Computational Toxicology 33 (2025) 100341

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.sciencedirect.com/journal/computational-toxicology

Full Length Article

Augmented allometric scaling: Predicting drug clearance in farm animals with machine learning using body weight

David Inauen [a,*], Leonie Sophie Lautz [b,1], Aalbert Jan Hendriks [c], Ronette Gehring [a]

[a] Utrecht University, Faculty for Veterinary Medicine, Institute for Risk Assessment Sciences, One Health Pharmacology, Yalelaan 104-106, 3584 CM Utrecht, the Netherlands
[b] Wageningen Food Safety Research, Wageningen University and Research, Akkermaalsbos 2, 6708 WB Wageningen, the Netherlands
[c] Radboud University Nijmegen, Department of Environmental Science, Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
PBK
Allometric scaling
QSAR
Pharmacokinetics
Veterinary medicine
Clearance

ABSTRACT

In farm animals, kinetic data of exogenous chemicals, such as pharmaceuticals, environmental pollutants or feed contaminants, are scarce. To allow extrapolation across chemicals and species this study developed a machine learning approach that integrated allometric scaling and quantitative structure–activity relationships to predict total body clearance in farm animals. Using body weight and molecular descriptors of chemicals, the models applied both linear and non-linear machine learning methods such as random forest to predict clearance. Data for intravenously administered chemicals were collected from literature from a variety of species. Molecular descriptors of these chemicals were computed. Log-transformed clearances were predicted for five farm animal species—cattle, sheep, goat, swine, horse—as well as dogs and cats for comparative analysis. Two models using machine learning methods were developed: a purely extrapolative machine learning model, and an approach titled "augmented allometric scaling" which, similarly to simple allometric scaling, used pre-existing data in other species to predict a chemicals' clearance in a target species. The extrapolative approach had large differences in training and test set metrics, while the latter approach demonstrated modestly improved predictive accuracy over simple allometric scaling in farm animals with up to 60.8% of predictions below a fold error of 2, compared to 51% given by allometry, with a difference of up to 0.5 fold errors. In dogs, the new approach performed comparably and worse in cats. This study highlights potentials and limits of machine learning in refining kinetic predictions in farm animals.

1. Introduction

Predicting the kinetics of chemicals in food-producing animals is crucial for assessing potential risks to animal health and human food safety. Of the kinetic parameters, total body clearance (CL) is particularly important, as it quantifies the efficiency of the body's metabolizing and excretory organs in eliminating exogenous chemicals. This parameter determines the relationship between external and internal exposure to drugs, environmental pollutants, pesticides, and feed additives [1].

Physiologically based kinetic models (PBK) have been used to predict tissue concentrations of chemicals in farm animals [2]. These models require precise input parameters, which often need to be predicted [3,4,5]. For human CL, there exists a wide range of predictive models, ranging from allometric scaling (AS) to machine learning (ML) [6,7,8,9]. Fewer such models have been developed in farm animals due to lack of data and resources. Especially the development of ML models proves to be difficult due to their data hunger.

AS is an empirical method that predicts kinetic parameters based on body weight (BW). It is grounded in the theory that physiological processes scale predictably with size among different species. For many small-molecule chemicals, correlation between averaged BW and total body CL has been demonstrated [10]. Consequently, AS is used in drug development and toxicology to scale from species with more data (e.g., laboratory animals) to species with sparser data (e.g., humans, farm animals). In general, AS is applicable to chemicals that are mainly excreted unchanged through urine or exhibit a high hepatic extraction ratio, while for chemicals with tubular reabsorption or heavy hepatic metabolism and low extraction ratios, AS fails to perform well as a

* Corresponding author.
E-mail addresses: d.inauen@uu.nl (D. Inauen), leonie.lautz@esqlabs.com (L.S. Lautz), jan.hendriks@ru.nl (A.J. Hendriks), r.gehring@uu.nl (R. Gehring).
[1] Present address: ESQlabs GmbH, Am Sportplatz 7, Saterland 26683, Germany.

https://doi.org/10.1016/j.comtox.2025.100341
Received 27 November 2024; Received in revised form 27 February 2025; Accepted 3 March 2025
Available online 4 March 2025

predictive tool [11]. In the latter case, CL is influenced by protein binding, and the expression and activity of transporters and enzymes, which vary unpredictably between species. Corrective approaches such as introducing brain weight, maximum life-span potential or the method of exponents have been considered to improve predictions [12]. Whether improvements from these approaches are from a biological basis is not clear [13]. AS may therefore only be used as a retrospective tool as there is currently no method to determine beforehand for which chemicals AS would be a feasible method.

The value of CL is determined by interactions between the physiological traits of the organism (biology) and the chemical properties of the chemical. Quantitative structure–activity relationships (QSARs) explore this relationship, assuming molecular properties inform about biological activity like kinetics. The relationship to activity is usually assumed to be a function of such properties, for example physicochemical descriptors or theoretical molecular descriptors. The recent surge in computational processing power, coupled with an ever-growing repository of biological data, has made it possible to apply ML to leveraging molecular properties in more sophisticated ways [14,15]. Examples of non-linear methods using molecular properties to predict plasma CL in humans include Random Forest (RF) [16,17,18,19], Support Vector Machine [16,20,21,22] and Gradient Boosting [16,19,23]. Other ML models have combined molecular properties with experimental CL data, offering hybrid approaches [22,23,24]. Those latter models implicitly integrated interspecies scaling techniques akin to AS, but pivot on kinetic parameters rather than BW for predictions.

CL of chemicals in farm animals is often unknown, due to a scarcity of experimental *in vivo* data. Addressing this, our study engaged in a hybrid approach, constructing regression models that incorporated the species' BW and a unique identifier for each species, and molecular properties of a wide range of different chemicals. This allowed the use of clearance data for the same chemical across species of varying BWs. Two different models were developed with the objective to determine whether such an approach could serve as a foundation for extrapolative predictions or if it could offer a viable alternative to AS in farm animals. Various regression methods used in ML—Extreme Gradient Boosting (XGBTREE), Support Vector Machine with a polynomial kernel (SVMPOLY), RF, Multiple Linear Regression (MLR), and Stochastic Gradient Boosting (GBM)—were applied on a dataset including a range of animal species and intravenously (IV) administered chemicals to predict CL on a *per species* basis for multiple farm animal species. While this approach shared similarities with QSAR models by assuming a relationship between molecular descriptors and clearance, our model also integrated BW and species identifiers.

2. Materials and methods

2.1. CL dataset

BW (kg) and CL (ml/min/kg) data were compiled manually by searching the PubMed and Google Scholar literature databases. The initial broad search strategy involved the keywords "pharmacokinetics" and/or "toxicokinetics" alongside specific species names. The next step was to narrow these search results down to include only IV kinetic studies that also provided essential BW data. A refined, targeted search was subsequently performed to augment the dataset with specific species-chemical pairs. A summary was provided in Table 1. To improve data uniformity, both BW and CL were log_{10}-transformed (denoted as logBW and logCL, respectively) to reduce skewness.

