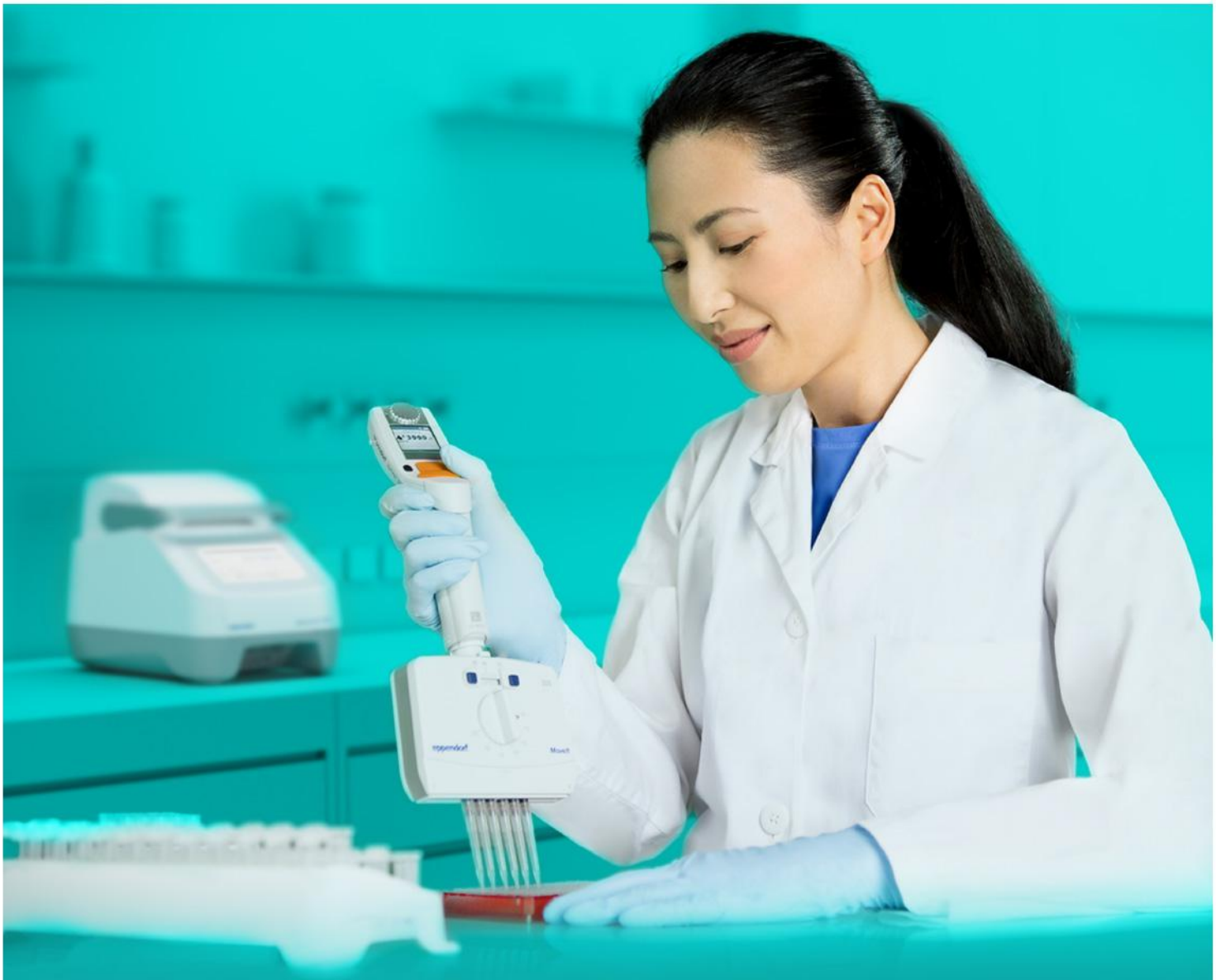
eppendorf



Reach Your 100%

Your success is our purpose

Your time is precious not only in terms of cost. However, a drug's journey from concept to patient takes years. You need to rely on products and services that ensure maximum efficiency at every step.

Our high-quality solutions are all engineered around the kind of functionality and intuitive performance necessary to speed up workflows, secure reproducibility, and accelerate the desired results.

Let's work together to make our world a better place.

www.eppendorf.com/pharma

Received: 21 October 2020 | Revised: 8 February 2021 | Accepted: 12 February 2021

DOI: 10.1111/bcp.14801

REVIEW ARTICLE - THEMED ISSUE

Machine learning in pharmacometrics: Opportunities and challenges

Mason McComb[1] | Robert Bies[1,2] | Murali Ramanathan[1,3]

[1]Department of Pharmaceutical Sciences, University at Buffalo, University at Buffalo, State University of New York, Buffalo, NY, USA

[2]Institute for Computational Data Science, University at Buffalo, NY, USA

[3]Department of Neurology, University at Buffalo, State University of New York, Buffalo, NY, USA

Correspondence
Murali Ramanathan, 355 Pharmacy Building, Department of Pharmaceutical Sciences, State University of New York, Buffalo, Buffalo, NY 14214-8033, USA.
Email: murali@buffalo.edu

The explosive growth in medical devices, imaging and diagnostics, computing, and communication and information technologies in drug development and healthcare has created an ever-expanding data landscape that the pharmacometrics (PMX) research community must now traverse. The tools of machine learning (ML) have emerged as a powerful computational approach in other data-rich disciplines but its effective utilization in the pharmaceutical sciences and PMX modelling is in its infancy. ML-based methods can complement PMX modelling by enabling the information in diverse sources of *big data*, e.g. population-based public databases and disease-specific clinical registries, to be harnessed because they are capable of efficiently identifying salient variables associated with outcomes and delineating their interdependencies. ML algorithms are computationally efficient, have strong predictive capabilities and can enable learning in the big data setting. ML algorithms can be viewed as providing a computational bridge from big data to complement PMX modelling. This review provides an overview of the strengths and weaknesses of ML approaches vis-à-vis population methods, assesses current research into ML applications in the pharmaceutical sciences and provides perspective for potential opportunities and strategies for the successful integration and utilization of ML in PMX.

KEYWORDS
artificial intelligence, drug delivery, machine learning, modelling and simulation, pharmacodynamics, pharmacokinetics, pharmacometrics

1 | UNMET RESEARCH CHALLENGES AND OPPORTUNITIES IN PHARMACOMETRICS

Pharmacometrics (PMX) is defined by the US Food and Drug Administration (FDA) as "*the science that quantifies drug, disease and trial information to aid efficient drug development and/or regulatory decisions*".[1] A wide range of quantitative techniques are used in drug development and can be considered under the PMX umbrella.

The explosive increase in new technologies and data throughout drug discovery and development has created challenges and opportunities for PMX. There is now an unmet need for innovative quantitative PMX methods that can provide insight and knowledge from the new data to accelerate and enhance drug development. There is great interest in integrating modelling and simulation, machine learning (ML) and artificial intelligence (AI) to leverage the increasing availability of *big data* to improve patients' outcomes to drug therapy.

In principle, ML and AI act as a bridge (Figure 1) between big data and pharmacometrics. ML provides computationally efficient approaches capable of processing big data that have powerful prediction and learning capabilities. ML can therefore be leveraged to handle the large sample sizes of big data to enable learning, hypothesis generation and model building. Integration of high-quality big data through ML and AI can supplement the smaller sample sizes used in PMX modelling and expand its parameter inference capabilities. In this review, we will critically investigate the utility of incorporating ML techniques in PMX for improving personalized medicine and drug development.

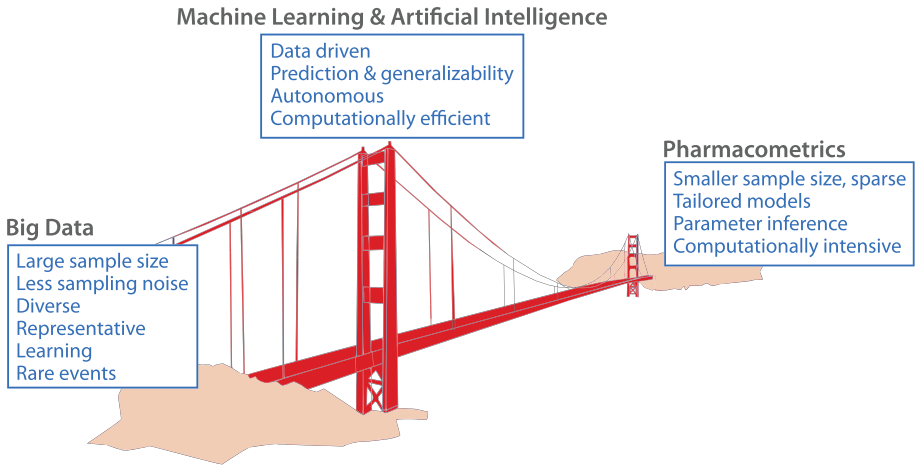


FIGURE 1 Machine learning (ML) is represented as the computational bridge between big data and pharmacometrics. The strengths and weaknesses of each area is shown in the text. Big data requires computationally efficient methods but enables learning. Because of the larger sample size of big data, the impact of sampling noise is reduced, better assessments of variability are obtained and it is more likely to contain indicators of low frequency and rare events. Integration of representative and high-quality big data with pharmacometric modelling via ML could provide insights into drug effects in diverse populations and in racial and ethnic minority groups. The high dimensionality of big data could enable consideration of a range of demographic, clinical, laboratory, environmental and genetic factors that contribute to drug outcomes. ML methods have strong generalizability and predictive capabilities and the integration of big data through ML could complement the pharmacologically informative findings and parameter inference capabilities from pharmacometric modelling, which typically have smaller sample sizes

2 | PHARMACOKINETIC AND PHARMACODYNAMIC MODELLING

Pharmacokinetic (PK) and pharmacodynamic (PD) modelling is the benchmark approach in the preclinical stages of drug development for describing time courses of drug concentrations and effects, interspecies scaling, and dose determination. PK/PD modelling relies primarily on compartmental models that are described through closed form algebraic solutions and/or systems of ordinary differential equations (ODEs). The ODEs are built on mass balance and information flow considerations that can be expressed in state–space form, which enables output noise to be incorporated.

Parameter estimation in PK/PD and population PK/PD modelling involves selection and subsequent optimization of an objective function that assesses the *goodness-of-fit* of the values from the proposed statistical model to the data. Examples of such objective functions include least squares, weighted least squares and log-likelihood functions. Maximum likelihood estimation (MLE, which minimizes the negative of the log-likelihood objective function) is the most widely used approach for obtaining system parameters in PK/PD modelling.[2–5]

Building PK/PD models is an iterative, empirically guided process that uses the qualitative features of time profiles, while also relying on principles of pharmacology and physiology to quantitatively represent the system. The complexity of these models is reduced to ensure parameter identifiability and the model selection process involves using measures of parsimony (e.g. Akaike information or AIC, Bayesian information criterion or BIC).

Data-driven metrics are also used as descriptors of PK time profiles in drug development. The mathematical framework in non-compartmental analysis (NCA) is the estimation of the statistical moments of the time profiles. NCA is typically used to assess bioavailability and bioequivalence.

Physiologically based PK (PBPK) models are anatomically based compartmental PK/PD models whose intercompartmental connectivity is based on physiological blood flow between organs. Because of their increased complexity, the majority of the system-specific parameters in PBPK models are fixed to physiological values to enable identification of the drug-relevant parameters of interest.[6–8] PBPK modelling can be viewed as a modelling platform and has found a useful niche for modelling antibody drugs.[9]

3 | POPULATION PK/PD

Population PK/PD is the most widely used approach for modelling PK/PD data from human studies.[2–5] The FDA provides a roadmap for population PK analyses in new drug applications and biologics license applications.[10] In PMX, population PK/PD is the key approach for enabling dose individualization and personalized medicine in the clinical setting. Population PK/PD is useful for modelling of sparse data sets and for integrating information from multiple studies.[11,12]

Nonlinear mixed effects (NLME) modelling provides a more elegant population modelling statistical framework by characterizing individual level data as a population.[13] NLME methods provide individual and population parameter estimates, and parse the interoccasion, interindividual (IIV) and residual components of variability. Calculating the NLME objective function is computationally intensive because the likelihood for each subject has to be integrated over parameter space during the optimization.[3]

4 | QUANTITATIVE SYSTEMS PHARMACOLOGY MODELLING

Vertically integrating PK/PD time profiles and clinical outcomes with the multitude of DNA, mRNA, protein and metabolite biomarkers and the corresponding biological pathways to construct multiscale system models has proven difficult.[14–16]

In PMX, the main strategy to overcome these challenges has been to re-purpose ODE-based compartmental models as quantitative systems pharmacology (QSP) models. QSP models are built by combining systems biology networks and pharmacodynamics modelling.[17] QSP models can have hundreds of compartments representing DNA, mRNA, protein and metabolite levels that are interconnected based on pharmacological knowledge of the activation and inhibition patterns that occur in signalling pathways of interest. QSP models are built with the goal of elucidating interactions between drug action, signalling pathways and the effector mechanisms.[16] QSP models require definition of a relevant biological network that defines interconnectivity of variables in the system and the corresponding pharmacologically relevant activation and inhibition relationships. QSP model building is subjective and requires expert insight to define the scope of biology and pharmacology. The quantitative behaviour of the network is assessed by using numerical methods such as Boolean networks,[18] agent-based models[19] or systems of ODE.[20] QSP models with high complexity are often pruned using heuristic approaches to reduce the variables in the system, and many parameters are fixed by the modeler. Connecting the model to clinical outcomes often requires link functions. QSP models have found utility as a mechanistic approach to qualitatively assess drug targeting, toxicity predictions and hypothesis generation in discovery and preclinical studies for drugs that target numerous interacting signalling pathways.

