour MR, Shirzad MB. An analysis of QSAR research based on machine learning concepts. Curr Drug Discov Technol. 2021;18(1):17–30.
2. Wu Z, Zhu M, Kang Y, Leung ELH, Lei T, Shen C, et al. Do we need different machine learning algorithms for QSAR modeling? A comprehensive assessment of 16 machine learning algorithms on 14 QSAR data sets. Brief Bioinforma. 2020;22(4):bbaa321.
3. Kumar V, Faheem M, Lee KW, et al. A decade of machine learning-based predictive models for human pharmacokinetics: Advances and challenges. Drug Discov Today. 2022;27(2):529–37.
4. Balani SK, Miwa GT, Gan LS, Wu JT, Lee FW. Strategy of utilizing in vitro and in vivo ADME tools for lead optimization and drug candidate selection. Curr Top Med Chem. 2005;5(11):1033–8.
5. Feinberg EN, Joshi E, Pande VS, Cheng AC. Improvement in ADMET prediction with multitask deep featurization. J Med Chem. 2020;63(16):8835–48.
6. Huang K, Fu T, Gao W, Zhao Y, Roohani Y, Leskovec J, et al. Artificial intelligence foundation for therapeutic science. Nat Chem Biol. 2022;18(10):1033–6.
7. Fang C, Wang Y, Grater R, Kapadnis S, Black C, Trapa P, et al. Prospective Validation of Machine Learning Algorithms for Absorption, Distribution, Metabolism, and Excretion Prediction: An Industrial Perspective. J Chem Inf Model. 2023;63(11):3263–74.