2.2. Molecular descriptors

The three-dimensional structures of the chemicals were generated using the *RDkit* package (v2022.9.4) in Python (v3.9.0), with SMILES strings sourced from PubChem—preferentially isomeric SMILES where available, otherwise the canonical version [25,26]. By default, *RDkit* performed a sanitization operation that ensured the structures conformed to Lewis dot representations, and explicit hydrogens were added. Subsequently, 1826 two- and three-dimensional molecular descriptors were calculated for each molecule using the *Mordred* package in Python (v1.2.0) [27], a convenient open-source tool designed for computing molecular descriptors.

Table 1
Summary of CL dataset. Medians (range) of BW and CL are provided.

Species	n chemicals	BW (kg)	CL (ml/min/kg)
Buffalo	6	350 (108–440)	4.17 (0.09–8.17)
Camel	13	385 (251–560)	1.70 (0.17–239.28)
Cat	39	4.1 (2.3–6.6)	4.95 (0.06–35.00)
Cattle	67	355 (30–800)	3.18 (0.01–88.50)
Dog	54	11.5 (6–32)	7.00 (0.25–167.00)
Donkey	14	143 (85–282)	3.49 (0.75–20.17)
Goat	57	35 (11–80)	3.72 (0.07–139.67)
Horse	60	500 (85.5–600.0)	11.8 (0.05–131.70)
Llama	15	112 (98.0–146.8)	1.33 (0.19–50.00)
Monkey	27	4.7 (0.7–12.7)	10.2 (0.83–69.40)
Mouse	11	0.02 (0.02–0.03)	60.9 (1.67–146.67)
Rabbit	24	3.1 (1.71–4.40)	3.54 (0.47–76.67)
Rat	55	0.28 (0.14–0.50)	26.1 (0.25–265.00)
Sheep	89	48 (15.0–118.9)	5.16 (0.01–190.00)
Swine	46	28.5 (2.6–217.3)	4.67 (0.16–74.83)

2.3. Workflow

The target species were cattle, sheep, goats, swine and horses—common farm animals in Europe—as well as dogs and cats, which were included to broaden the comparative scope of the study.

The models were developed by reusing the dataset for each target species, obtaining new test and training sets each time. The following procedures were repeated for each target species.

2.3.1. Train test split

The CL data of each chemical were merged with the corresponding molecular descriptors. The test and training sets were obtained by the following procedure: the test set constituted of 75 % of the target species CL data and the training set constituted of the non-target species together with the remaining 25 % of the target species data, as shown on the left-hand side of Fig. 1. Leaving 25 % of the target species data inside the training set ensured that the model could train on data points with the label of the target species. Using this method, sufficiently large test sets could be obtained for the extrapolative model. The 75–25 split of the target species data was performed by the maxDissim function from the *caret* package (v6.0–94) in R (v4.3.1) which selected a set of chemicals which were structurally diverse from an initial randomly selected subset of the target species data [28,29]. For discussion purposes, phenylbutazone and salicylate were manually included in all applicable test sets due to their potential unsuitability for AS application [30,31,32,33,34].

2.3.2. Modeling approaches

Two different approaches were considered: The first approach was a classical "*extrapolative*" ML model, demonstrating the model's capability to predict to "unseen" chemicals. The second approach was a restriction of the first, titled "*augmented AS*", where the test chemicals were known, that is, the model trained on those chemicals and only BW and the species label were varied when predicting the CL of a chemical. This approach incorporated information of other chemicals, as opposed to the simple AS approach where each chemical is treated in isolation. The obvious disadvantage of this approach was that unlike an extrapolative ML model it would not be able to generalize to new chemicals and thus similarly to AS was dependent on pre-existing data of the chemical in other animal species. Data availability is an issue in the veterinary sciences, thus the prediction performance in case of absence or presence of kinetic data in other species was of interest. The training and test sets of

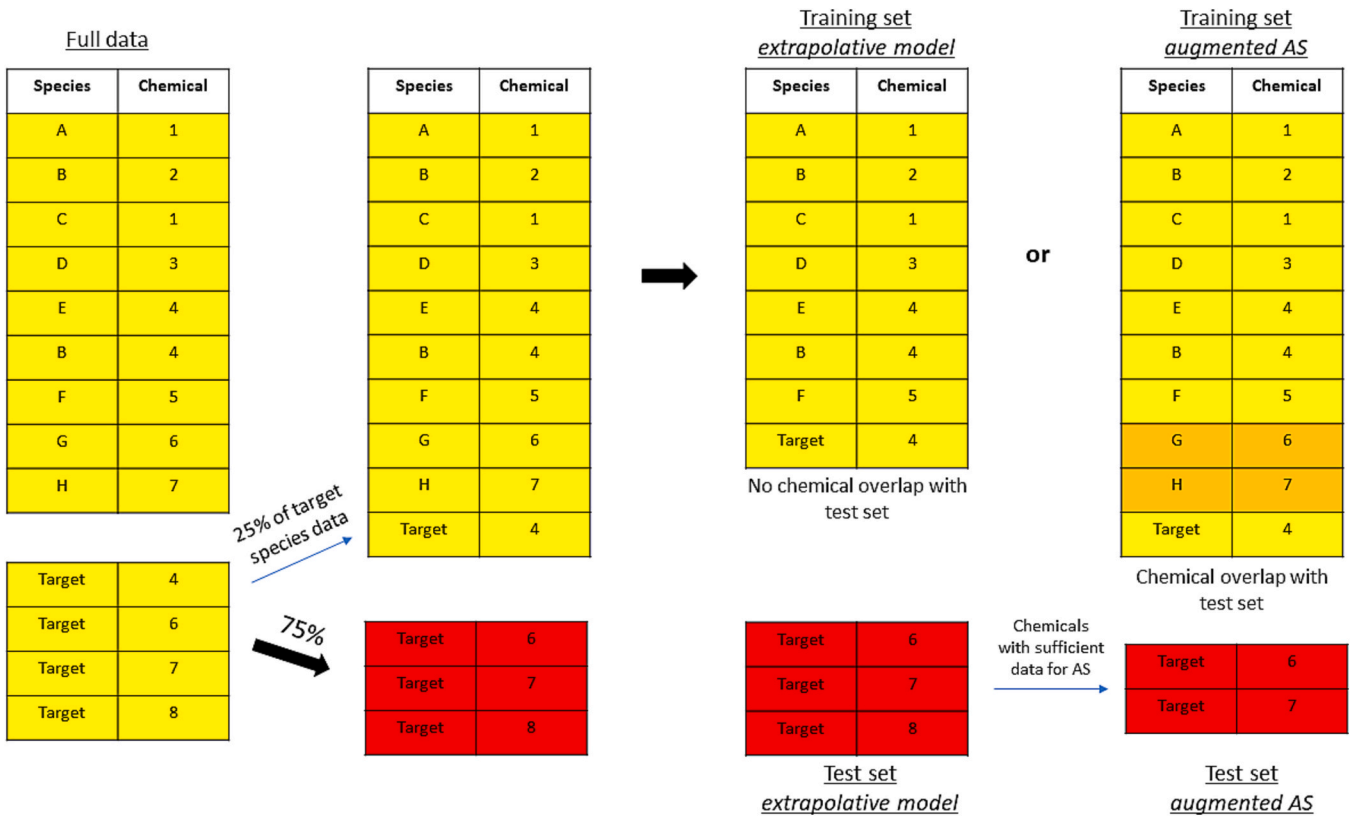

Fig. 1. Structure of the species-chemical data. Simplified schematic of train-test split. The right-hand side depicts the extrapolative model and augmented AS model.

the two approaches are highlighted on the right-hand side of Fig. 1.