Despite the success of NLME in population PK/PD modelling, it has proven difficult to extend and translate that success to characterize individual and population profiles in larger pharmacological systems and to QSP models. While the computational resources needed for integration and optimization in NLME are indeed extensive, it is the underlying process of model development, which is highly dependent on human intervention and trial-and-error, that substantially limits the extension of NLME to, and utility for large complex systems modelling applications. Human intervention is critical and indispensable for evaluating the results from both small and large models. However, manually building complex yet parsimonious models that synthesize knowledge from multidimensional data sets is particularly challenging. QSP-based models take a bottom-up approach by using user-selected biochemical networks and pharmacological activation patterns for horizontal integration. However, vertical integration to higher-order pathological outcomes, generalizability of networks between tissues and to disease-states and parsimony and parameter estimation remain unresolved issues. Population modelling methods cannot currently be used in the QSP modelling setting despite the availability of rich data sets that could potentially enable integrated modelling across both modelling modalities. There is an unmet need for innovative PMX methods that can bridge the chasm between population PK/PD and QSP models.[21]

5 | MACHINE LEARNING AND PMX: OPPORTUNITIES, APPLICATIONS AND CHALLENGES

Advances in engineering and in computing hardware and software have had a transformative effect on pharmaceutical sciences and clinical research from the bench to the bedside. Massive amounts of data are being generated in the drug development setting due to advances in in silico chemical discovery and property prediction, high-throughput synthesis and screening, and bioanalytical methods including whole genome DNA sequencing, RNASeq, single molecular array assays and array-based mRNA expression profiling and mass spectrometric methods for small molecule quantitation, proteomics, methylation profiling and metabolomics. There has been concomitant impact of the explosive growth in medical devices, computing, communication and information technologies on healthcare. This has enabled new diagnostic tools, imaging techniques, electronic health records, telemedicine and digital medicine. These developments have created an ever-expanding data landscape that the PMX research must navigate.[22]

There is increasing interest in expanding the scope of PMX to integrate the rapidly growing clinical and biomedical datasets and the need for novel quantitative modelling techniques capable of learning across all areas of drug development.[23] Several recent reviews have attempted to provide a specific rationale for incorporating ML in population PK/PD modelling.[24–27] A consortium called Accelerating Therapeutics for Opportunities in Medicine (ATOM, https://atomscience.org) was funded by Cancer Moonshot programme of the federal government in 2016 to harness the potential of ML for drug discovery. ATOM involves the pharmaceutical company GSK, Lawrence Livermore National Laboratory of the Department of Defense, Frederick National Laboratory of the National Cancer Institute and the University of California, San Francisco. Its goal is to develop shared, publicly available tools for drug discovery. ATOM has developed an AI-based cheminformatics platform called ATOM Modeling Pipeline to historical drug discovery data on 500 failed GSK drugs, bioassay data and molecular properties.[28] The adaptation, integration and application of big data and ML in PMX has yet to reach fruition.

In the following sections, we provide an overview of ML techniques and critically assess the strengths and weaknesses of ML methods for drug development and PMX applications. The scope of the survey does not include extensive introduction and training in the science behind ML/AI given THAT there are numerous resources in textbooks and online tutorials.

6 | SURVEY OF ML METHODS

6.1 | Learning and ML

When large volumes of data are available, computer algorithms that are capable of learning become a natural and more powerful choice. As defined by Mitchell, "*A computer program is said to learn from experience E with respect to some class of tasks T and performance measure*

P if its performance at tasks in T, as measured by P, improves with experience E".[29] ML is an implementation of AIthat synthesizes knowledge and enables improvement of algorithms autonomously from data. Broadly, statistical learning involves estimating a function that defines the relationship between inputs (referred to as independent variables or predictors) and outputs (dependent variables, outcomes or responses).[30,31] In the context of big data, however, not all available predictor variables will be equally useful for optimal task performance because the set of useful or informative variables called *features* is unknown and will be present among irrelevant and redundant variables. ML is a subset of AI that utilizes computer algorithms to learn patterns from data to solve modelling problems. Feature selection algorithms identify informative features and are employed to preprocess data in ML. The main goal of ML is to identify patterns that are generalizable and predictive whereas the focus in statistics—and statistical modelling approaches such as NLME—is to enable inference regarding the population from sample data.[32]

The learning framework in ML algorithms can be categorized as supervised, unsupervised and reinforcement learning based on the nature and extent of feedback available for learning. Supervised learning identifies patterns from a training dataset when an outcome variable is provided. Classification and regression algorithms are supervised learning methods that relate the input variables to the predefined, labelled classes or outcomes, respectively. Population PK/PD modelling can be viewed as supervised learning process. Unsupervised learning is used to discover patterns in a dataset that does not include an outcome variable. Clustering, dimensionality reduction and probabilistic graphical modelling are examples of unsupervised learning methods. Reinforcement learning is an emerging ML method that is used for decision making based on systems of rewards and penalties.

6.2 | Supervised learning

Supervised learning provides models for regression and classification of labelled data. These models are then applied to new unlabelled data for prediction or to categorize their class. A simple and familiar supervised approach is logistic regression (LR), which can be used in ML for binary classification problems. The logistic function underlying LR expresses the log odds ratio of the labelled binary outcome variable as a linear combination of the input variables.[33,34] Pros and cons: LR coefficients are easily interpreted to provide assessments of feature importance; However, LR performs poorly in the presence of high levels of multicollinearity or if the binary classes in the data are difficult to separate with a linear boundary.

Algorithms that support both multigroup classification and regression include k-nearest neighbour (kNN), support vector machines (SVM), artificial neural networks (ANNs), classification and regression trees (CARTS) and naïve Bayes (NB) classifiers.

kNN is a nonparametric supervised approach commonly used for classifying multicategorical data.[35] kNN uses a lazy-learning method that evaluates distance of the test data points to the corresponding kNNs in the training set using a distance metric such as Euclidean, Manhattan, cosine distances, Pearson correlation or other metrics.[36] Pros and cons: kNN is a fast method; however, the kNN distance measure can render it sensitive to data scaling and normalization.

An SVM separates multidimensional data with a linear decision boundary or hyperplane that maximizes the margin between the decision boundary and the classified data.[37] The term support vector refers to the proximal data points that are critical for positioning the decision boundary. Kernel SVM extends to nonlinear boundaries by mapping the original data to a higher dimension using polynomial, sigmoidal or radial basis functions.[38] Support vector regression (SVR) uses a margin of tolerance around a decision boundary to provide predictions of the data.[39] Pros and cons: SVM can handle high dimensional data (i.e. when the number of dimensions exceeds the number of samples) well, but its performance is poor when the classes in data overlap.

6.2.1 | ANNs

Neural networks (NNs) are a particularly powerful and versatile ML approach whose mathematical framework is biologically inspired by the nervous system.[40–42] NNs are used for supervised learning and have broad applications in prediction, image analysis, natural language processing, facial recognition and autonomous vehicles.[43]

The quintessential NN is composed of 3 layers of nodes or neurons. Each node in the input layer contains several weighted connections (arcs) to the hidden layer of nodes, which in turn connects to the output with its weighted arcs. These arc weights are the parameters optimized in the learning process.[44] When >1 hidden layer is present, the NN is called a deep NN. The values at each node are obtained from the dot product of its input nodes and input weights, which are then scaled by a bias parameter and transformed through an activation function. Deep learning techniques use deep NNs to identify a hierarchy of features useful for the task. Pros and cons: While NNs are a powerful tool that can describe highly nonlinear relationships, the broader adoption of NNs in PMX has been challenging because they can be *black boxes*, which can limit interpretation and extraction of mechanistic process information.[45–50]

6.2.2 | Decision trees

Decision trees or CARTs, which were pioneered by Breiman,[51] provide an alternative ML approach to classification and regression problems that differs distinctively from traditional parametric approaches of statistics. A CART is a tree graph whose leaves represent the values of the discrete or continuous target outcome variable. The hierarchy of nonleaf nodes in a CART contain rules that iteratively partition the incoming values of the predictor variable along 1 of the edges until leaf nodes are reached.[51] CARTs provide the conceptual framework for several other ML algorithms such random forest regression (RFR), bagged and boosted decision trees, which reduce bias.

RFR is a powerful ML tool that is composed of several hundreds to thousands of CARTs, i.e. an ensemble or forest of trees, that are built from randomly bagged samples of the input data.[52] Although an individual CART may be biased, an ensemble of CARTs can provide superior classification and regression performance. Additionally, boosting algorithms are methods of weighting the ensemble of CARTs to further improve the performance of RFR.

Gradient boosted machines (GBM) are built from the training data that iteratively combines CART while reducing a loss function with each successive addition.[53] In adaptive boosting (AdaBoost), the weights on the training data are iteratively updated with each successive CART.[54] Extreme Gradient boosting (XGBoost) is an ensemble that uses several tools to form strong learners including CART hyperparameter weighting, regularization and nongradient-based optimization for loss functions.[55] Pros and cons: RFR can handle both categorical and continuous variables, conduct classification and regression and is effective at reducing overfitting; however, the algorithm is comparatively slow.

6.2.3 | Probabilistic supervised classifier

The NB classifier is a simple and efficient classification algorithm that utilizes Bayes' theorem with the strong assumption (hence, the *naïve* in NB) that the input variables are conditionally independent given the class label.[56,57] NB can be considered as a simplistic form of a Bayesian network (BN). The prior probabilities for the classes and the conditional probabilities can be set using the Bernoulli, multinomial and Gaussian distribution to accommodate binary, discrete and continuous outcome classification, respectively.[58,59] Pros and cons: NB can be surprisingly effective and powerful for many classification problems because it does not require traversing parameter spaces; however, it has limitation as a regression tool.[60] RFR outperforms NB.

6.3 | Unsupervised learning

Unsupervised classification encompasses a diverse range of commonly used clustering algorithms including *k*-means, hierarchical clustering, Gaussian mixture models and density-based spatial clustering of applications with noise (DBSCAN).

The *k*-means algorithm partitions data into *k* clusters. The value of *k* is provided by the user and the algorithm iteratively refines the centroid values of the *k* clusters until the within-cluster variance is minimized.[61] Pros and cons: the simple *k*-means algorithm is a special case of the expectation maximization algorithm that performs better when the clusters have equal variances and spherical shape. It is relatively easy to implement but requires a user-provided value for *k*. While there are approaches to select an appropriate value for *k*, this can be a problem particularly when the data sets to be analysed are not static.

Hierarchical clustering sequentially partitions the data into clusters.[62] Agglomerative and divisive clustering are 2 methods for hierarchical clustering that organize the data by successively grouping all individual data points or by iteratively dividing all data into smaller clusters, respectively. The distance metrics used to group data points in each cluster include Euclidean, Manhattan or cosine distance. Different clusters are grouped through complete, average or Ward's linkage measures.[44] Pros and cons: Unlike the *k*-means algorithm, hierarchical clustering does not require prior information on the number of clusters as input. However, it can be difficult to determine the appropriate number of clusters from the dendrogram used to present hierarchical clustering results. Hierarchical clustering is also computationally intensive (because of the pairwise distance or similarity calculations needed) and sensitive to noise and outliers.