8. Przybylak K, Madden J, Covey-Crump E, Gibson L, Barber C, Patel M, et al. Characterisation of data resources for in silico modelling: benchmark datasets for ADME properties. Expert Opin Drug Metab Toxicol. 2018;14(2):169–81.
9. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: a benchmark for molecular machine learning. Chem Sci. 2018;9(2):513–30.
10. Wieder O, Kohlbacher S, Kuenemann M, Garon A, Ducrot P, Seidel T, et al. A compact review of molecular property prediction with graph neural networks. Drug Discov Today Technol. 2020;37:1–12.
11. Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural message passing for quantum chemistry. In: International conference on machine learning. Sydney: JMLR.org; 2017. p. 1263–72.
12. Ma J, Sheridan RP, Liaw A, Dahl GE, Svetnik V. Deep neural nets as a method for quantitative structure-activity relationships. J Chem Inf Model. 2015;55(2):263–74.
13. Tropsha A, Isayev O, Varnek A, Schneider G, Cherkasov A. Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR. Nat Rev Drug Discov. 2024;23(2):141–55.
14. Roscher R, Bohn B, Duarte MF, Garcke J. Explainable machine learning for scientific insights and discoveries. IEEE Access. 2020;8:42200–16.
15. Lisboa P, Saralajew S, Vellido A, Fernández-Domenech R, Villmann T. The coming of age of interpretable and explainable machine learning models. Neurocomputing. 2023;535:25–39.
16. Burkart N, Huber MF. A survey on the explainability of supervised machine learning. J Artif Intell Res. 2021;70:245–317.
17. Gallego V, Naveiro R, Roca C, Rios Insua D, Campillo NE. AI in drug development: a multidisciplinary perspective. Mol Divers. 2021;25:1461–79.
18. Yang G, Rao A, Fernandez-Maloigne C, Calhoun V, Menegaz G. Explainable AI (XAI) in biomedical signal and image processing: promises and challenges. In: 2022 IEEE International Conference on Image Processing (ICIP). Bordeaux: IEEE; 2022. p. 1531–1535. https://doi.org/10.1109/ICIP46576.2022.9897629.
19. Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York: Association for Computing Machinery; 2016. p. 1135–44.
20. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17. Red Hook: Curran Associates Inc.; 2017. p. 4768–77.
21. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29:1189–232.
22. Rodríguez-Pérez R, Bajorath J. Interpretation of Compound Activity Predictions from Complex Machine Learning Models using Local Approximations and Shapley Values. J Med Chem. 2019;63(16):8761–77.
23. Wojtuch A, Jankowski R, Podlewska S. How can SHAP values help to shape metabolic stability of chemical compounds? J Cheminformatics. 2021;13:1–20.
24. Anjum M, Khan K, Ahmad W, Ahmad A, Amin MN, Nafees A. New SHapley Additive ExPlanations (SHAP) Approach to Evaluate the Raw Materials Interactions of Steel-Fiber-Reinforced Concrete. Materials. 2022;15(18):6261.
25. Ghose AK, Viswanadhan VN, Wendoloski JJ. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. J Comb Chem. 1999;1(1):55–8.
26. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem. 2002;45(12):2615–23.
27. Breiman L. Random Forests. Mach Learn. 2001;45:5–32.
28. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst. 2017;30:52.
29. Sutton CD. Classification and Regression Trees, Bagging, and Boosting. Handb Statist. 2005;24:303–29.
30. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics. 2010;26(10):1340–7.
31. Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. Knowl Inf Syst. 2014;41:647–65.
32. Shapley LS. A Value for N-Person Games. Santa Monica: RAND Corporation; 1952. p. 295. https://doi.org/10.7249/P0295.
33. Wildman SA, Crippen GM. Prediction of Physicochemical Parameters by Atomic Contributions. J Chem Inf Comput Sci. 1999;39(5):868–73.
34. Ertl P, Rohde B, Selzer P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. J Med Chem. 2000;43(20):3714–7.
35. Labute P. A widely applicable set of descriptors. J Mol Graph Model. 2000;18(4–5):464–77.
36. Hall LH, Kier LB. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modeling. Rev Comput Chem. 1991;2:367–422.
37. Rogers D, Hahn M. Extended-Connectivity Fingerprints. J Chem Inf Model. 2010;50(5):742–54.
38. Morgan HL. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. J Chem Doc. 1965;5(2):107–13.
39. Obach R. The prediction of human clearance from hepatic microsomal metabolism data. Curr Opin Drug Discov Dev. 2001;4(1):36–44.
40. Varma MV, Perumal OP, Panchagnula R. Functional role of P-glycoprotein in limiting peroral drug absorption: optimizing drug delivery. Curr Opin Chem Biol. 2006;10(4):367–73.
41. Broccatelli F, Larregieu CA, Cruciani G, Oprea TI, Benet LZ. Improving the prediction of the brain disposition for orally administered drugs using BDDCS. Adv Drug Deliv Rev. 2012;64(1):95–109.
42. Jiang L, Kumar S, Nuechterlein M, Reyes M, Tran D, Cabebe C, et al. Application of a high-resolution in vitro human MDR1-MDCK assay and in vivo studies in preclinical species to improve prediction of CNS drug penetration. Pharmacol Res Perspect. 2022;10(1):e00932.

43. Sun H, Shah P, Nguyen K, Yu KR, Kerns E, Kabir M, et al. Predictive models of aqueous solubility of organic compounds built on A large dataset of high integrity. Bioorg Med Chem. 2019;27(14):3110–4.
44. Cheng T, Li Q, Wang Y, Bryant SH. Binary classification of aqueous solubility using support vector machines with reduction and recombination feature selection. J Chem Inf Model. 2011;51(2):229–36.
45. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 1997;23(1–3):3–25.
46. M Honorio K, L Moda T, D Andricopulo A. Pharmacokinetic properties and in silico ADME modeling in drug discovery. Med Chem. 2013;9(2):163–176.
47. Wang NN, Deng ZK, Huang C, Dong J, Zhu MF, Yao ZJ, et al. ADME properties evaluation in drug discovery: Prediction of plasma protein binding using NSGA-II combining PLS and consensus modeling. Chemom Intell Lab Syst. 2017;170:84–95.
48. Alshehri AS, You F. Deep learning to catalyze inverse molecular design. Chem Eng J. 2022;444:136669.
49. Sridharan B, Goel M, Priyakumar UD. Modern machine learning for tackling inverse problems in chemistry: molecular design to realization. Chem Commun. 2022;58(35):5316–31.
50. Iovanac NC, MacKnight R, Savoie BM. Actively Searching: Inverse Design of Novel Molecules with Simultaneously Optimized Properties. J Phys Chem A. 2022;126(2):333–40.

Publisher’s Note