2.3.3. Filtering procedure

To reduce the number of predictive features and thereby prevent overfitting and correlated descriptors, a filtering procedure was applied to the training set to exclude certain molecular descriptors [16,35]. This process led to a unique set of descriptors for each target species due to variations in their respective training sets. The steps of the filtering procedure were as follows:

1. Molecular descriptors with NA values and/or that belonged to the *Mordred* class "Autocorrelation" were removed.
2. Descriptors with low variance descriptors were removed by the "nearZeroVar" function of the *caret* package. A descriptor was considered low variance if the ratio of the frequencies of the most common value over the second most common value exceeded 95/5, and if the proportion of unique values was less than 10 % of the total count.
3. Descriptors with fewer than 5 % non-zero entries were removed.
4. Correlated descriptors were handled as follows: Pairwise Spearman correlation coefficients were calculated for each descriptor. The descriptor correlated with the largest number of other descriptors (absolute correlation > 0.7) was added to the final feature set. In case of a tie, the descriptor that appeared first in the *Mordred* listing was chosen. All descriptors correlated to the selected one were removed, and the pairwise correlations were recalculated on the reduced dataset. This process continued until all descriptors were either included in the final feature set or removed. The Spearman correlation coefficient was preferred over the Pearson coefficient due to its ability to capture non-linear relationships between descriptors.

2.3.4. Training

After filtering, a feature selection algorithm was applied to the training set using the Boruta function (doTrace = 2; maxRuns = 200; excluding tentative features) from the R package *Boruta* (v8.0.0) [36]. This algorithm, based on RF, ranked candidate features by comparing the model's performance on the original data against a random permutation of the features and selected individual features based on a specified cutoff value. The Boruta algorithm, utilizing the filtered molecular descriptors and logBW as predictors, targeted logCL as the variable to predict. Preliminary runs revealed that the choice of random seed during Boruta feature selection did not greatly impact prediction accuracy, thus all feature selections were performed using seed 0. Finally, an additional categorical feature identifying the species was added to the Boruta-selected features. Ultimately, CL was predicted using three classes of predictors: BW, molecular descriptors and a species identifier, setting the present approach apart from QSAR which relies solely on molecular information. No further pre-processing was applied to the features. The extrapolative and augmented AS models were trained using the *caret* package with the XGBTREE, SVMPOLY, RF, MLR and GBM methods, all set to default tuning hyperparameters. The MLR method assumed a linear relationship between the predictive features and CL while the other methods introduced non-linearity to capture the complex dynamics of drug metabolism across species. Training was conducted using repeated 10-fold cross-validation (10 repeats) with seed 0.

2.3.5. Simple allometric scaling

The allometric relationship between CL and BW was given by the equation $CL = a^{*}BW^{b}$, where a was the allometric coefficient and b was the allometric exponent. Using logarithmic transformations, this relationship could be expressed as $logCL = loga + b^{*}logBW$. This allowed a and b to be computed by simple linear regression. Note that for the allometric regression, CL was measured as ml/min. For evaluation, CL was converted back to ml/min/kg.

This method required a minimum of three data points from different animal species for each chemical, excluding the species of interest (target species). A subset of the test chemicals was formed, composed of

the test chemicals which had data available in at least three non-target species. For each of those chemicals, simple AS predictions in the target species were performed by means of linear regression.

2.3.6. Evaluation

As evaluation metrics, the root mean squared error (RMSE), coefficient of determination (R2) and, to compare with AS, fold errors (FE) were computed. The RMSE was defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n_j}(CL_i - \widehat{CL}_i)^2}{n_j}} \quad (1)$$

and the R2 was defined as

$$R2 = 1 - \frac{\sum_{i=1}^{n_j}(CL_i - \widehat{CL}_i)^2}{\sum_{i=1}^{n_j}(CL_i - \overline{CL})^2} \quad (2)$$

with n_j the number of data points in the j-th (target species) test set, CL the observed CL and $\widehat{CL} = 10^{\widehat{\log CL}}$, where $\widehat{\log CL}$ was the predicted value of the log_{10}-transformed CL of the test set. $\overline{CL}$ denoted the mean of the observed values. Note that in the case of simple linear regression, the above R2 definition is equivalent to the squared correlation between observed and predicted values.

The FE of the i-th prediction was defined as

$$FE_i = \begin{cases} \frac{\widehat{CL}_i}{CL_i}, & \widehat{CL}_i > CL_i \\ \frac{CL_i}{\widehat{CL}_i}, & \widehat{CL}_i \leq CL_i \end{cases} \quad (3)$$

and the absolute average fold error (AAFE) as

$$AAFE = 10^{\frac{\sum_{i=1}^{n_j}\left|\log_{10}\frac{\widehat{CL}_i}{CL_i}\right|}{n_j}} \quad (4)$$

For comparison to simple AS, the model's accuracy was quantified by reporting the percentage of chemicals with a prediction error less than 2-fold error ("perc < 2fold"), offering insight on how many chemicals were predicted within error margins that were considered acceptable for chemical risk assessment [38]. The AAFE reflected the average prediction error across chemicals, influenced by the magnitude of the individual errors. The predictions from the augmented AS model were compared with the simple AS predictions on a subset of the test set, refer to right-hand side of Fig. 1. In contrast to simple AS, the augmented AS model was trained on a broader data set, using chemicals both contained in the test set and outside of it.

3. Results

3.1. Dataset

The comprehensive dataset that emerged from the search strategy included CL values for 244 IV-administered chemicals across 15 mammalian species, including laboratory animals (mice, rats, monkeys and rabbits), cats and dogs, and various farm animals (details provided in the supplementary material (SM) 3). For categorization purposes, both beef and dairy cattle were consolidated under the label of "cattle".

The data availability varied, with some species missing data for certain chemicals and some chemicals missing data for certain species. When multiple CL values for the same species-chemical pairing appeared, likely due to varying study conditions or population characteristics such as age and sex, the mean BW and CL were calculated. Where studies provided only a range of BW, the median value was used. In instances of wide BW variation within the same species attributable to different age groups, data from only one age group were considered, with a preference for adult animals (refer to SM 3 for details). In two salicylate studies—one in dogs and one in cats—BW values were not reported. Therefore, BW was imputed using the average of all other dog respectively cat BW values in the dataset.