Gaussian mixture models cluster the data using a weighted sum of distributions. MLE methods use either expectation-maximization (EM) or maximum *a posterori* to train the model to find the set of parameters for the distributions including the mean, covariance and the weights.[56,63]

6.4 | Unsupervised probabilistic graphical models

Probabilistic graphical models are an unsupervised learning ML method built on graph theory. Probabilistic graphical models represent the joint probability distribution as a graph whose nodes represent random variables and edges describe the interdependencies among the random variables.

6.4.1 | BNs

In other work,[64] we evaluated BNs, which are directed acyclic graphs (DAG) that model the joint probability distribution by factorization of all marginal and conditional distributions.[65,66] BNs are constructed in 2 steps, structural learning from the data to construct the DAG and parameter learning from the DAG.

Structural learning is typically implemented using constraint-based, score-based, or a hybrid of the 2 methods. Constraint-based methods rely on the notion of dependence separation (d-separation), where 2 variables are independent when evidence from a third variable breaks their association, while restricting the edges to as few as possible in the DAG.[67] In score-based algorithms, the DAG is constructed by iteratively adding and removing edges until an optimal score (on e.g. BIC, mutual information) is achieved. Parameter learning can be performed using MLE or Bayesian methods.[68] Structural and parameter learning can be applied in a training and testing model development strategy. Pros and cons: the DAG identified by BNs can provide a hypothesis-generation framework for exploring potential cause-effect relationships, but there can be other DAG structures that are equivalent in modelling the empirically observed data distribution; however, BNs cannot model cyclic dependencies such as those caused by feedback loops and are computationally intensive.

Markov networks are undirected graphs, which can be cyclic, whose cliques factorize the joint probability distribution.[69] As in BN,

the nodes of the Markov networks represent random variables and its edges represent probabilistic interactions.[70] Factor functions are used to represent the interactions between random variables, and the product of local weighted factor functions is used to represent the joint probability distribution. Pros and cons: Markov networks allow for more flexibility, but parameterization of the network is less intuitive and more computationally intensive than BN.

7 | REINFORCEMENT LEARNING

Ribba et al. review reinforcement learning (RL) approaches in PMX for precision dosing.[21] The underlying framework of RL is based on Markov decision processes and it is closely related to optimal control theory. The objective in RL is to maximize the value function to a controller or agent that exerts actions to steer the state of its environment (the system) toward a target. Controller actions that alter the state of the system appropriately (inappropriately) are rewarded (penalized) and accrue in the value function.[71] The fundamental concepts of RL can be easily mapped to clinical decision processes in precision medicine wherein the medical professional can be viewed as the controller or agent, the patient or patient population as the system, the treatment and dosing decisions as actions, the disease state or biomarker as the state and the reward function as clinical utility. RL readily lends itself to temporal processes because the agent and the system interact in a sequential manner that results in short-term changes to the state and long-term optimization of the value function. Pros and Cons: ODE-based PK/PD models are readily implemented in RL,[72,73] but the sample size requirements for reinforcement learning are high.[21]

8 | PMX *VS*. ML MODELLING

Table 1 provides a high-level summary of the contrasts between PMX and ML modelling that need to be considered to effectively integrate the 2 modelling strategies.

TABLE 1 Comparison of modelling strategies in pharmacometrics *vs*. machine learning

Characteristic	Pharmacometrics	Machine learning
Goal	Parameter estimation and variability characterization	Prediction
Foundation	User driven, data driven or mechanistic	Data driven
Workflow	Iterative workflow, manual review and curation	Structured workflow and automated curation
Method	Fit for purpose, restricted scope	Generalizable and versatile
Updating	Static: new models required	Dynamic: learning and retraining

The main goal in PMX is inference of parameters of the structural, variability and covariate model parameters in the population, whereas the goal of ML is prediction of outcomes. The PMX structural model is user-driven and guided by the modeller's (domain) knowledge of mechanism. ML approaches are data driven and some techniques, e.g. NNs can have excellent prediction performance, but can be perceived as black boxes because it is difficult to extract mechanistic insight.

Table 2 compares the data available for PMX modelling to the big data problems that ML algorithms excel at solving. Big data are colloquially described by the 6 Vs, volume, velocity, variety, veracity, variability and value. Volume refers to the size of the datasets; velocity refers to speed at which the data are created, processed and analysed; variety is the heterogeneous structure that arises from combining multiple data sources; veracity pertains to the robustness, reliability and quality of data and collection methods; variability is related with the stochasticity and noise of the data; and value is utility of the information from the data.[75]

The most obvious difference between ML and PMX data sets is the volume. The volume or size of the data sets can be characterized by the number of dimensions (d) and the number of observations (n), which is dependent both on the number of subjects and the number of time points (T).

As a rule-of-thumb, a data set can be considered *big* when $n > 1000$ and $d > 50$.[74] Additionally, a data set can be *tall* when it contains large number of observations and a small number of predictor variables (small d, large n), and *wide* when it contains a small number of observations and a large number of predictor variables (large d, small n) or both (large d, large n). PMX data sets are much smaller and more sparse than the big data required for ML. The choice of ML methods within big data can depend on whether it is *tall* or *wide*.[76]

PMX and clinical data are usually generated at a much slower velocity and, likewise, variety is typically limited in drug development datasets, where serially sampled drug concentrations, selected biomarkers and endpoints are typically measured. Patient specific

TABLE 2 Comparison of pharmacometric data to *big data*

Definition	Pharmacometric data	Big data
Volume	Small	Large
Sample size, n	$5 < n < 1000$	$n > 1000$
Time points, T	$5 < T < 100$	Variable
Dimensions, d	$5 < d < 50$	$d > 50$[74]
Velocity	Batch acquisition	Automated acquisition
Variety	Limited, structured, homogeneous, correlated	Variable, heterogeneous
Veracity	High integrity	Variable, often user dependent
Value	Directly relevant to outcomes	Inconsistent

covariates are also included but may not be monitored throughout a study. The veracity of the data collected for PMX modelling is very high, with collection methods being controlled by trained individuals supported by standard operating procedures. While interoccasion variability and IIV and instrument (residual) variability is present in PMX data, the primary goal of population PK/PD modelling is to characterize and enable population inference regarding these sources of variability. Many sources of big data have very high signal-to-noise ratio as they may be obtained from using high quality collection methods that digitally recorded events from calibrated sensors. In addition to making learning possible, the large sample sizes in big data significantly reduce sampling noise, which is dominant in sparse PMX data. The population diversity and variability can also be better estimated by using ML methods on big data sets. However, care must be taken to avoid over-fitting in ML modelling, as it can compromise the generalizability and the predictive performance.

8.1 | Model validation

Model validation is an integral part of the ML method. The test–training approach wherein the available data is partitioned into train–test splits of 70:30, 75:25 or 80:20. The model of interest is trained on the larger subset of data, while the test subset is used to compare the performance of the model to competing approaches, typically using metrics such as mean square error (MSE), sensitivity, accuracy, area under the receiver operator curve (ROC), AIC or BIC. Importantly, learning curves, which plot method accuracy against training sample size are used to evaluate the effectiveness with which ML algorithms learn.

In k-fold cross validation, the data are divided into k parts, with the first $k - 1$ parts used to train the model, and kth part is used to test the model. This is repeated k times, and the average of the test scores across the k repeats is used as the model criteria. If $k = n$, where n is the number of observations, this is termed leave-1-out-cross-validation. The values of algorithm hyperparameters, which control the learning process in the ML method, are often *tuned* using k-fold cross validation sampling.

Bootstrapping is a resampling method in which m samples are randomly selected with replacement to create the training set and the unselected samples are used for testing. This process is repeated, usually from 1000–10 000 times, and the model criteria is averaged across all sets. Based on the asymptotic characteristics of bootstrap resampling, a ratio of 63.2% of the data set for training and 36.8% for testing is frequently utilized.[77]

Because PMX data are typically limited in size, these model validation methods are not routinely used during PMX model development. Furthermore, there is a lack of consensus on the definition and approaches for PMX model validation.[78] Model validation in PMX can be interpreted as assessing uncertainty in parameter estimates, goodness-of-fit plots and metrics. Additionally, the term validation can be used to describe how a model performs when challenged with a new external dataset.[79]

PMX modelling and drug development problems differ in distinct ways from the big data problems at which ML methods have excelled and proven their mettle. It is therefore important to systematically identify specific ML methods best suited for individual application areas in drug development and to adapt and refine them.

9 | APPLICATIONS OF ML IN PHARMACEUTICAL SCIENCES

Now that we have formalized the distinctive approaches used in ML and PMX, we review the trends, applications, limitations and promise of ML in pharmaceutical sciences.

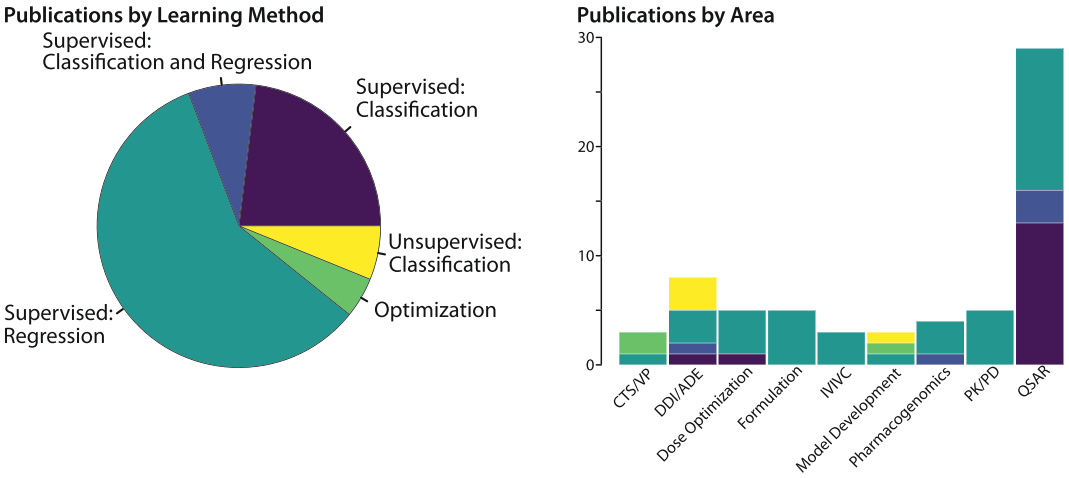


FIGURE 2 Publications of machine learning applications in pharmaceutical sciences from1995–2020. The publications by learning method (pie chart, left) show the relative proportions of machine learning methods: supervised, unsupervised or optimization. Additional dissection of supervised and unsupervised machine learning include classification, regression, or both applications. The bar graph (right) shows the number of publications (y-axis) by area in pharmaceutical sciences (x-axis). CTS, clinical trial simulation; VP, virtual population; DDI, drug–drug interaction; ADE, adverse drug event; IVIVC, in vitro in vivo correlation; PK/PD, pharmacokinetics/pharmacodynamics; QSAR, quantitative structural activity relationship

As of July 2020, a Web of Science search of "machine learning" nested within the search of "pharmacokinetics or pharmacodynamics" yielded over 100 publications. The publications were categorized to 1 of the following areas of the pharmaceutical sciences: PK/PD, dose optimization, quantitative structure–activity relationship (QSAR); adverse drug event (ADE) prediction and drug–drug interactions (DDIs) and clinical trial simulation (CTS). Figure 2 summarizes the categorization of 65 ML publications in pharmaceutical sciences from 1995–2020. Supervised classification, regression, or a combination of the 2 approaches are the most frequently used ML methods and QSAR emerged as the most frequent application area. Table S1 summarizes the pharmaceutical sciences applications and ML algorithms that have been investigated in the literature.

As noted, previously several recent reviews have focused on identifying the most promising and appropriate application areas and ML tools.[24–27] Talevi et al. provided an accessible primer on ML concepts that included a case study that employed RFR for investigating the structure–activity relationships of inhibitors of the **putrescine** transporter of trypanosome parasites.[27,80] Hutchinson et al. proposed an implementation framework with 2 hypothetical examples that utilized deep learning to perform global parameter sensitivity analysis and for combining PK parameters with imaging and omics data.[25] Koch et al. used CART-based ML approaches for covariate selection in the context of a simulated data and clinically relevant example of phototherapy for bilirubinaemia in neonates.[81] In a commentary, Chaturvedula et al. highlighted the use of genetic algorithms (GAs) for model selection in population modelling and deep learning for target identification in drug repurposing.[24]

9.1 | ML in PK/PD and PopPK

The earliest applications of ML in pharmaceutical sciences investigated whether NN methods could be a surrogate for traditional PMX modelling.

9.1.1 | ML studies of the PK/PD and PopPK of antibiotics

PK data for the antibiotics: **gentamycin**, **tobramycin** and **arbekacin** have been assessed using NN methods.[82–85]

NNs performed as well or better than NLME models for predicting peak concentrations of gentamycin for 111 patients that included patient covariates (age, height, weight, body surface area, serum creatinine, and creatinine clearance).[82] The predictive performance of NNs was similar to NLME for steady-state peak and trough (16.5 *vs*. 18.6%) concentrations. However, NLME outperformed ANN when extrapolating beyond the measure dosage.

NN was used to predict the peak and trough plasma concentrations of tobramycin using data from 101 paediatric patients.[84] Dosing information (dose, dosing interval, time of blood drawn) and patient demographic variables (age, weight, sex, cystic fibrosis or cancer) were input into the NN. The predictive performance of NN (33.9% absolute error at dose initiation and 37.3% absolute error at steady state) *vs*. NLME (39.9% overall absolute error) were comparable.

The predictive performance of NN was compared to linear and logistic regression for dose efficacy optimization of arbekacin in 30 burn patients with the covariates (dose, BMI, serum creatinine, parenteral fluid amount, and burn severity).[85] NN outperformed both linear and logistic regression on all outcomes predictions based on AIC.

Smith et al. combined bench research, ML and mechanism-based mathematical modelling strategies to investigate optimal dosing strategies for treating carbapenem-resistant *Acinetobacter baumannii*.[86] The GA was used to optimize a combination dosing regimen, which resulted in a 19.6 g/d, 2-hour infusion every 5 hours of **meropenem** with 5.17 mg/kg/d every 6 hours of **polymyxin B**.

Feretzakis et al.[87] compared the performance of ML algorithms to assess the dependence of antibiotic resistance/sensitivity status on easily obtained demographic, bacterial and drug characteristics such as gender, age, type of antibiotic, Gram stain status and type of sample in a large hospital microbiology laboratory database. The goal was to identify the ML algorithms for predictive use in an intensive care setting. They found that NN had a modestly higher area under the ROC curve than RFR and kNN with $k = 5$. However, this descriptive report did not identify any of the characteristics contributing to antibiotic resistance/sensitivity status. ML has been used to mine molecular structure databases created for drug repurposing to enable antibiotic discovery: a **c-Jun N-terminal kinase**[88] inhibitor was identified as a candidate antibiotic using an NN trained on a dataset of known bacterial growth inhibitors and tested on a larger dataset of candidate drugs; its antibiotic properties against several resistant bacterial strains were evaluated in animal studies and its unique mechanism of action were identified.[89] These results are promising but research is needed before the broader utility of ML for antibiotic discovery via molecule repurposing can be considered established.

9.1.2 | ML studies of the PK/PD and PopPK of other drug classes

Population PK data for **tacrolimus**, an immunosuppressive drug used in organ transplantation that requires dose optimization because of its narrow therapeutic index and large IIV, has been investigated in several ML studies.[90–93] Tang et al. performed a head-to-head comparison of 8 different ML methods: NN; multiple linear regression; CART; multivariate adaptive regression splines (MARS); boosted regression tree; SVR; RFR; lasso regression; and Bayesian additive regression trees for tacrolimus dose optimization.[90] The analyses included several patient covariates, pharmacogenomic data and concurrent therapy. CART outperformed all other ML methods with a mean absolute error of 0.73 (95% CI: 0.63–0.82). Gim et al. compared CART, RFR, and least absolute shrinkage and selection operator for determining the influence of single nucleotide polymorphism (SNP) for predicting of tacrolimus Cmax and AUC in healthy Korean males.[91] Their results

showed that all ML methods identified CYP3A5 SNP, *rs*776746 as the best predictor of tacrolimus exposure. Other research groups have also assessed the potential of ML methods for tacrolimus dose optimization.[92,93]

ML methods has also been applied to digoxin dosing due to its narrow therapeutic range and DDIs, which can be easily preventable with active monitoring. The appropriateness of the initial dosing of **digoxin** was assessed posthoc from 307 Taiwanese inpatient records using 10 predictor variables, while serum digoxin concentrations above <0.9 ng/mL were classified as inappropriate and used as the response variable.[94] The predictive performance was assessed using the ROC AUC. RF showed the greatest predictive performance (ROC AUC = 0.912) followed by MLP (0.813), CART (0.791) and C4.5 (0.784). Thus, it was concluded that the initial dose of digoxin can be accurately predicted for increased safety.

Using an ensemble approach that included ML techniques with NLME was investigated for predicting the concentrations of the analgesic **remifentanil** in 30 patients with sex, weight, height, age, BSA, lean body mass, infusion rate, infusion duration and remifentanil dose as covariates. The average predictions from the ensemble were compared to each method alone.[95] The NN-NLME ensemble (MSE = 55.17) outperformed NLME (MSE = 95.77), NN (MSE = 57.12), SVM (MSE = 181.20) and the NN-NLME-SVM ensemble (MSE = 70.80).

9.2 | ML applications in QSAR

QSAR methods are used for lead optimization in early drug discovery.[96] The predictive performance of ML algorithms makes it an attractive alternative to statistical regression-based approaches for QSAR analysis. Additionally, QSAR analyses are well suited and easily implemented with ML based regression and classification methods.

For example, Mishra et al. utilized QSAR to extend Lipinski's[97] molecular descriptors based on a catalogue of 60 000 molecules.[98] Cortes-Ciriano used compared the predictive performance of: GBM, partial least squares, RFR and SVR for predicting the potency of a broad range compounds from their physiochemical descriptors.[99] QSAR has also been used to predict plasma protein binding,[100–104] blood brain barrier permeability,[105–107] clearance,[108–110] volume of distribution,[111–113] half-life,[114] bioavailability[115] and toxicity.[116] ML-based QSAR has shown utility in in drug metabolism and PK to predict compound metabolism,[117,118] transporter transcriptional upregulation,[119] uptake, efflux and inhibition.[120–126]

9.3 | ML for DDIs

DDIs are a serious concern in drug development that lead to approximately 30% of clinical ADE.[127] ML approaches are well suited for identifying complex associations from large databases and have been used for predicting DDIs.

The FDA Adverse Event Reporting System (FAERS) contains reports of ADE[128] to support the FDA's postmarketing safety surveillance programme for approved drug and therapeutic biologic products (https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard). The FAERS outcomes are death, life threatening, hospitalization (initial or prolonged), disability, congenital anomaly, or require intervention (if medical or surgical intervention was required to prevent permanent damage) and other. Several groups have utilized the FAERS in conjunction with other public-domain data sources as input data for ML methods to gain insight into potential DDIs. Zhang et al.[129] integrated information from FAERS, PubChem[130] for chemical structure information, DrugBank[131] for drug targets, enzymes, and transporters, KEGG[132] for biological pathway information and SIDER[133] for side effect information. Several ML approaches have been utilized, including: convolutional neural networks,[134] SVM, kNN and NN,[135] NB and CART,[136] GBM,[137] and RFR.[138]

While the FAERS data sets are large, the utility of these approaches is limited by data quality. The limitations include reporting bias, the coarse granularity of the demographic and clinical information and the nonrepresentation of treated subjects who did not experience adverse events. FAERS data are retrospective and are biased because report submission is voluntary. The FDA emphasizes that the database should not be used to obtain incidence statistics for adverse events. The FDA does not require a causal relationship between an adverse event to be established or proven for inclusion; there may be insufficient information for evaluating even serious adverse events and it is not possible to trace multiple AER on the same patient over time. The FAERS database also presents drug information in the vocabulary used by the submitter and therefore contains spelling errors, trade names and other variations that complicate direct utilization of the data. Text-mining of the literature has also been used to identify DDIs.[139,140]

9.4 | ML for CTSs

CTS is a method to increase the potential success of clinical trials by identifying and evaluating design and implementation inefficiencies to reduce the time, resources and financial burdens in drug development.[141,142] Allen et al.[143] described a method to generate virtual patients that simulates a biologically plausible range of output from a desired model followed by the use of a modified sum-of-squared errors cost function that compares the simulated values to real-world data to obtain a new virtual patient population. This method was further developed to harness used ML global optimization algorithms, including: simulated annealing and nested simulated annealing; Metropolis–Hasting; and GA.[144,145] Additionally, new approaches for translating the information gained from randomized controlled trials to specific target populations using RF are being developed.[146]

10 | INTEGRATION OF ML IN PMX

10.1 | Covariate selection

Covariate selection is an area of population modelling within PMX model development that may be promising for incorporating ML methods. Statistical methods for covariate analysis are constructed on linear regression-based selection methods such as stepwise regression, all subset regression, ridge regression and least absolute shrinkage and selection operator.[147]

Previous work in our group demonstrated a novel approach to integrate pharmacogenomics data in PK/PD modelling using information theoretic approaches.[148] This method was used to simultaneously evaluate gene–environmental interactions using PK/PD, clinical outcomes and genome-wide pharmacogenetic data. Novel and known interactions between warfarin and gemcitabine were identified using the *K*-way interaction information metric.

Koch et al. considered simulation scenarios where the implementation of CART assisted in the selection of both time dependent and time independent ML analysis.[81] For their time-independent analysis, 1-compartment models with a defined covariate relationship between elimination rate constant (k_{el}) and simulated covariates were simulated. NCA was used to calculate half-lives for all 23 patients and dichotomous outcome variable based on a half-life threshold was obtained. CART was used to identify the most influential covariates (i.e. risk factors) associated with higher half-life without knowledge of the model used to generate the data Koch's time-dependent example included age as an additional predictor variable in the decision tree model. They found that CART was useful for identifying and ranking the covariates.

Hall et al. applied MARS, a nonparametric piecewise ML method for identifying linear and nonlinear relationships, for covariate selection.[149] MARS identified ranges of weight and age that were predictors of absorption, clearance and volume of distribution.

10.2 | ML approaches for modelling and optimization

The earliest studies of ML evaluated NN strategies. The predictive performance of NN was compared to population-based NLME[82] and individual level PK/PD[83] models. Using simulated datasets from PK-linked PD models, Gobburu et al.[83] found that model misspecification was abrogated by using NN as a model-independent approach to predict PK/PD concentration–time profiles. However, NN performed poorly with sparse and noisy data.

Bies et al. used GA for automated model selection including choice of covariate relationships, intervariability models, and residual models.[150] GA outperformed stepwise covariate modelling approaches based on lower objective criteria.

In a novel strategy, Chen et al.[151] incorporated the neural network architecture for describing nonlinear relationships into the ODE framework. Rachkauckas et al.[152] incorporated this as a feature in the Julia pharmaceutical sciences package. While this method offers the potential to model inherently unknown relationships within a network of ODE, it may carry a level of overfitting risk.

Bunte et al. evaluated combined parameter estimation and clustering approach for population PK modelling of **prednisolone**.[153] The parameters for a 2-compartment model of prednisolone were obtained using MLE and combined with Gaussian mixture modelling for clustering with the EM algorithm.

10.3 | ML for precision medicine

There is emerging interest in using RL for designing strategies for precision medicine.[21] RL has been investigated in the context of treating sepsis,[154] anaemia in renal failure patients on haemodialysis,[155–159] **propofol** anaesthesia[72] and chemotherapy.[160]

11 | CASE STUDY

In the following section, a simulation case study is performed to demonstrate the utility of unsupervised ML methods to guide PMX modelling. The goal of this case study was to: (i) identify *target* biomarkers to guide efficacy of a drug candidate; (ii) propose a structural PK/PD model for further analysis; and (iii) identify potential subject variability for covariate modelling.

11.1 | PMX simulation method and results

Concentration–time PK profiles were simulated for 200 subjects with the 2-compartment model (Figure 3A) after oral administration of 1000 mg of drug. The typical value for the population clearance was simulated with a linear covariate model:

$$\mathrm{TVCL} = 0.35 + 0.03 \times \mathrm{COV1} + 0.075 \times \mathrm{COV2}$$

where *COV*1 and *COV*2 were subject-specific covariates generated from normal distributions with a mean of 20 and 50 units, respectively, with 20% CV. Between-subject variability was applied to *CL* and the volume in the central compartment (*V*) using exponential variation with η set to 0.16. Residual variability of 20% was included on the drug plasma concentration (C1) using a proportional error model. The 200 simulated drug plasma concentration (C1) are shown in Figure 3B.