3.2. Predictions

3.2.1. Extrapolative ML model

There were stark differences between training and test RMSE and R2 values as seen in Table 2, indicating that this ML approach was not suitable to extrapolate on "unseen" chemicals: The XGBTREE method in horses achieved the overall highest test set R2 value of 0.41, while R2 values of all other methods-species combinations were lower or even negative. The best perc < 2fold and AAFE ranges in farm animals were recorded for XGBTREE (32.4 %-50 %, 2.6–3.1), RF (37.1 %-50 %, 2.5–2.9) and GBM (41.3 %-48.6 %, 2.5–3.1), with RF marginally outperforming the other methods (SM 1 Fig. 8). SVMPOLY and MLR consistently showed inferior results compared to these methods. The performance for dogs was on a par with that observed in the larger farm animal species, while the results for cats were notably poorer. The model exhibited large outlier errors, for example, the maximum FE for MLR in cattle was 1355, whereas for RF it was 312.

Select chemicals are discussed: phenylbutazone was consistently

Table 2
Extrapolative model: Results for target species on test set. Column "Features" indicates the number of features after the Boruta procedure, including BW and the species identifier. RMSE and R2 on training and test sets are given.

Method	Features	N train	RMSE train	R2 train	N test	RMSE test	R2 test
Cattle							
XGBTREE	76	379	0.33	0.79	50	0.74	−0.14
SVMPOLY	76	379	0.31	0.81	50	0.62	0.21
RF	76	379	0.2	0.92	50	0.7	−0.02
MLR	76	379	0.43	0.63	50	0.87	−0.56
GBM	76	379	0.28	0.85	50	0.74	−0.12
Sheep							
XGBTREE	79	369	0.3	0.82	68	0.64	0.15
SVMPOLY	79	369	0.3	0.83	68	0.61	0.23
RF	79	369	0.21	0.92	68	0.6	0.26
MLR	79	369	0.42	0.68	68	0.69	0.01
GBM	79	369	0.27	0.86	68	0.59	0.29
Goat							
XGBTREE	74	411	0.21	0.91	45	0.6	0.12
SVMPOLY	74	411	0.32	0.8	45	0.55	0.27
RF	74	411	0.21	0.92	45	0.57	0.21
MLR	74	411	0.44	0.63	45	0.58	0.17
GBM	74	411	0.29	0.83	45	0.57	0.2
Swine							
XGBTREE	74	423	0.21	0.92	35	0.61	0.06
SVMPOLY	74	423	0.31	0.81	35	0.64	−0.06
RF	74	423	0.2	0.92	35	0.58	0.14
MLR	74	423	0.44	0.62	35	0.6	0.09
GBM	74	423	0.29	0.83	35	0.54	0.25
Horse							
XGBTREE	79	440	0.22	0.89	46	0.55	0.41
SVMPOLY	79	440	0.39	0.64	46	0.69	0.07
RF	79	440	0.21	0.9	46	0.59	0.32
MLR	79	440	0.4	0.62	46	0.7	0.05
GBM	79	440	0.3	0.79	46	0.56	0.38
Dog							
XGBTREE	81	420	0.21	0.91	41	0.51	0.35
SVMPOLY	81	420	0.33	0.78	41	0.52	0.31
RF	81	420	0.2	0.92	41	0.5	0.36
MLR	81	420	0.41	0.66	41	0.9	−1.04
GBM	81	420	0.29	0.84	41	0.5	0.36
Cat							
XGBTREE	80	466	0.21	0.91	30	0.74	−0.12
SVMPOLY	80	466	0.07	0.99	30	0.76	−0.17
RF	80	466	0.2	0.92	30	0.72	−0.05
MLR	80	466	0.42	0.64	30	0.88	−0.6
GBM	80	466	0.3	0.81	30	0.78	−0.26

overpredicted (FE > 10) across all farm animals by all ML methods, as seen in the observed vs predicted plots in SM 1 Figs. 9–15. Carprofen was considerably overpredicted in horses and cats, with identical predictions for R and S stereoisomers due to their highly similar molecular structures. The same was also true for ketoprofen R and S predictions. Salicylate was accurately predicted in horses and moderately well in goats, swine and dogs (FE < 5) but was strongly overpredicted in cats. Bromide in sheep and dimethyl sulfoxide in horses were strongly overpredicted. The observed trends indicated that the extrapolative model tended to overpredict chemicals with a low CL, across all species.

3.2.2. Augmented AS model

R2 values of the different methods on the test sets ranged from 0.39 up to 0.76, as given in SM 1 Table 1. These higher values compared to the extrapolative model were achieved due to partial bias of the model as the same chemicals were present both in the training and test sets which is a limitation posed in this approach. The simple AS performance in farm animals showed a perc < 2fold range of 35.7 %-57.1 % and an AAFE range of 2–3.2, as seen in Fig. 2. The linear method MLR outperformed AS in cattle, sheep, goats and horses with perc < 2fold and AAFE ranges of 35.7 %-60.9 % and 1.8–2.3. Among the non-linear methods, further improvements were observed across all farm animals: XGBTREE (42.9 %-62.5 %, 1.9–2.2), SVMPOLY (28.6 %-66.7 %, 1.9–2.7), RF (35.7 %-75 %, 1.7–2.3), and GBM (42.9 %-70.8 %, 1.9–2.1). AS and ML results for horses were notably worse compared to other farm animals. In dogs, RF and GBM performed comparably to AS. In cats, AS performed better than all ML methods. For all evaluation in this section, tramadol in horse was not included since simple AS returned an outlier FE of 363886. Observed vs predicted plots are given in SM 1 Figs. 16–22.

On average, RF and GBM were the best performing ML methods, followed by XGBTREE. RF and GBM outperformed simple AS by an average of 0.35 and 0.32 FEs, with better results than AS in 55.8 % and 58.7 % of predictions, respectively, as seen in Table 3. On farm animals, these differences increased to 0.5 and 0.46 (61.8 % and 59.8 % predictions better than simple AS). While the ML approach performed better than simple AS in larger species such as cattle and horse, it performed worse in cats.

Direct comparisons for individual predictions can be made (SM 1 Tables 2 and 3). In case of extreme predictions by AS (FE > 10), improvements could be found, such as diminazene in cattle (RF: 1.72, GBM: 1.29) or carprofen in horses (RF: 2.36 R, 2.08 S, GBM: 2.27 R, 2.08 S). An exception was trimethoprim in sheep (RF: 15.67, GBM: 12.29). For AS predictions with FEs between 2 and 10, RF and GBM mostly demonstrated similar or lower FEs, with exceptions of carprofen R (RF: 5.69, GBM: 3.08) and salicylate (RF: 10.67, GBM: 13.78) in cats. For salicylate in cats, all ML methods underperformed relative to AS. Examples of worse performance than AS were found in xylazine hydrochloride in sheep (RF: 2.97, GBM: 2.48), salicylate (RF: 4.95, GBM: 4.84) and trimethoprim (RF: 3.43, GBM: 2.72) in horse, acivicin (RF: 2.66, GBM: 2.27) and tiludronate (RF: 2.4, GBM: 2.98) in dogs, and clavulanic acid (RF: 6.32, GBM: 5.17) and doxycycline (RF: 4.06, GBM: 3.49) in cats. Most improvements were found in the chemicals where both the simple AS and augmented AS approach did not perform within a FE of 2 (SM 1 Fig. 23). In the case of RF, those chemicals (n = 44) had an AAFE of 3.56, compared to 4.62 of the simple AS predictions; for GBM (n = 39), the respective AAFE was 3.86, compared to 4.81 of AS.