C1 from the PK model was linked to the PD model through an initial biomarker (*B1*, Figure 3C) using an indirect response model with stimulation of the *B1* production rate ($k_{in,B1}$). The concentration of *B1* was linked to a second biomarker (*B2*, Figure 3D) by similar stimulation of the *B2* production rate ($k_{in,B2}$). Finally, the concentration of *B2* was linked to a third biomarker (*B3*, Figure 3E) using an indirect response model with inhibition of the production rate constant ($k_{in,B3}$).

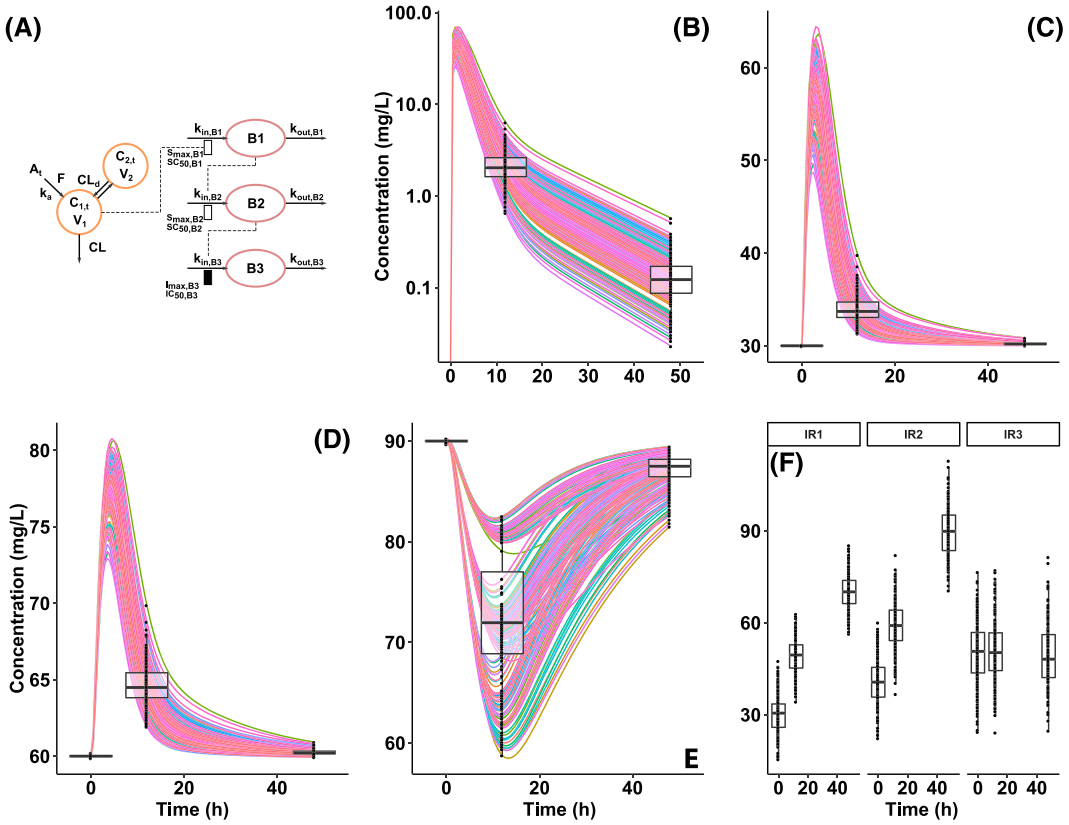


FIGURE 3 Pharmacometrics simulation results for subsequent machine learning modelling methods. (A) Schematic of a 2-compartment pharmacokinetic model (yellow circles) linked to a pharmacodynamic model (peach ovals) composed of 3 indirect response models that modulate biomarkers *B1*, *B2* and *B3*. (B) Simulated concentration–time profiles from the pharmacokinetic model for the 200 subjects (coloured lines, log base10 scale), with the *observed* data indicated by the black dots and box plots at 0, 12 and 48 hours. The pharmacodynamic model simulations for B1 (C), B2 (D) and B3 (E) for the 200 subjects (coloured lines, log base10 scale) are shown, with the *observed* data indicated by the black dots and box plots at 0, 12 and 48 hours. (F) The simulated irrelevant biomarkers, *IR1*, *IR2* and *IR3* serve as the *observed* negative control biomarkers indicated by the black dots and box plots at 0, 12 and 48 hours

The maximum fractional extent of inhibition for *B3* ($I_{max,B3}$) for each subject was also provided using a covariate model:

$$I_{max,B3} = 0.3 + 0.3 \times SNP1 + 0.03 \times SNP2$$

where *SNP*1 and *SNP*2 were binary variables indicating the presence of an SNP. Four groups based on the genotype combinations *SNP*1 and *SNP*2 were generated, producing 3 phenotype groups with an $I_{max,\ B3}$ of 0.3, 0.6 or 0.9. The patterns in the time profiles of *B3* are shown in Figure 3E.

Three additionally irrelevant variables (*IR1*, *IR2* and *IR3*) were simulated at 3 time points to act as negative controls to assess ML modelling. The mean at times 0, 12 and 48 hours were set to 30, 50 and 70 mg/mL for *IR1*, 40, 60 and 90 mg/mL for *IR2*, and 50, 50 and 50 mg/mL for *IR3*, and all with 20% CV (Figure 4F).

11.2 | ML modelling methods and results

Three time points (0, 12 and 48 h) were selected for each *C1*, *B1*, *B2*, *B3*, *IR1*, *IR2* and *IR3* from the PMX simulations to serve as input data for ML modelling. These are highlighted as box plots in Figures 1B-F.

Unsupervised RF using the *randomForestSRC* R package with default hyperparameters was used to identify the candidate target variables for modelling. The model was fit to the entire observed dataset, excluding the concentration time points. Minimal depth was used as a variable importance measure where a threshold of 5, determined by the mean of depth distribution throughout the forest, was used to characterize variables as important if they were smaller than the mean depth.[161] $B2_{t\ =\ 48}$, $B1_{t\ =\ 48}$, $B1_{t\ =\ 12}$, $B2_{t\ =\ 12}$, $B3_{t\ =\ 48}$, $B3_{t\ =\ 12}$, $IR3_{t\ =\ 48}$, $B1_{t\ =\ 0}$, $B2_{t\ =\ 0}$, $IR3_{t\ =\ 12}$, and $IR1_{t\ =\ 0}$, were identified as candidate variables (Figure 3A).

Structural BN modelling was performed to construct a graphical model of the interdependencies of the candidate biomarkers identified from the RF. Three structural BN models were fit to the candidate biomarkers at each of the observed times (0, 12 and 48 h) using the hill-climbing structural learning algorithm in the *bnlearn* R package. Biomarkers that did not contain associations (edges) in the graph were considered irrelevant, and dropped from subsequent analysis.

The BN at time 0 displayed no connections (no shown), probably due to 0 concentration for all subjects and limited differences of variables at baseline.

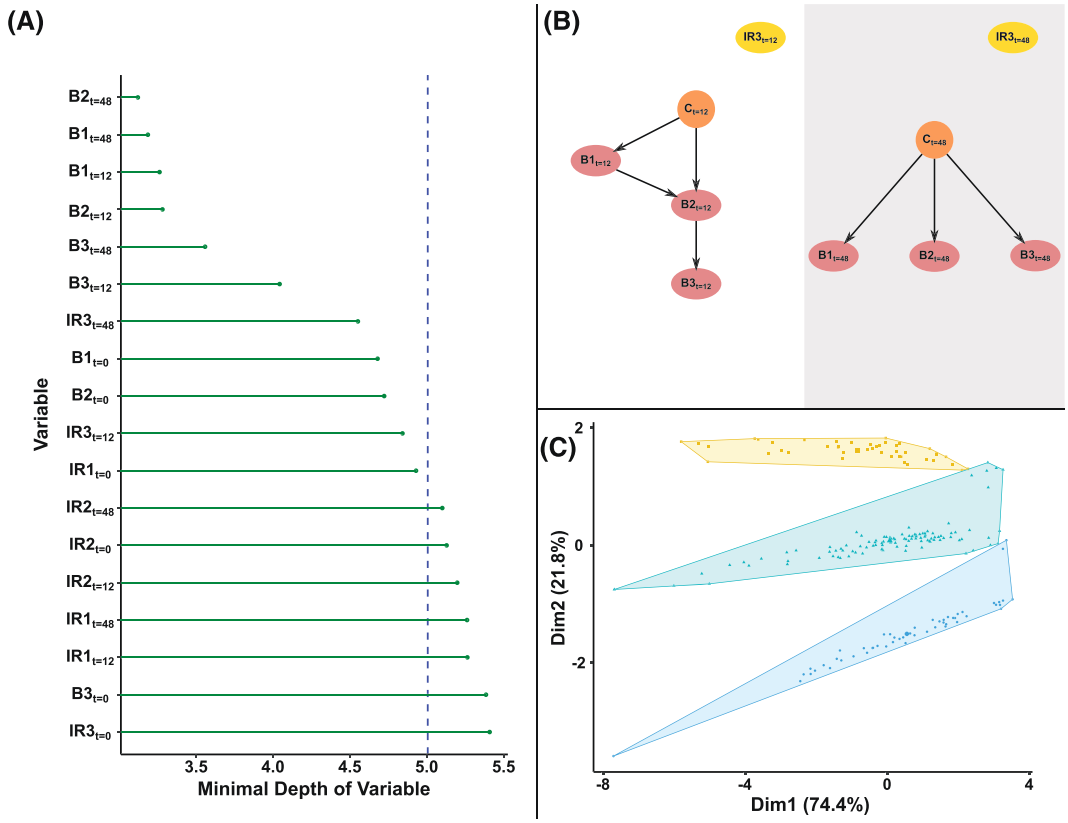


FIGURE 4 Machine learning modelling results from the *observed* pharmacometrics dataset using unsupervised machine learning methods. Minimal depth of all biomarkers from the unsupervised random forest (A, *y*-axis) where lower numbers (*x*-axis) indicate closer to the root node and more important, while variables that are higher than the threshold minimal depth (5, blue hashed line) indicate unimportant variables. (B) Resulting Bayesian network at 12 hours (left, white background) and 48 hours (right, grey background) show the associations of the plasma concentration (orange circles) with the true biomarkers (peach circles) and no associations with the irrelevant variables (yellow, ovals). K-means clustering identified 3 clusters (C, blue cluster 1, green cluster 2, yellow cluster 3) for all identified candidate biomarkers

The BN at 12 hours identified appropriate dependencies (Figure 4B) among the candidate variables. The association of $C_{t = 12}$ and $B1_{t = 12}$ highlights the PK/PD relationship between concentration in the plasma stimulating the production of $B1_{t = 12}$. The association between $C_{t = 12}$ and $B2_{t = 12}$ due to the concomitant increase in the levels of both variables at this time point. Furthermore, the catenary response from $B1_{t = 12}$ to $B2_{t = 12}$ to $B3_{t = 12}$ was identified and provides the insight into the original model. The irrelevant variable $IR_{t = 12}$ was not associated with any of the other variables.

The BN at 48 hours $C_{t = 48}$ showed associations with all true biomarkers and no associations to $IR3_{t = 48}$. However, the catenary associations among the biomarkers in the PD model were not identified because *B2* and *B3* returned to near baseline for all subjects at 48 hours. These results demonstrate that the time sequence of events and the associations can be captured using unsupervised ML approaches and leveraged during PMX modelling.

The set of biomarkers identified via the BN associations ($B1_{t = 12}$, $B2_{t = 12}$, $B3_{t = 12}$, $B1_{t = 48}$, $B2_{t = 48}$ and $B3_{t = 48}$.) were analysed with the K-means clustering algorithms using the *stats* R package. Elbow and silhouette plots confirmed 3 clusters were optimal (not shown). The 3 clusters from K-means were visualized using the first 2 principal components; principal component analysis was conducted using the *factoextra* R package. Figure 4C shows the 3 clusters that can be visualized using scatter plot of the first 2 principal components. This demonstrates utility of K-means for identifying subpopulations in a PMX dataset.

11.3 | Discussion

Unsupervised ML methods can aid pattern recognition within large datasets to identify the most relevant variables and guide in the PMX modelling building process in a data-driven manner. The 3 ML methods used in this Case Study were RF, BN and K-means clustering. The results show that candidate target biomarkers can be identified using RF, while model structures for subsequent PMX modelling can be suggested using BN and further insight into subject variability within the dataset can be obtained using K-means clustering.

12 | ENVISIONING THE FUTURE OF ML IN PMX

Given the rapid integration of ML and AI methods in other fields, it is very likely that these methods could have transformative roles across

every aspect of PMX and the drug development process. While recognizing the promise and the untapped potential of ML, it is also important to acknowledge the strengths, limitations and distinctive differences in the underlying capabilities of both established PMX vis-à-vis emerging ML methods in order to identify and define problems that are barriers on the critical path that need to be addressed in pharmaceutical sciences research.

A word about sample size requirements and power because among those in the PMX and pharmaceutical sciences community unfamiliar with ML, there is sometimes the unrealistic expectation regarding the effectiveness of ML methods. It should be emphasized that as in statistical testing, the sample size for a given power (which is equivalent to $1 - \beta$, where β is Type II error or the probability of false negatives) for ML methods are also determined by effect size, variability (covariance or correlation in the multivariate setting) considerations. An additional consideration in ML methods for high dimensional data is the so-called *curse of dimensionality* (wherein the addition of extra dimensions to a mathematical space exponentially increases the hypervolume in which the data is distributed), which increases the computational complexity and degrades the power of algorithms. Generally, learning curves for ML methods show improvements in performance that show a power law dependence on increasing training sample size before reaching a plateau.

For practitioners, it is often helpful have handy heuristics or rule-of-thumb for the sample size per candidate feature and also a sense of the power law exponent for the ML method's learning curve to guide data collection decisions. A widely used rule of thumb for sample size in linear regression is ≥10 samples per parameter in the model. A study by van der Ploeg et al.[162] compared the effective sample size requirements of ML methods for binary classification including SVM, NN and RFR to LR and CART for clinical data sets found that LR performed well once there were 20–50 events per variable whereas SVM, NN and RFR were more *data hungry* requiring nearly 10 times more data. The training data set sample size requirements for deep learning methods may be larger, e.g. in image classification applications, a rule-of-thumb of 1000 training images per class is typically employed.[163] However, the learning curves for deep learning methods appear to show improvement with additional training data even after traditional ML methods have reached plateau.[87,164–166] In addition, some caution may be warranted when using sample sizes estimated for ML algorithms in image and signal processing applications, which have data with high signal to noise ratio, to guide PMX modelling, which has multiple sources of variability.

Nonetheless, with a few exceptions, statistical power is often not carefully assessed in many ML applications in PMX and pharmaceutical sciences because the overall focus is to highlight the promise and predictive potential of ML. Sucheston et al.[167] systematically compared the power of their information theoretic algorithm for gene–environment interactions to LR and to multifactor dimensionality reduction method for a range of effect sizes and interaction models. They found that their information theoretic ML method had greater power than multifactor dimensionality reduction and was comparable to LR. More recently, our group has compared the power and false positive rate of our generalized pharmacometric modelling method, which combines RFR with BN, to enable covariate modelling in PMX.[64]

ML algorithms also have utility for exploratory analysis of high dimensional data in PMX even when sample sizes are small. For example, supervised and unsupervised clustering techniques are widely used for *omics* data and used to generate visually effective heat plots. However, it must be appreciated that the while the resultant models can indeed aid hypothesis generation, they may not be robust or useful for predictive purposes; some of the features identified can be expected to be volatile and difficult to replicate.

Figure 1, which envisions ML methods as a bridge between big data and PMX, was guided by the possibility that sample size constraints may potentially limit the utilization of currently available ML methods in stand-alone PMX applications that do not have adequate big data. Furthermore, the studies to date suggest that ML methods cannot supplant population PK/PD methods, which can consider the pharmacology of the drug and incorporate system physiology in the structural model. However, ML can complement population PK/PD modelling by serving as a bridge for modelling and learning from big data and by facilitating hypothesis generation. We posit that the predictive capabilities of ML in particular could have a notable impact on dose individualization and personalized medicine.

As a practical matter that enables generalizability, there is also a need to identify the most reliable and robust ML-based workflow for each PMX application. It may be necessary to analyse data with several appropriate ML algorithms and advance only those findings that are common to multiple ML methods. Our research results that use the results from RFR as inputs for Bayesian network modelling demonstrates the utility of considering a pipeline of ML methods.

It is worth considering the most promising niches in PMX modelling that could potentially benefit the most from ML-based strategies. In population PK/PD modelling, the covariate modelling component be an attractive target. For example, our work suggests that ML algorithms could be useful for building covariate models from high dimensional genotyping methods.[64,148] Similarly, large population-based big data sets could be used extrapolate covariate models built from small PMX studies to minority populations and also to populations that are more diverse. The FDA and the PMX modelling community have recognized that the characteristics of the patients enrolled in drug development clinical trials are quite frequently not representative of the population at large.[10,168] Another impactful area would be to consider the natural history of the biomarkers of physiological processes across the lifespan using well-curated and representative big data sets such as population-based National Health and Nutritional Examination Survey (NHANES, https://www.cdc.gov/nchs/nhanes/index.htm). Because ML methods are capable of handling correlated high-dimensional data and identifying the interdependencies amongst the salient variables, they may be capable of providing better approaches for creating representative *virtual patients* for clinical trial simulations. However, another potentially interesting but unexplored ML application is model-based meta-analysis in population PK/PD. It would be interesting to investigate whether ML can be explored for building a model for the population models for drugs in a given therapeutic class

to enable decision making in the regulatory setting where large amounts of the needed data might be available. Our results demonstrate that ML-based preprocessing of big data could have a transformative impact in guiding the development of QSP models.[64] ML methods can provide structural networks for QSP models that are guided by the interdependencies in biochemical and physiological biomarkers. Although an ML-based strategy would require good quality input data from population-based big data sets and clinical registries, it could reduce the subjectivity and bias present in the manually curated biochemical networks and literature review workflows that are currently used for building QSP structural models.

Although ML has the potential to have a transformative impact on PMX and drug development, the utilization of ML methods is at an early stage. Much research on the choice of the specific problems, algorithms, workflows and integration of findings is required to harness the promise and move the field forward. There is, however, a great deal of ongoing research interest in leveraging ML methods in the pharmaceutical sciences that could render the methods and references in this review outdated. The time is ripe to set up a consortium of industry, academic, contract research organizations and regulatory agencies to advance ML in the pharmaceutical sciences.

12.1 | Nomenclature of targets and ligands

Key protein targets and ligands in this article are hyperlinked to corresponding entries in http://www.guidetopharmacology.org, and are permanently archived in the Concise Guide to PHARMACOLOGY 2019/20 (Alexander et al., 2019 a,b).

CONFIDENTIALITY

Use of the information in this manuscript for commercial, non-commercial, research or purposes other than peer review not permitted prior to publication without expressed written permission of the author.

ACKNOWLEDGEMENTS

Support from Grant MS190096 from the Department of Defense Multiple Sclerosis Research Program for the Office of the Congressionally Directed Medical Research Programs (CDMRP) to the Ramanathan laboratory is gratefully acknowledged.

COMPETING INTERESTS

Mason McComb: None. Dr Robert Bies received research funding from the Food and Drug Administration and the National Institutes of Health and the Department of Defense. He serves as an expert consultant to Janssen Pharmaceuticals and has received research support from Merck and Company. Dr Murali Ramanathan received research funding from the National Multiple Sclerosis Society, Department of Defense, National Science Foundation and National Institute of Neurological Diseases and Stroke. He received funding for unrelated research from Otsuka Pharmaceutical and Development. He receives royalty from a self-published textbook.

DATA AVAILABILITY STATEMENT

Not applicable.

ORCID

Mason McComb https://orcid.org/0000-0001-9871-8616
Robert Bies https://orcid.org/0000-0003-3818-2252
Murali Ramanathan https://orcid.org/0000-0002-9943-150X

REFERENCES

1. Wang Y. Division of Pharmacometrics. FDA 2018.
2. Ette EI, Williams PJ. Population pharmacokinetics I: background, concepts, and models. *Ann Pharmacother*. 2004;38(10):1702-1706.
3. Bauer RJ, Guzy S, Ng C. A survey of population analysis methods and software for complex pharmacokinetic and pharmacodynamic models with examples. *AAPS j*. 2007;9(1):E60-E83.
4. Duffull SB, Wright DF, Winter HR. Interpreting population pharmacokinetic-pharmacodynamic analyses - a clinical viewpoint. *Br J Clin Pharmacol*. 2011;71(6):807-814.
5. Bonate PL. Recommended reading in population pharmacokinetic pharmacodynamics. *AAPS j*. 2005;7(2):E363-E373.
6. Slob W, Janssen PH, van den Hof JM. Structural identifiability of PBPK models: practical consequences for modeling strategies and study designs. *Crit Rev Toxicol*. 1997;27(3):261-272.
7. Hsieh NH, Reisfeld B, Bois FY, Chiu WA. Applying a Global Sensitivity Analysis Workflow to Improve the Computational Efficiencies in Physiologically-Based Pharmacokinetic Modeling. *Front Pharmacol*. 2018;9:588.
8. Yates JW, Jones RD, Walker M, Cheung SY. Structural identifiability and indistinguishability of compartmental models. *Expert Opin Drug Metab Toxicol*. 2009;5(3):295-302.
9. Shah DK, Betts AM. Towards a platform PBPK model to characterize the plasma and tissue disposition of monoclonal antibodies in preclinical species and human. *J Pharmacokinet Pharmacodyn*. 2012;39(1):67-86.
10. FDA. Enhancing the Diversity of Clinical Trial Populations — Eligibility Criteria, Enrollment Practices, and Trial Designs Guidance for Industry. In: Research CfDEa, ed. 2019.
11. Owen JS, Fiedler-Kelly O. *Introduction to Population Pharmacokinetic/Pharmacodynamic Analysis With Nonlinear Mixed Effects Models*. Wiley; 2014.
12. Sun H, Fadiran EO, Jones CD, et al. Population pharmacokinetics. A regulatory perspective. *Clin Pharmacokinet*. 1999;37(1):41-58.
13. Davidian M, Giltinan DM. Nonlinear Models for Repeated Measurement Data: An Overview and Update. *JABES*. 2003;8(4):387-419.
14. Bai JPF, Earp JC, Pillai VC. Translational Quantitative Systems Pharmacology in Drug Development: from Current Landscape to Good Practices. *AAPS J*. 2019;21(4):72.
15. Helmlinger G, Sokolov V, Peskov K, et al. Quantitative Systems Pharmacology: An Exemplar Model-Building Workflow With Applications in Cardiovascular, Metabolic, and Oncology Drug Development. *CPT Pharmacometrics Syst Pharmacol*. 2019;8(6):380-395.
16. Bradshaw EL, Spilker ME, Zang R, et al. Applications of Quantitative Systems Pharmacology in Model-Informed Drug Discovery: Perspective on Impact and Opportunities. *CPT Pharmacometrics Syst Pharmacol*. 2019;8(11):777-791.
17. Iyengar R, Zhao S, Chung SW, Mager DE, Gallo JM. Merging systems biology with pharmacodynamics. *Sci Transl Med*. 2012;4(126):126-127.
18. Bloomingdale P, Nguyen VA, Niu J, Mager DE. Boolean network modeling in systems pharmacology. *J Pharmacokinet Pharmacodyn*. 2018;45(1):159-180.
19. Cosgrove J, Butler J, Alden K, et al. Agent-Based Modeling in Systems Pharmacology. *CPT Pharmacometrics Syst Pharmacol*. 2015;4(11):615-629.