3.3. Analysis of feature importance

For concision, feature importances solely for one of the best performing methods, RF, were computed, with details available in SM 1 Tables 4–8. The importance of each feature was determined by comparing mean squared errors of the normal model and models where the feature values were randomly permuted. The most influential features were Mm, Mi, Mz, which summarize information on the chemical's molecular composition such as molecular weight, number of atoms, and the number of bonds. LogBW emerged as the second most critical descriptor overall.

Another consistently high ranked descriptor was LogS, the logarithm of a chemical's water solubility. The analysis also highlighted the relevance of topological charge indexes (JGI), which reflect correlations with the charge distribution within a molecule [39]. Other important descriptors were PEOE_VSA, SlogP_VSA, and Estate_VSA, which measure the contributions of atoms in the molecule to the Van der Waals surface area, depending on their partial charge, logP, or E-state values, respectively [40]. The VSA_Estate bins atomic Estate values depending on the Van der Waals contributions. These findings align with other ML studies to predict CL in humans, which also deemed LogS, SlogP_VSA and PEOE_VSA as important [16]. Each descriptor listed represents a cluster of correlated molecular descriptors that are filtered out during the preprocessing phase. The descriptor labelling species showed relatively low importance in this context.

4. Discussion

In the present study, two models based on ML were developed to predict total body CL of chemicals in farm animals using molecular descriptors and body weight. The study showcased a purely extrapolative model and an augmented approach to AS, each serving distinct predictive capabilities.

The extrapolative approach demonstrated little success in generalizing, particularly due to large overpredictions and outliers when dealing with lower CL chemicals. This may have been caused by an imbalanced dataset, with relatively fewer low CL chemicals being represented. This limitation points to potential improvements, such as introducing a more balanced dataset or refined selection of molecular and physico-chemical descriptors that may enhance prediction accuracy.

Based on the average FE values, the augmented AS model using ML methods performed better than simple AS on the aggregate of the five farm animal species by up to 0.5 AAFE while it performed comparably in dogs and worse in cats. Overall, both simple and augmented AS in horse performed comparably worse than in ruminants such as cattle, sheep or goat. The dataset consisted in large parts of ruminant data, thus possibly improving the predictions for these related species. For horse, only little data of related species (e.g. donkeys) was available. In similar fashion, cats were the only hypercarnivore in the dataset. Mock simulations removing cat data from the training set showed that the extrapolative model using GBM returned similar or worse RMSE in all species except cattle, indicating that inclusion of dissimilar species still proves useful.

The average reduction of AAFE in all animals by 0.2 using the MLR method over the simple linear regression of AS was likely due to the introduction of additional features in the form of molecular descriptors, and an additional reduction of 0.1 AAFE was found through non-linear methods. The lowest FEs were presented by RF and GBM, with improvements of 1.06 and 0.95 AAFE in cases where both simple and augmented AS did not return FEs below two. *A priori* it is unknown on which chemicals AS would perform well or unsatisfactory, thus the ML approach may be an alternative less likely to give high errors, based on the aggregate molecular information of a multitude of chemicals.

A notable case is phenylbutazone, which exhibits high variability in CL across animal species of similar BW; for instance, CL ranges from 0.02 ml/min/kg in cattle [31] to 0.4 ml/min/kg in horses [32] and up to 6 ml/min/kg in donkeys [30]. Phenylbutazone's minimal excretion in urine, which is only about 1 %-2% in humans [41] and horses [42], poses challenges for AS predictions. While AS achieved FEs below 5 in sheep, goats, swine, and horses, it resulted in a substantially higher FE of 16.5 in cattle. In contrast, the augmented AS model refined predictions in cattle, achieving FE ranges from 6.8 to 12.8 across various ML methods, while maintaining comparable FEs to AS in the other species.

A counter-example is the case of salicylate in cats, where AS achieved

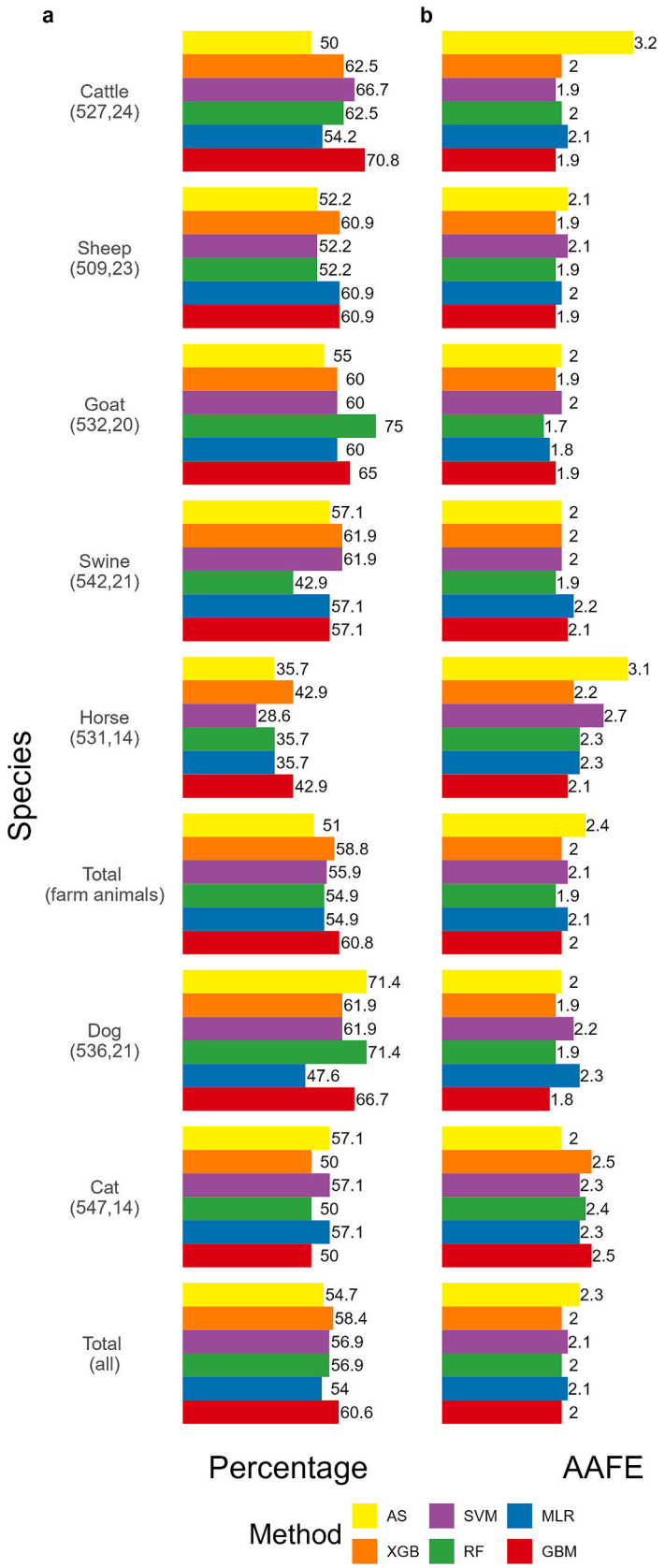

Fig. 2. Summary of FEs on all chemicals with simple AS predictions, and the augmented AS predictions using the ML methods. Total is the weighted mean. The values below each species name indicate the sizes of the training sets of the model, resp. the number of allometrically predicted chemicals. Horse is without tramadol.

Table 3
Percentage of predictions where the augmented AS model performed equal or better than simple AS.

Species	n test	RF (%)	GBM (%)
Cattle	24	79.2	75
Sheep	23	52.2	52.2
Goat	20	70	60
Swine	21	47.6	42.9
Horse	14	57.1	71.4
Farm animals only	102	61.8	59.8
Dog	21	42.9	61.9
Cat	14	28.6	21.4
All species	137	55.8	58.7

a FE of 4.9, but all ML methods performed worse. Research indicates that cats eliminate salicylate more slowly (0.065 ml/min/kg [34], 0.118 ml/min/kg [33]) compared to, for example, dogs (0.25 ml/min/kg [34]), which are of similar BW. This reduced clearance in cats is likely due to their inability to convert salicylate to salicylurate through glycination [43,44]. To enhance ML predictions, the model could be refined by incorporating additional species-specific identifiers, such as dietary habits, phylogenetic relations, or metabolic profiles. Additionally, expanding the database to include species that are either phylogenetically closer to cats or share similar dietary patterns—such as ferrets, wild felines and marine mammals—might improve prediction accuracy.

The ML predictions between stereoisomers for chemicals like ketoprofen and carprofen were indistinguishable, despite known significant differences in the kinetics of their stereoisomers [45,46,47,48,49,50,51]. This reflects a challenge in the model's handling of structural differences. The issue arises because only a limited number of molecular descriptors differentiated between stereoisomers, and these differences were often lost during the filtering process. To improve the model's ability to distinguish between stereoisomers, it would be beneficial to intentionally include descriptors that specifically indicate R or S configurations. Furthermore, expanding the dataset to include a greater variety of chemicals with stereoisomers could enhance the model's discriminatory power and refine its predictions.

Simple allometry studies in farm animals showed variability in prediction accuracy with perc < 2fold values of 65.1 % and AAFEs of 1.86 when predicting clearance based on three large species, and 47.2 % and 3.17 when based on three to four smaller pre-clinical species [37]. Comparatively, a comprehensive review on allometric clearance prediction studies in humans showed perc < 2fold values ranging from 46 % to 81 % (excluding an outlier study at 17 %), with AAFE values between 1.28 and 3.23, underscoring the variability in traditional allometric methods [52]. Meanwhile, published ML models for human clearance predictions have used extensive datasets and demonstrate varied effectiveness. For instance, the model by Wang et al. [16] achieved a perfect perc < 2fold of 100 % across 254 test chemicals (R2: 0.875). In contrast, Lombardo et al. [18] reported perc < 2fold values of 54 % and 70 % for metabolically and renally cleared chemicals, with AAFEs of 2.6 and 1.7, derived from 643 and 329 test chemicals, respectively, using a model based on clearance mechanism assignment. Another model by Iwata et al. [23] reached a perc < 2fold of 66.5 % and an AAFE of 1.92 across 45 test chemicals, employing a multimodal ML model that utilized various pre-clinical data.

To our knowledge, this is the first *in silico* study for predicting bioactivity in farm animals that integrates both body weight with molecular descriptors. Prior studies have not combined these attributes; for example, one study applied allometric approaches with molecular descriptors for predicting LD50 but did not integrate these attributes [53], while a recent ML study predicted plasma half-lives on a number of food animal species using molecular descriptors and fingerprints [54]. The introduction of the BW feature allows for the combination of data across different species. The key advantage of this model lies in its flexibility. Any new data point, regardless of species or chemical, can be added to the training set as long as the essential information is known (CL, BW and molecular descriptors). This flexibility allows the training set to be expanded indefinitely by incorporating data from new species or chemicals, even with just a single data point. Unlike simple AS, which can only improve with additional experimentation on the same chemical, the present ML model is informed by all chemicals simultaneously. Furthermore, species are differentiated not only by BW but also a unique identifier.

The main limitation of the augmented AS approach was that CL data of the same chemical in other species was required to predict for a target species. The ML model was already trained on the molecular descriptors of a chemical, which means that the augmented AS model has data leakage in the molecular descriptors. BW and the species identifier become the only varying features and predictions for a chemical are made on this lower-dimensional subspace of the regression. This is similar to simple AS, which only uses BW as the single predictor. The difference between simple AS and the augmented AS approach using ML is that the former performs simple linear regression with a single chemical, whereas regression of the augmented model is performed by training on data composed of multiple chemicals. Contrast this to the extrapolative model, which avoids data leakage and where the full feature space is used for predictions. Thus, direct comparison between the two ML approaches is not feasible.

Some improvements over simple AS were found, but there is room to further develop the model: In addition to molecular descriptors, fingerprints and other molecular information could be added. Further, the CL dataset could be extended by adding more chemical studies and animals species, for example experimental animal species. More species specific information such as phylogenetics, nutrition and environmental factors could be added. Predictive accuracy, however, is also limited by experimental error; sensitivity of predictions depends on the ML method used [55].

Future research should prioritize enhancing CL prediction for environmental pollutants, pesticides and feed additives in farm animals, considering the complexities introduced by different exposure routes. This could involve adding a feature that codes for route of administration. Additionally, assessing the model's performance using pre-clinical and human data could broaden its applications, potentially including pediatric dosing.

Funding sources

This study is financed by the TK Plate 2.0 project of the European Food Safety Authority (contract: OC/EFSA/SCER/2021/07).

CRediT authorship contribution statement

David Inauen: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Leonie Sophie Lautz:** Writing – review & editing, Investigation, Data curation. **Aalbert Jan Hendriks:** Writing – review & editing. **Ronette Gehring:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the consortium members of the TK Plate 2.0 project for their input.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.comtox.2025.100341.

Data availability

The data collected and used for the models is provided in the supplementary materials at https://doi.org/10.5281/zenodo.14222086.

References

[1] P.L. Toutain, A. Bousquet-Mélou, Plasma clearance, J. Vet. Pharmacol. Ther. 27 (6) (2004) 415–425, https://doi.org/10.1111/j.1365-2885.2004.00605.x.

[2] L.S. Lautz, R. Oldenkamp, J.L. Dorne, A.M.J. Ragas, Physiologically based kinetic models for farm animals: Critical review of published models and future perspectives for their use in chemical risk assessment, Toxicol. In Vitro 60 (2019) 61–70, https://doi.org/10.1016/j.tiv.2019.05.002.

[3] L.S. Lautz, S. Hoeks, R. Oldenkamp, A.J. Hendriks, J.L.C.M. Dorne, A.M.J. Ragas, Generic physiologically based kinetic modelling for farm animals: Part II. Predicting tissue concentrations of chemicals in swine, cattle, and sheep, Toxicol. Lett. 318 (2020) 50–56, https://doi.org/10.1016/j.toxlet.2019.10.008.