20. Krzyzanski W, van Hasselt JGC. Special issue: mathematical pharmacology. *J Pharmacokinet Pharmacodyn*. 2018;45(1):1.
21. Ribba B, Dudal S, Lave T, Peck RW. Model-Informed Artificial Intelligence: Reinforcement Learning for Precision Dosing. *Clin Pharmacol Ther*. 2020;107(4):853-857.
22. Qian T, Zhu S, Hoshida Y. Use of big data in drug development for precision medicine: an update. *Expert Rev Precis Med Drug Dev*. 2019;4(3):189-200.
23. Mentre F, Friberg LE, Duffull S, et al. Pharmacometrics and Systems Pharmacology 2030. *Clin Pharmacol Ther*. 2020;107(1):76-78.
24. Chaturvedula A, Calad-Thomson S, Liu C, Sale M, Gattu N, Goyal N. Artificial Intelligence and Pharmacometrics: Time to Embrace, Capitalize, and Advance? *CPT Pharmacometrics Syst Pharmacol*. 2019;8(7):440-443.
25. Hutchinson L, Steiert B, Soubret A, et al. Models and Machines: How Deep Learning Will Take Clinical Pharmacology to the Next Level. *CPT Pharmacometrics Syst Pharmacol*. 2019;8(3):131-134.
26. Reker D, Shi Y, Kirtane AR, et al. Machine Learning Uncovers Food- and Excipient-Drug Interactions. *Cell Rep*. 2020;30(11):3710-3716.e3714
27. Talevi A, Morales JF, Hather G, et al. Machine Learning in Drug Discovery and Development Part 1: A Primer. *CPT Pharmacometrics Syst Pharmacol*. 2020;9(3):129-142.
28. Hinkson VI, Madej B, Stahlberg EA. Accelerating Therapeutics for Opportunities in Medicine: A Paradigm Shift in Drug Discovery. *Front Pharmacol*. 2020;11:770.
29. Mitchell TM, Carbonell JG, Michalski RS. *Machine Learning: A Guide to Current Research (The Springer International Series in Engineering and Computer Science)*. Kluwer Academic Publishers; 1986.
30. Hastie T, Tibshirani R, Friedman JH. *The elements of statistical learning: data mining, inference, and prediction*. 2nded. New York, NY: Springer; 2009.
31. James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning: with applications in R*. New York: Springer; 2013.
32. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018;15(4):233-234.
33. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Comput Graph Stat*. 2002;35(5–6):352-359.
34. Sperandei S. Understanding logistic regression analysis. *Biochem Med (Zagreb)*. 2014;24(1):12-18.
35. Cover T. Estimation by the nearest neighbor rule. *IEEE Trans Inf Theory*. 1968;14(1):50-55.
36. Abu Alfeilat HA, Hassanat ABA, Lasassmeh O, et al. Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. *Big Data*. 2019;7(4):221-248.
37. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273-297.
38. Kumar B, Vyas O, Vyas R. A comprehensive review on the variants of support vector machines. *World Sci Publ Company*. 2019;33(25):1950303.
39. Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V. Support Vector Regression Machines. *Adv Neural Inf Proces Syst*. 1997;9:155-161.
40. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Phys Rev*. 1958;65(6):386-408.
41. McCulloch WS, Pitts WA. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys*. 1943;5(4):115-133.
42. Hebb DO. The Organization of Behavior: A Neuropsychological Theory. 1949.
43. Schmidhuber J. Deep learning in neural networks: An overview. *CoRR*. 2014;abs/1404.7828.
44. Bonaccorso G. *Machine Learning Algorithms*. Livery Place, 35 Livery Street, Birmingham B3 2PB. UK: Packt Publishing Ltd; 2017.
45. Benitez JM, Castro JL, Requena I. Are Artificial Neural Networks Black Boxes? *IEEE Trans Neural Netw*. 1997;8(5):1156-1164.
46. Dayhoff JE, DeLeo JM. Artificial neural networks: opening the black box. *Cancer*. 2001;91(8 Suppl):1615-1635.
47. Heckerling PS, Gerber BS, Tape TG, Wigton RS. Entering the black box of neural networks. *Methods Inf Med*. 2003;42(3):287-296.
48. Rodvold DM, McLeod DG, Brandt JM, Snow PB, Murphy GP. Introduction to artificial neural networks for physicians: taking the lid off the black box. *Prostate*. 2001;46(1):39-44.
49. Sussillo D, Barak O. Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Comput*. 2013;25(3):626-649.
50. Zhang H, Weng TW, Chen PY, Hsieh CJ, Daniel L. Efficient Neural Network Robustness Certification with General Activation Functions. *arXiv*. 2018;1811.00866.
51. Breiman L, Friedman J, Stone C, Olshen R. Classification and Regression Trees. 1984.
52. Breiman L. Random Forests. *Mach Learn*. 2001;45(1):5-32.
53. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat*. 2001;29(5):1189-1232.
54. Freund Y, Scharpire R. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci*. 1997;55(1):119-139.
55. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *CoRR*. 2016;abs/1603.02754.
56. Duda RO, Hart PE. Pattern Classificationo and Scene Analysis. Stanford Research Institute: John Wiley & Sons; 1973.
57. Langley P, Sage S. Induction of Selective Bayesian Classifiers. Paper presented at: 10th Conference on Uncertainty in Artificial Intelligence (UAI-1994). 2013.
58. Hand D, Yu K. Idiot's Bayes - Not so stupid after all? 2001;69(3):385-398.
59. Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Mach Learn*. 1997;29(2–3):131-163.
60. Frank E, Trigg L, Holmes G, Witten I. Technical Note: Naive Bayes for Regression. *Mach Learn*. 2000;41(1):5-25.
61. Lloyd SP. Least squares quantization in pcm. *IEEE Trans Inf Theory*. 1982;28(2):129-137.
62. Anderberg MR. Cluster Analysis for Applications. 1973.
63. Reynolds D. Gaussian Mixture Models. *Encyclopedia of Biometrics*. 2009.
64. McComb M, Ramanathan M. Generalized Pharmacometric Modeling, a Novel Paradigm for Integrating Machine Learning Algorithms: A Case Study of Metabolomic Biomarkers. *Clin Pharmacol Ther*. 2020;107(6):1343-1351.
65. Pearl J. Bayesian Networks. In: Arbib M, ed. *Handbook of Brain Theory and Neural Networks*; 2003.
66. Pearl J. Probabilistic Reasoning in Intelligent Systems. 1988.
67. Koski T, Noble JM. A Review of Bayesian Networks and Structural Learning. *Math Appl*. 2012;40(1):53-103.
68. Scutari M. Learning Bayesian Networks with bnlearn R Package. *J Stat Softw*. 2010;35(3):1-22.
69. Gogate V, Webb W, Domingos P. Learning Efficient Markov Networks. Paper presented at: Advances in Neural Information Processing Systems 23 (NIPS 2010) 2010; Hyatt Residency, Vancouver, CA.
70. Koller D, Friedman N. Probabilistic Graphical Models. 2009.
71. Szepesvari C. *Algorithms for Reinforcement Learning*. Morgan & Claypool Publishers; 2009.
72. Moore B, Pyeatt L, Kulkarni V, Panousis P, Padrez K, Doufas A. Reinforcement Learning for Closed-Loop Propofol Anesthesia: A Study in Human Volunteers. *J Mach Learn Res*. 2014;15:655-696.
73. Padmanabhan R, Meskin N, Haddad WM. Reinforcement learning-based control of drug dosing for cancer chemotherapy treatment. *Math Biosci*. 2017;293:11-20.

74. Fokoue E. A Taxonomy of Big Data for Optimal Predictive Machine Learning and Data Mining. 2016.
75. Ristevski B, Chen M. Big Data Analytics in Medicine and Healthcare. *J Integr Bioinform*. 2018;15(3):20170030.
76. Johnstone IM, Titterington DM. Statistical challenges of high-dimensional data. *Philos Transact a Math Phys Eng Sci*. 2009;367 (1906):4237-4253.
77. Xu Y, Goodacre R. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *J Anal Test*. 2018;2(3):249-262.
78. Sherwin CM, Kiang TK, Spigarelli MG, Ensom MH. Fundamentals of population pharmacokinetic modelling: validation methods. *Clin Pharmacokinet*. 2012;51(9):573-590.
79. Musuamba FT, Bursi R, Manolis E, et al. Verifying and Validating Quantitative Systems Pharmacology and In Silico Models in Drug Development: Current Needs, Gaps, and Challenges. *CPT Pharmacometrics Syst Pharmacol*. 2020;9(4):195-197.
80. Alexander SPH, Kelly E, Mathie A, et al. The Concise Guide to PHARMACOLOGY 2019/20: Transporters. *Br J Pharmacol*. 2019; 176(Suppl 1):S397-S493.
81. Koch G, Pfister M, Daunhawer I, Wilbaux M, Wellmann S, Vogt JE. Pharmacometrics and Machine Learning Partner to Advance Clinical Data Analysis. *Clin Pharmacol Ther*. 2020;107(4):926-933.
82. Brier ME, Zurada JM, Aronoff GR. Neural network predicted peak and trough gentamicin concentrations. *Pharm Res*. 1995;12(3): 406-412.
83. Gobburu JV, Chen EP. Artificial neural networks as a novel approach to integrated pharmacokinetic-pharmacodynamic analysis. *J Pharm Sci*. 1996;85(5):505-510.
84. Chow HH, Tolle KM, Roe DJ, Elsberry V, Chen H. Application of neural networks to population pharmacokinetic data analysis. *J Pharm Sci*. 1997;86(7):840-845.
85. Yamamura S, Kawada K, Takehira R, et al. Artificial neural network modeling to predict the plasma concentration of aminoglycosides in burn patients. *Biomed Pharmacother*. 2004;58(4):239-244.
86. Smith NM, Lenhard JR, Boissonneault KR, et al. Using machine learning to optimize antibiotic combinations: dosing strategies for meropenem and polymyxin B against carbapenem-resistant Acinetobacter baumannii. *Clin Microbiol Infect*. 2020;26(9): 1207-1213.
87. Feretzakis G, Loupelis E, Sakagianni A, et al. Using Machine Learning Techniques to Aid Empirical Antibiotic Therapy Decisions in the Intensive Care Unit of a General Hospital in Greece. *Antibiotics (Basel)*. 2020;9(2):50.
88. Alexander SPH, Fabbro D, Kelly E, et al. The Concise Guide to PHARMACOLOGY 2019/20: Enzymes. *Br J Pharmacol*. 2019;176 (Suppl 1):S297-S396.
89. Stokes JM, Yang K, Swanson K, et al. A Deep Learning Approach to Antibiotic Discovery. *Cell*. 2020;180(4):688-702.e613
90. Tang J, Liu R, Zhang YL, et al. Application of Machine-Learning Models to Predict Tacrolimus Stable Dose in Renal Transplant Recipients. *Sci Rep*. 2017;7(1):42192.
91. Gim JA, Kwon Y, Lee HA, et al. A Machine Learning-Based Identification of Genes Affecting the Pharmacokinetics of Tacrolimus Using the DMET (TM) Plus Platform. *Int J Mol Sci*. 2020; 21(7):2517.
92. Van Looy S, Verplancke T, Benoit D, et al. A novel approach for prediction of tacrolimus blood concentration in liver transplantation patients in the intensive care unit through support vector regression. *Crit Care*. 2007;11(4):R83.
93. Thishya K, Vattam KK, Naushad SM, Raju SB, Kutala VK. Artificial neural network model for predicting the bioavailability of tacrolimus in patients with renal transplantation. *PLoS One*. 2018;13(4): e0191921.
94. Hu YH, Tai CT, Tsai CF, Huang MW. Improvement of Adequate Digoxin Dosage: An Application of Machine Learning Approach. *J Healthc Eng*. 2018;2018:3948245.
95. Poynton MR, Choi BM, Kim YM, et al. Machine learning methods applied to pharmacokinetic modelling of remifentanil in healthy volunteers: a multi-method comparison. *J Int Med Res*. 2009;37(6): 1680-1691.
96. Mitchell JB. Machine learning methods in chemoinformatics. *Wiley Interdiscip Rev Comput Mol Sci*. 2014;4(5):468-481.
97. 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*. 2001; 46(1–3):3-26.
98. Mishra H, Singh N, Lahiri T, Misra K. A comparative study on the molecular descriptors for predicting drug-likeness of small molecules. *Bioinformation*. 2009;3(9):384-388.
99. Cortes-Ciriano I. Benchmarking the Predictive Power of Ligand Efficiency Indices in QSAR. *J Chem Inf Model*. 2016;56(8): 1576-1587.
100. Yap CW, Chen YZ. Quantitative Structure-Pharmacokinetic Relationships for drug distribution properties by using general regression neural network. *J Pharm Sci*. 2005;94(1):153-168.
101. Zhu XW, Sedykh A, Zhu H, Liu SS, Tropsha A. The use of pseudo-equilibrium constant affords improved QSAR models of human plasma protein binding. *Pharm Res*. 2013;30(7):1790-1798.
102. Yuan Y, Chang S, Zhang Z, et al. A novel strategy for prediction of human plasma protein binding using machine learning techniques. *Chemom Intel Lab Syst*. 2020;199:103962.
103. Sun L, Yang H, Li J, et al. In Silico Prediction of Compounds Binding to Human Plasma Proteins by QSAR Models. *ChemMedChem*. 2018; 13(6):572-581.
104. Watanabe R, Esaki T, Kawashima H, et al. Predicting Fraction Unbound in Human Plasma from Chemical Structure: Improved Accuracy in the Low Value Ranges. *Mol Pharm*. 2018;15(11): 5302-5311.
105. Suenderhauf C, Hammann F, Huwyler J. Computational prediction of blood-brain barrier permeability using decision tree induction. *Molecules*. 2012;17(9):10429-10445.
106. Zhang X, Liu T, Fan X, Ai N. In silico modeling on ADME properties of natural products: Classification models for blood-brain barrier permeability, its application to traditional Chinese medicine and in vitro experimental validation. *J Mol Graph Model*. 2017;75: 347-354.
107. Brito-Sanchez Y, Marrero-Ponce Y, Barigye SJ, et al. Towards Better BBB Passage Prediction Using an Extensive and Curated Data Set. *Mol Inform*. 2015;34(5):308-330.
108. Yap CW, Li ZR, Chen YZ. Quantitative structure-pharmacokinetic relationships for drug clearance by using statistical learning methods. *J Mol Graph Model*. 2006;24(5):383-395.
109. Demir-Kavuk O, Bentzien J, Muegge I, Knapp EW. DemQSAR: predicting human volume of distribution and clearance of drugs. *J Comput Aided Mol des*. 2011;25(12):1121-1133.
110. Gobburu JV, Shelver WH. Quantitative structure-pharmacokinetic relationships (QSPR) of beta blockers derived using neural networks. *J Pharm Sci*. 1995;84(7):862-865.
111. Freitas AA, Limbu K, Ghafourian T. Predicting volume of distribution with decision tree-based regression methods using predicted tissue: plasma partition coefficients. *J Chem*. 2015;7(1):6.
112. Louis B, Agrawal VK. Prediction of human volume of distribution values for drugs using linear and nonlinear quantitative structure pharmacokinetic relationship models. *Interdiscip Sci*. 2014;6(1): 71-83.
113. Wang Y, Liu H, Fan Y, et al. In Silico Prediction of Human Intravenous Pharmacokinetic Parameters with Improved Accuracy. *J Chem Inf Model*. 2019;59(9):3968-3980.