[4] W.-C. Chou, et al., An interactive generic physiologically based pharmacokinetic (igPBPK) modeling platform to predict drug withdrawal intervals in cattle and swine: A case study on flunixin, florfenicol, and penicillin G, Toxicol. Sci. 188 (2) (2022) 180–197, https://doi.org/10.1093/toxsci/kfac056.

[5] A. Paini, et al., Next generation physiologically based kinetic (NG-PBK) models in support of regulatory decision making, Comput. Toxicol. 9 (2019) 61–72, https://doi.org/10.1016/j.comtox.2018.11.002.

[6] W.-C. Chou, Z. Lin, Machine learning and artificial intelligence in physiologically based pharmacokinetic modeling, Toxicol. Sci. (2022) kfac101, https://doi.org/10.1093/toxsci/kfac101.

[7] Z. Lin, W.-C. Chou, Machine learning and artificial intelligence in toxicological sciences, Toxicol. Sci. 189 (1) (2022) kfac075–kfac119, https://doi.org/10.1093/toxsci/kfac075.

[8] F. Lombardo, et al., Comprehensive assessment of human pharmacokinetic prediction based on in vivo animal pharmacokinetic data, Part 2: clearance, J. Clin. Pharmacol. 53 (2) (2013) 178–191, https://doi.org/10.1177/0091270012440282.

[9] Danishuddin, V. Kumar, M. Faheem, K. Woo Lee, A decade of machine learning-based predictive models for human pharmacokinetics: advances and challenges, Drug Discov. Today 27 (2) (2022) 529–537, https://doi.org/10.1016/j.drudis.2021.09.013.

[10] Q. Huang, R. Gehring, L.A. Tell, M. Li, J.E. Riviere, Interspecies allometric meta-analysis of the comparative pharmacokinetics of 85 drugs across veterinary and laboratory animal species, J. Vet. Pharmacol. Ther. 38 (3) (2015) 214–226, https://doi.org/10.1111/jvp.12174.

[11] Q. Huang, J.E. Riviere, The application of allometric scaling principles to predict pharmacokinetic parameters across species, Expert Opin. Drug Metab. Toxicol. 10 (9) (2014) 1241–1253, https://doi.org/10.1517/17425255.2014.934671.

[12] I. Mahmood, Application of allometric principles for the prediction of pharmacokinetics in human and veterinary drug development, Adv. Drug Deliv. Rev. 59 (11) (2007) 1177–1192, https://doi.org/10.1016/j.addr.2007.05.015.

[13] P.L. Bonate, D. Howard, Critique of prospective allometric scaling: does the emperor have clothes? J. Clin. Pharmacol. 40 (4) (2000) 335–340, https://doi.org/10.1177/00912700022009017.

[14] T.A. Soares, A. Nunes-Alves, A. Mazzolari, F. Ruggiu, G.-W. Wei, K. Merz, The (Re)-evolution of quantitative structure–activity relationship (QSAR) studies propelled by the surge of machine learning methods, J. Inform. Model. 62 (22) (2022) 5317–5320, https://doi.org/10.1021/acs.jcim.2c01422.

[15] Z. Wang, et al., *In silico* prediction of chemical respiratory toxicity via machine learning, Comput. Toxicol. 18 (May 2021) 100155, https://doi.org/10.1016/j.comtox.2021.100155.

[16] Y. Wang, et al., In silico prediction of human intravenous pharmacokinetic parameters with improved accuracy, J. Chem. Informat. Model. 59 (9) (2019) 3968–3980, https://doi.org/10.1021/acs.jcim.9b00300.

[17] G. Berellini, N.J. Waters, F. Lombardo, In silico prediction of total human plasma clearance, J. Chem. Inform. Model. 52 (8) (2012) 2069–2078, https://doi.org/10.1021/ci300155y.

[18] F. Lombardo, R.S. Obach, M.V. Varma, R. Stringer, G. Berellini, Clearance mechanism assignment and total clearance prediction in human based upon in silico models, J. Med. Chem. 57 (10) (2014) 4397–4405, https://doi.org/10.1021/jm500436v.

[19] F. Miljković, et al., Machine Learning Models for Human In Vivo Pharmacokinetic Parameters with In-House Validation, Mol. Pharm. 18 (12) (2021) 4520–4530, https://doi.org/10.1021/acs.molpharmaceut.1c00718.

[20] U. Fagerholm, S. Hellberg, J. Alvarsson, O. Spjuth, In silico predictions of the human pharmacokinetics/toxicokinetics of 65 chemicals from various classes using conformal prediction methodology, Xenobiotica 52 (2) (2022) 113–118, https://doi.org/10.1080/00498254.2022.2049397.

[21] V.K. Gombar, S.D. Hall, Quantitative structure-activity relationship models of clinical pharmacokinetics: clearance and volume of distribution, J. Chem. Inform. Model. 53 (4) (2013) 948–957, https://doi.org/10.1021/ci400001u.

[22] W. Huang, et al., Prediction of human clearance based on animal data and molecular properties, Chem. Biol. Drug Des. 86 (5) (2015) 990–997, https://doi.org/10.1111/cbdd.12567.

[23] H. Iwata, et al., Predicting total drug clearance and volumes of distribution using the machine learning-mediated multimodal method through the imputation of various nonclinical data, J. Chem. Inform. Model. 62 (17) (2022) 4057–4065, https://doi.org/10.1021/acs.jcim.2c00318.

[24] T. Wajima, K. Fukumura, Y. Yano, T. Oguma, Prediction of human clearance from animal data and molecular structural parameters using multivariate regression analysis, J. Pharm. Sci. 91 (12) (2002) 2489–2499, https://doi.org/10.1002/jps.10242.

[25] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. Accessed: Oct. 17, 2023. [Online]. Available: https://www.python.org/.

[26] *RDkit Open-source cheminformatics*. Accessed: Oct. 17, 2023. [Online]. Available: https://www.rdkit.org/.

[27] H. Moriwaki, Y.-S. Tian, N. Kawashita, T. Takagi, Mordred: a molecular descriptor calculator, J. Cheminf. 10 (1) (2018) 4, https://doi.org/10.1186/s13321-018-0258-y.

[28] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2023. Accessed: Oct. 17, 2023. [Online]. Available: https://www.R-project.org/.

[29] Kuhn and Max, Building predictive models in R using the caret package, J. Stat. Softw. 28 (5) (2008) 1–26, https://doi.org/10.18637/jss.v028.i05.

[30] N.S. Matthews, K.E. Peck, T.S. Taylor, K.L. Mealey, Pharmacokinetics of phenylbutazone and its metabolite oxyphenbutazone in miniature donkeys, Am. J. Vet. Res. 62 (5) (2001) 673–675, https://doi.org/10.2460/ajvr.2001.62.673.

[31] P. De Backer, R. Braeckman, F. Belpaire, M. Debackere, Bioavailability and pharmacokinetics of phenylbutazone in the cow, J. Vet. Pharmacol. Ther. 3 (1) (1980) 29–33, https://doi.org/10.1111/j.1365-2885.1980.tb00476.x.

[32] H.K. Knych, R.M. Arthur, D.S. McKemie, K. Seminoff, B. Hamamoto-Hardman, P. H. Kass, Phenylbutazone blood and urine concentrations, pharmacokinetics, and effects on biomarkers of inflammation in horses following intravenous and oral administration of clinical doses, Drug Test. Anal. 11 (6) (2019) 792–803, https://doi.org/10.1002/dta.2553.

[33] Parton, Balmer, Boyle, Whittem, Machon, The pharmacokinetics and effects of intravenously administered carprofen and salicylate on gastrointestinal mucosa and selected biochemical measurements in healthy cats, J. Vet. Pharmacol. Ther. 23 (2) (2000) 73–79, https://doi.org/10.1046/j.1365-2885.2000.00253.x.

[34] L.E. Davis, B.A. Westfall, Species differences in biotransformation and excretion of salicylate, Am. J. Vet. Res. 33 (6) (1972) 1253–1262.

[35] A.E. Comesana, T.T. Huntington, C.D. Scown, K.E. Niemeyer, V.H. Rapp, A systematic method for selecting molecular descriptors as features when training models for predicting physiochemical properties, Fuel 321 (2022) 123836, https://doi.org/10.1016/j.fuel.2022.123836.

[36] M. Kursa, A. Jankowski, W. Rudnicki, Boruta - A system for feature selection, Fund. Inform. 101 (2010) 271–285, https://doi.org/10.3233/FI-2010-288.

[37] I. Mahmood, M. Martinez, R.P. Hunter, Interspecies allometric scaling. Part I: prediction of clearance in large animals, J. Vet. Pharmacol. Ther. 29 (5) (2006) 415–423, https://doi.org/10.1111/j.1365-2885.2006.00786.x.

[38] I. P. on C. Safety and I.-O. P. for the S. M. of Chemicals, *Characterization and application of physiologically based phamacokinetic models in risk assessment*. World Health Organization, 2010. Accessed: Jun. 12, 2024. [Online]. Available: https://iris.who.int/handle/10665/44495.

[39] J. Galvez, R. Garcia, M.T. Salabert, R. Soler, Charge indexes. New topological descriptors, J. Chem. Inf. Comput. Sci. 34 (3) (1994) 520–525, https://doi.org/10.1021/ci00019a008.

[40] P. Labute, A widely applicable set of descriptors, J. Mol. Graph. Model. 18 (4) (2000) 464–477, https://doi.org/10.1016/S1093-3263(00)00068-1.

[41] J. Aarbakke, Clinical pharmacokinetics of phenylbutazone, Clin. Pharmacokinet. 3 (5) (1978) 369–380, https://doi.org/10.2165/00003088-197803050-00003.

[42] T. Tobin, et al., Phenylbutazone in the horse: a review, J. Vet. Pharmacol. Ther. 9 (1) (1986) 1–25, https://doi.org/10.1111/j.1365-2885.1986.tb00008.x.

[43] M.H. Court, Feline drug metabolism and disposition: pharmacokinetic evidence for species differences and molecular mechanisms, Veter. Clin. North Am. Small An. Pract. 43 (5) (2013) 1039–1054, https://doi.org/10.1016/j.cvsm.2013.05.002.

[44] L.S. Lautz, M.Z. Jeddi, F. Girolami, C. Nebbia, J.L.C.M. Dorne, Metabolism and pharmacokinetics of pharmaceuticals in cats (Felix sylvestris catus) and implications for the risk assessment of feed additives and contaminants, Toxicol. Lett. 338 (2021) 114–127, https://doi.org/10.1016/j.toxlet.2020.11.014.

[45] P. Lees, F.S. Aliabadi, M.F. Landoni, Pharmacodynamics and enantioselective pharmacokinetics of racemic carprofen in the horse, J. Vet. Pharmacol. Ther. 25 (6) (2002) 433–448, https://doi.org/10.1046/j.1365-2885.2002.00436.x.

[46] P.M. Taylor, et al., Pharmacodynamics and enantioselective pharmacokinetics of carprofen in the cat, Res. Vet. Sci. 60 (2) (1996) 144–151, https://doi.org/10.1016/s0034-5288(96)90009-0.

[47] P. Delatour, R. Foot, A.P. Foster, D. Baggo, P. Lees, Pharmacodynamics and chiralpharmacokinetics of carprofen in calves, Br. Vet. J. 152 (2) (1996) 183–198, https://doi.org/10.1016/S0007-1935(96)80073-X.

[48] M.F. Landoni, P. Lees, Pharmacokinetics and pharmacodynamics of ketoprofen enantiomers in calves, Chirality 7 (8) (1995) 586–597, https://doi.org/10.1002/chir.530070806.

[49] P. Lees, P.M. Taylor, F.M. Landoni, A.K. Arifah, C. Waters, Ketoprofen in the Cat: Pharmacodynamics and Chiral Pharmacokinetics, Vet. J. 165 (1) (2003) 21–35, https://doi.org/10.1016/S1090-0233(02)00168-5.

[50] MF Landoni, W. Comas, N. Mucci, G. Anglarilli, D. Bidal, P. Lees, Enantiospecific pharmacokinetics and pharmacodynamics of ketoprofen in sheep, J. Vet. Pharmacol. Ther. 22 (6) (1999) 349–359, https://doi.org/10.1046/j.1365-2885.1999.00209.x.
[51] H.K. Knych, R.M. Arthur, S. Steinmetz, D.S. McKemie, Pharmacokinetics of ketoprofen enantiomers following intravenous and oral administration to exercised Thoroughbred horses, Vet. J. 207 (2016) 196–198, https://doi.org/10.1016/j.tvjl.2015.09.018.
[52] P. Zou, et al., Applications of human pharmacokinetic prediction in first-in-human dose estimation, AAPS J 14 (2) (2012) 262–281, https://doi.org/10.1208/s12248-012-9332-y.
[53] Ý.M. Tich, V. Trcka, Z. Roth, A.M. Krivucov, QSAR analysis and data extrapolation among mammals in a series of aliphatic alcohols, Environ. Health Perspect. 61 (1985) 321–328, https://doi.org/10.1289/ehp.61-1568771.
[54] P.-Y. Wu, et al., Development of machine learning-based quantitative structure-activity relationship models for predicting plasma half-lives of drugs in six common food animal species, Toxicol. Sci. p. kfae125 (2024) Sep, https://doi.org/10.1093/toxsci/kfae125.
[55] I. Cortes-Ciriano, A. Bender, T.E. Malliavin, Comparing the influence of simulated experimental errors on 12 machine learning algorithms in bioactivity modeling using 12 diverse data sets, J. Chem. Inform. Model. 55 (7) (2015) 1413–1425, https://doi.org/10.1021/acs.jcim.5b00101.

Glossary

AAFE: absolute average fold error
AS: allometric scaling
BW: body weight
CL: clearance
FE: fold error
GBM: stochastic gradient boosting
IV: intravenous
ML: machine learning
MLR: multiple linear regression
PBK model: physiologically based kinetic model
perc < 2fold: percentage of predictions with fold error lower than 2
QSAR: quantitative structure–activity relationship
RF: random forest
RMSE: root mean squared error
R2: coefficient of determination
SM: supplementary material
SVMPOLY: support vector machine with polynomial kernel
XGBTREE: extreme gradient boosting