114. Lu J, Lu D, Zhang X, et al. Estimation of elimination half-lives of organic chemicals in humans using gradient boosting machine. *Biochim Biophys Acta*. 2016;1860(11 Pt B):2664-2671.
115. Schneckener S, Grimbs S, Hey J, et al. Prediction of Oral Bioavailability in Rats: Transferring Insights from in Vitro Correlations to (Deep) Machine Learning Models Using in Silico Model Outputs and Chemical Structure Parameters. *J Chem Inf Model*. 2019;59(11):4893-4905.
116. Bloomingdale P, Mager DE. Machine Learning Models for the Prediction of Chemotherapy-Induced Peripheral Neuropathy. *Pharm Res*. 2019;36(2):35.
117. Tomberg A, Johansson MJ, Norrby PO. A Predictive Tool for Electrophilic Aromatic Substitutions Using Machine Learning. *J Org Chem*. 2019;84(8):4695-4703.
118. Cai Y, Yang H, Li W, Liu G, Lee PW, Tang Y. Computational Prediction of Site of Metabolism for UGT-Catalyzed Reactions. *J Chem Inf Model*. 2019;59(3):1085-1095.
119. Ekins S, Kortagere S, Iyer M, et al. Challenges predicting ligand-receptor interactions of promiscuous proteins: the nuclear receptor PXR. *PLoS Comput Biol*. 2009;5(12):e1000594.
120. Yang M, Chen J, Shi X, et al. Development of in Silico Models for Predicting P-Glycoprotein Inhibitors Based on a Two-Step Approach for Feature Selection and Its Application to Chinese Herbal Medicine Screening. *Mol Pharm*. 2015;12(10):3691-3713.
121. Prachayasittikul V, Prachayasittikul V. P-glycoprotein transporter in drug development. *EXCLI J*. 2016;15:113-118.
122. Bing H, Young L, Bing-Ke L, et al. Predicting and Virtually Screening Breast Cancer Targeting Protein HEC1 Inhibitors by Molecular Descriptors and Machine Learning Methods. *Biophys Chem*. 2015;31(9):1795-1802.
123. Shaikh N, Sharma M, Garg P. Selective Fusion of Heterogeneous Classifiers for Predicting Substrates of Membrane Transporters. *J Chem Inf Model*. 2017;57(3):594-607.
124. Sedykh A, Fourches D, Duan J, et al. Human intestinal transporter database: QSAR modeling and virtual profiling of drug uptake, efflux and interactions. *Pharm Res*. 2013;30(4):996-1007.
125. Poongavanam V, Haider N, Ecker GF. Fingerprint-based in silico models for the prediction of P-glycoprotein substrates and inhibitors. *Bioorg Med Chem*. 2012;20(18):5388-5395.
126. Baidya ATK, Ghosh K, Amin SA, et al. In silico modelling, identification of crucial molecular fingerprints, and prediction of new possible substrates of human organic cationic transporters 1 and 2. *New J Chem*. 2020;44(10):4129-4143.
127. Vilar S, Friedman C, Hripcsak G. Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media. *Brief Bioinform*. 2018;19(5):863-877.
128. FDA. FDA Adverse Event Reporting System. 2014.
129. Zhang W, Chen Y, Liu F, Luo F, Tian G, Li X. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinf*. 2017;18(1):18.
130. Medicine NLo. PubChem. 2020.
131. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):D1074-D1082.
132. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30.
133. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Res*. 2016;44(D1):D1075-D1079.
134. Liu S, Tang B, Chen Q, Wang X. Drug-Drug Interaction Extraction via Convolutional Neural Networks. *Comput Math Methods Med*. 2016;2016:6918381.
135. Hunta S, Yooyativong T, Aunsri N. A novel integrated action crossing method for drug-drug interaction prediction in non-communicable diseases. *Comput Methods Programs Biomed*. 2018;163:183-193.
136. Borella E, Poggesi I, Magni P. Prediction of the Effect of Renal Impairment on the Pharmacokinetics of New Drugs. *Clin Pharmacokinet*. 2018;57(4):505-514.
137. Davazdahemami B, Delen D. A chronological pharmacovigilance network analytics approach for predicting adverse drug events. *J am Med Inform Assoc*. 2018;25(10):1311-1321.
138. Raja K, Patrick M, Elder JT, Tsoi LC. Machine learning workflow to enhance predictions of Adverse Drug Reactions (ADRs) through drug-gene interactions: application to drugs for cutaneous diseases. *Sci Rep*. 2017;7(1):3690.
139. Wu HY, Chiang CW, Li L. Text mining for drug-drug interaction. *Methods Mol Biol*. 2014;1159:47-75.
140. Percha B, Garten Y, Altman RB. Discovery and explanation of drug-drug interactions via text mining. *Pac Symp Biocomput*. 2012;2012:410-421.
141. Holford N, Ma SC, Ploeger BA. Clinical trial simulation: a review. *Clin Pharmacol Ther*. 2010;88(2):166-182.
142. Gal J, Milano G, Ferrero JM, et al. Optimizing drug development in oncology by clinical trial simulation: Why and how? *Brief Bioinform*. 2018;19(6):1203-1217.
143. Allen RJ, Rieger TR, Musante CJ. Efficient Generation and Selection of Virtual Populations in Quantitative Systems Pharmacology Models. *CPT Pharmacometrics Syst Pharmacol*. 2016;5(3):140-146.
144. Rieger TR, Allen RJ, Bystricky L, et al. Even More Efficient Generation and Selection of Virtual Populations in Quantitative Systems Pharmacology Models. *J Pharmacokinet Pharmacodyn*. 2017;44:S115-S115.
145. Rieger TR, Allen RJ, Bystricky L, et al. Improving the generation and selection of virtual populations in quantitative systems pharmacology models. *Prog Biophys Mol Biol*. 2018;139:15-22.
146. Goldstein BA, Phelan M, Pagidipati NJ, Holman RR, Pencina MJ, Stuart EA. An outcome model approach to transporting a randomized controlled trial results to a target population. *J am Med Inform Assoc*. 2019;26(5):429-437.
147. Hutmacher MM, Kowalski KG. Covariate selection in pharmacometric analyses: a review of methods. *Br J Clin Pharmacol*. 2015;79(1):132-147.
148. Knights J, Chanda P, Sato Y, et al. Vertical Integration of Pharmacogenetics in Population PK/PD Modeling: A Novel Information Theoretic Method. *CPT Pharmacometrics Syst Pharmacol*. 2013;2(2):1-10.
149. Hall RG 2nd, Pasipanodya JG, Swancutt MA, Meek C, Leff R, Gumbo T. Supervised Machine-Learning Reveals That Old and Obese People Achieve Low Dapsone Concentrations. *CPT Pharmacometrics Syst Pharmacol*. 2017;6(8):552-559.
150. Bies RR, Muldoon MF, Pollock BG, Manuck S, Smith G, Sale ME. A genetic algorithm-based, hybrid machine learning approach to model selection. *J Pharmacokinet Pharmacodyn*. 2006;33(2):195-221.
151. Chen RTQ, Rubanova Y, Bettencourt J, Duvenaud D. Neural Ordinary Differential Equations. 2018.
152. Rackauckas C, Innes M, Ma Y, Bettencourt J, White L, Dixit V. DiffEqFlux.jl — A Julia Library for Neural Differential Equations.
153. Bunte K, Smith DJ, Chappell MJ, et al. Learning pharmacokinetic models for in vivo glucocorticoid activation. *J Theor Biol*. 2018;455:222-231.
154. Komorowski M. Clinical management of sepsis can be improved by artificial intelligence: yes. *Intensive Care Med*. 2020;46(2):375-377.
155. Gaweda AE, Muezzinoglu MK, Aronoff GR, Jacobs AA, Zurada JM, Brier ME. Individualization of pharmacological anemia management using reinforcement learning. *Neural Netw*. 2005;18(5–6):826-834.
156. Gaweda AE, Jacobs AA, Brier ME, Zurada JM. Pharmacodynamic population analysis in chronic renal failure using artificial neural

networks--a comparative study. *Neural Netw*. 2003;16(5–6): 841-845.
157. Barbieri C, Molina M, Ponce P, et al. An international observational study suggests that artificial intelligence for clinical decision support optimizes anemia management in hemodialysis patients. *Kidney Int*. 2016;90(2):422-429.
158. Barbieri C, Bolzoni E, Mari F, et al. Performance of a Predictive Model for Long-Term Hemoglobin Response to Darbepoetin and Iron Administration in a Large Cohort of Hemodialysis Patients. *PLoS One*. 2016;11(3):e0148938.
159. Escandell-Montero P, Chermisi M, Martinez-Martinez JM, et al. Optimization of anemia treatment in hemodialysis patients via reinforcement learning. *Artif Intell Med*. 2014;62(1):47-60.
160. Yauney G, Rana A, Wong LC, Javia P, Muftu A, Shah P. Automated Process Incorporating Machine Learning Segmentation and Correlation of Oral Diseases with Systemic Health. *Annu Int Conf IEEE Eng Med Biol Soc*. 2019;2019:3387-3393.
161. Ehrlinger J. ggRandomForests: Random Forests for Regression. *arXiv:150107196v2*. 2015.
162. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol*. 2014;14(1):137.
163. Warden P. How many images do you need to tain a neural network? In: WordPress.com, ed. *Pete Warden's Blog* 2017.
164. Hestness J, Narang S, Ardalani N, et al. Deep Learning Scaling is Predictable, Empirically. *arXiv:171200409*. 2017.
165. Joulin A, van der Maaten L, Jabri A, Vasilache N. Learning Visual Features from Large Weakly Supervised Data. *arXiv:151102251*. 2015.
166. Lei S, Zhang H, Wang K, Su Z. How Training Data Affect the Accuracy and Robustness of Neural Networks from Image Classification. *ICLR 2019 Conference Blind Submission* 2019.
167. Sucheston L, Chanda P, Zhang A, Tritchler D, Ramanathan M. Comparison of information-theoretic to statistical methods for gene-gene interactions in the presence of genetic heterogeneity. *BMC Genomics*. 2010;11(1):487.
168. Clark LT, Watkins L, Pina IL, et al. Increasing Diversity in Clinical Trials: Overcoming Critical Barriers. *Curr Probl Cardiol*. 2019;44(5): 148-172.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: McComb M, Bies R, Ramanathan M. Machine learning in pharmacometrics: Opportunities and challenges. *Br J Clin Pharmacol*. 2022;88(4):1482-1499. https://doi.org/10.1111/bcp.14801